# Inter-Model Analysis of Tsunami-Induced Coastal Currents

**Patrick J. Lynett[a], Kara Gately[b], Rick Wilson[c], Luis Montoya[a], Diego Arcas[e,f], Betul Aytore[g], Yefei Bai[h], Jeremy D. Bricker[i,z], Manuel J. Castro[j], Kwok Fai Cheung[h], C. Gabriel David[k], Gozde Guney Dogan[g], Cipriano Escalante[j], José Manuel González-Vida[m], Stephan T. Grilli[n], Troy W. Heitmann[h], Juan Horrillo[o], Utku Kânoğlu[p], Rozita Kian[g], James T. Kirby[q], Wenwen Li[r], Jorge Macías[j], Dmitry J. Nicolsky[s], Sergio Ortega[t], Alyssa Pampell-Manis[o], Yong Sung Park[u], Volker Roeber[i], Naeimeh Sharghivand[p], Michael Shelby[n], Fengyan Shi[q], Babak Tehranirad[q], Elena Tolkova[v], Hong Kie Thio[r], Deniz Velioğlu[g], Ahmet Cevdet Yalçıner[g], Yoshiki Yamazaki[h], Andrey Zaytsev[w], Y. Joseph Zhang[x]**

[a] University of Southern California, Tsunami Research Center, Los Angeles, CA 90089
[b] NOAA, National Tsunami Warning Center, Tsunami Research Center, Palmer, AK 99645
[c] California Geological Survey, Seismic Hazards Mapping Program - Tsunami Projects, Sacramento CA 95814
[d]University of Washington, Dept. of Applied Mathematics, Seattle, WA 98195
[e]NOAA Center for Tsunami Research, 7600 Sand Point Way NE, Seattle, WA 98115
[f]University of Washington, JISAO, 3737 Brooklyn Ave. NE, Seattle, WA 98105
[g]Middle East Technical University, Department of Civil Engineering, Ocean Engineering Research Center, Dumlupinar Boulevard, No:1 Cankaya, Ankara, Turkey, 06800
[h]University of Hawaii, School of Ocean and Earth Science and Technology, Department of Ocean and Resources Engineering, Holmes Hall 402, 2540 Dole Street, Honolulu, Hawaii, USA 96822
[i]Tohoku University, International Research Institute of Disaster Science (IRIDeS), 468-1 E304 AzaAoba, Aramaki, Aoba-ku, Sendai, Japan, 980-0845
[j]University of Málaga, Facultad de Ciencias, Departamento de Análisis Matemático, Campus de Teatinos, s/n, Málaga, Spain, 29080
[k]Franzius-Institute for Hydraulic, Estuarine and Coastal Engineering, Leibniz University of Hanover, Nienburger Straße 4, Hanover, Germany, 30167
[l]University of Washington, Dept. of Earth & Space Sciences, Seattle, WA 98195
[m]University of Málaga, E.T.S. Telecomunicación, Departamento de Matemática Aplicada, Campus de Teatinos, s/n, Málaga, Spain, 29080
[n]University of Rhode Island, Department of Ocean Engineering, Narragansett, RI 02882, USA
[o]Texas A&M University at Galveston, Tsunami Research Group, Offshore and Coastal System Engineering, 200 Seawolf Parkway, Galveston,TX, 77553
[p]Middle East Technical University, Department of Engineering Sciences, Dumlupinar Boulevard No:1, Cankaya, Ankara, Turkey, 06800
[q]University of Delaware, Center for Applied Coastal Research, Dept. of Civil and Environmental Engineering, Newark DE 19716 USA
[r]AECOM, 915 Wilshire Blvd, Los Angeles, CA 90017
[s]University of Alaska Fairbanks, Geophysical Institute, 903 Koyokuk Drive, Fairbanks, AK 99775-7320
[t]University of Málaga, Laboratorio de Métodos Numéricos, SCAI, Campus de Teatinos, s/n, Málaga, Spain,29080
[u]University of Dundee, Division of Civil Engineering, Perth Road, Dundee, United Kingdom, DD1 4HN
[v]NorthWest Research Associates, 4118 148th Ave NE Redmond, WA 98052-5164, USA
[w]Special Research Bureau for Automation of Marine Researches, Far Eastern Branch of Russian Academy of Sciences, 693013 Russia Uzhno-Sakhalinsk Gorkiy str. 25
[x]Virginia Institute of Marine Science, College of William & Mary, Center for Coastal Resource Management, 1375 Greate Road, Gloucester Point, VA 23062-1346, USA
[z]Current address: Delft University of Technology, Department of Hydraulic Engineering, Netherlands

## Abstract

To help produce accurate and consistent maritime hazard products, the National Tsunami Hazard Mitigation Program organized a benchmarking workshop to evaluate the numerical modeling of tsunami currents. Thirteen teams of international researchers, using a set of tsunami models currently utilized for hazard mitigation studies, presented results for a series of benchmarking problems; these results are summarized in this paper. Comparisons focus on physical situations where the currents are shear and separation driven, and are thus de-coupled from the incident tsunami waveform. In general, we find that models of increasing physical complexity provide better accuracy, and that low-order three-dimensional models are superior to high-order two-dimensional models. Inside separation zones and in areas strongly affected by eddies, the magnitude of both model-data errors and inter-model differences can be the same as the magnitude of the mean flow. Thus, we make arguments for the need of an ensemble modeling approach for areas affected by large-scale turbulent eddies, where deterministic simulation may be misleading. As a result of the analyses presented herein, we expect that tsunami modelers now have a better awareness of their ability to accurately capture the physics of tsunami currents, and therefore a better understanding of how to use these simulation tools for hazard assessment and mitigation efforts.

## Introduction

On March 9 and 10, 2015 in Portland, Oregon, the National Tsunami Hazard Mitigation Program (NTHMP) organized a benchmarking workshop to evaluate the numerical modeling of tsunami currents. Five different benchmarking datasets were organized, two of which will be discussed in detail in this paper. These datasets were selected based on characteristics such as: 1) geometric complexity; 2) currents that are shear/separation driven (and thus are de-coupled from the incident wave forcing); 3) tidal coupling; and 4) interaction with the built environment. Information about all of these benchmark problems can be found at the workshop website: http://coastal.usc.edu/currents_workshop/.

While tsunami simulation models have been well validated against wave height and runup (e.g. Synolakis et al., 2008), comparisons with speed data are much less common (e.g. Lynett et al., 2012). As model results are increasingly being used to estimate or indicate damage to coastal infrastructure (e.g. Yeh et al., 2005; Lynett et al., 2014; Chock, 2016; Chock et al., 2016; Tokimatsu et al., 2016), understanding the accuracy and precision of speed predictions becomes important. In the context of this paper, which focuses on coastal (not overland) currents, predictions of flow speed are integral for the estimation of forces on vessels (Okal et al., 2006; Suppasri et al., 2013), impacts at harbor structures such as floating docks (Keen et al., 2017), and prediction of sediment, boulder, and debris transport (Sugawara et al., 2014). Furthermore, as the structure under consideration exists (or will exist) at a unique location, the engineer or planner must have confidence that the tsunami speed prediction at that specific location is accurate. Thus, the tsunami community is left with the formidable challenge of providing a local speed prediction that should be accurate and precise enough to base a structural design on; yet, we know from observations of historical and recent tsunamis (e.g. Borrero et al., 2015 provides a review of events in harbors) that tsunami flow can be highly chaotic. It is a primary goal of this paper to present an assessment of accuracy and precision in simulating tsunami-induced currents using the state-of-the-art numerical models.

In this paper, we summarize the major results of the community modeling exercise. First, two of the five benchmark problems will be described in detail; these two problems will be the focus here. Next, information about the thirteen numerical models which simulated the benchmark problems is provided. Following this background, an extensive statistical analysis of the model-data and inter-model comparisons is presented, and conclusions and recommendations for future work are given

## Overview of Benchmark Problems

### *Benchmark Problem #1 (BM#1): Steady Flow over Submerged Obstacle*

This benchmark is based on work by Lloyd and Stansby (1997a,b) (L&S). While there are many controlled experimental datasets looking at the wake behind a cylinder, there are very few that examine the wake behind a sloping obstacle in the context of shallow water flow. While L&S performed a wide range of different island configurations, here we focus only on one where the obstacle (the island) was entirely submerged initially and throughout the experiment. Since the obstacle remains submerged at all times, the wake is physically generated through a spatially variable bottom stress (i.e. gradients in bottom friction). The aim of this benchmark is to test a

model's ability to generate a separation region and the resulting oscillatory wake, for an idealized and simplified case.

Test case SB4_02 in the Lloyd and Stansby Part II (1997b) paper is used. Steady upstream flow enters a long and shallow tank with the obstacle placed roughly in the mid position of the flume; see Figure 1 for a plan-view layout of the experimental tank. The steady upstream discharge (depth-averaged) velocity is $U = 0.115\ m/s$, water depth is $h = 0.054\ m$, the Reynolds number of the mean flow is $Re = 6210$, and the Froude number of the mean flow is $Fr = 0.16$. A conical island is placed on a flat bottom. The side slopes of the conical island are ~8 degrees, and the ratio of the water depth to the island height ($h/h_i$ in the L&S paper) = 1.10. The height of the island is 0.049 m and the diameter at the base of the island is 0.75 m. Using the island diameter at mid-height as the characteristic length scale relevant to shedding (following L&S), the Strouhal number for this wake is 0.33.

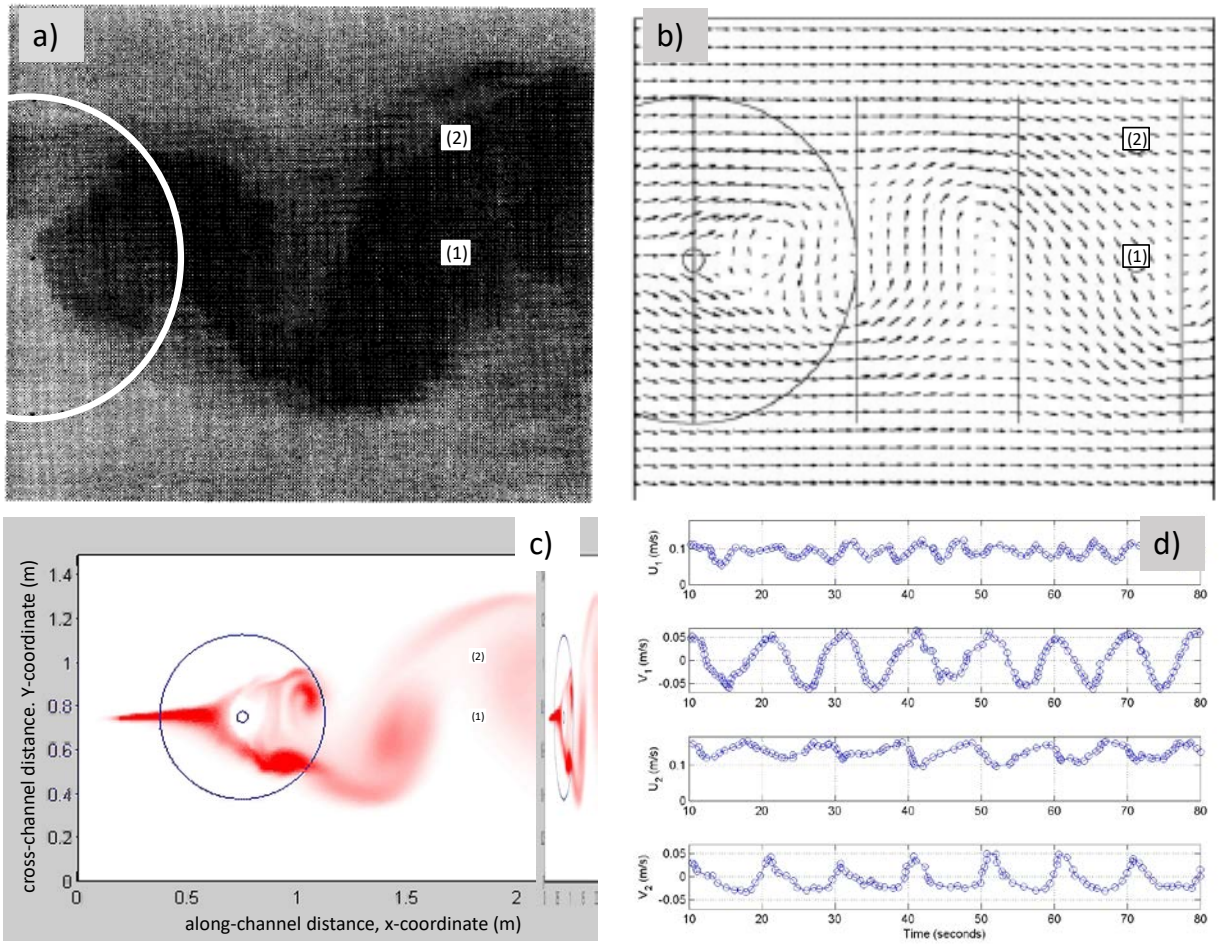


**Figure 1: Summary of numerical and experimental data from BM#1: a) is a dye visualization from the experiment (modified from L&S); b) shows the PIV-extracted surface velocity field from the experiment (modified from L&S); c) shows a numerical simulation including scalar dye transport to visualize the vortex street; and d) shows the experimental data at the two time series locations. In plots a), b) and c), the two time series locations are shown by the numbers.**

For this benchmark, the horizontal components of velocity at two different locations behind the island are compared. Data comparison locations are shown in Fig. 1; Point (1) is located along the centerline of the tank, 1.02 m behind the center of the island, and Point (2) is located at the same x-location as Point (1), but 0.27 m offset in the positive cross-tank (y) direction. The time series data is provided in Figure 1d), where $U_1$ and $V_1$ are the x- and y-components, respectively, of velocity at location (1) and $U_2$ and $V_2$ are the x- and y- components at location (2). There are clear oscillations in the velocity signal related to the vortex shedding, and the magnitude of these oscillations are, as expected, on the order of the mean flow speed. Also note that the individual oscillations vary, as no shedding event is the same as the previous one.

Modelers were asked to provide results for at least three different numerical configurations:

1. Simulation result with dissipation sub-models included, using the roughness information included in the L&S paper to best determine the friction factor. In the L&S paper, the friction factor is estimated to be 0.006 (as a dimensionless pipe-flow-like drag coefficient) or a Manning's n value of 0.01 $s/m^{1/3}$. If a Re-dependent friction factor formulation was used in a model, then a roughness height, $k_s$, of ~$1.5e^{-6}$ m is recommended.
2. Simulation result with optimized agreement based on tuning of dissipation model coefficients (e.g. friction factor). Note that this simulation could be skipped if the modelers did not wish to optimize their comparisons based on agreement with the experimental time series data.
3. Simulation result with all dissipation sub-models not included (e.g. a physically inviscid simulation). The purpose of this test is to understand the relative importance of numerical vs physical (as approximated by the governing equations used in each numerical model) effects on vorticity generation and dissipation for this class of comparison.

*Benchmark Problem #2 (BM#2): Tsunami Currents in Hilo Harbor*

This benchmark is based on a field dataset from the velocity data recorded in Hilo Harbor, Hawaii, resulting from the 2011 Tohoku tsunami. The aim of this benchmark is to understand the importance of model resolution and numerics on the prediction of tsunami currents. While modelers will aim to achieve the best agreement with the measured data, this is not the primary goal of this exercise. Some of the questions that this benchmark attempts to address include:

1. What level of accuracy and precision can we expect from a model with regard to modeling currents on complex bathymetry?
2. Will a model converge with respect to speed predictions and model resolution?
3. What is the variation across hydrodynamic different models (e.g. hydrostatic vs non-hydrostatic), using the same wave forcing, resolution, and bottom friction (or approximate equivalent when using different bottom stress models) ?

To attempt to most clearly answer these questions, this field case will be somewhat idealized, or reduced in complexity, to give the modeling results the best chance of an "apples-to-apples" comparison. For this benchmark, free surface elevation (from tide stations) and velocity information (from Acoustic Doppler Current Profiler (ADCPs)) are compared. Data for this benchmark test is discussed in detail in Arcos and LeVeque (2014) and Cheung et al. (2013).

Figure 2 shows a plot of the bathymetry from Hilo Harbor, Hawaii. The data is provided in [lat, long] on a 1/3 arcsec grid, taken from the NOAA NCEI Tsunami DEM database. Note that shown on this figure are also the simulation “control point” (white dot; upper-most dot in the figure), the two ADCP locations (black dots) and the tidal station (yellow dot; lower-most dot in the figure). As mentioned above, this problem has been "reduced" in an attempt to isolate differences in the employed incident wave forcing. For the bathymetry data, this "reduction" manifests as a flattening of the bathymetry at a depth of 30 meters; in the offshore portion of the bathymetry grid, there are no depths greater than 30 m.

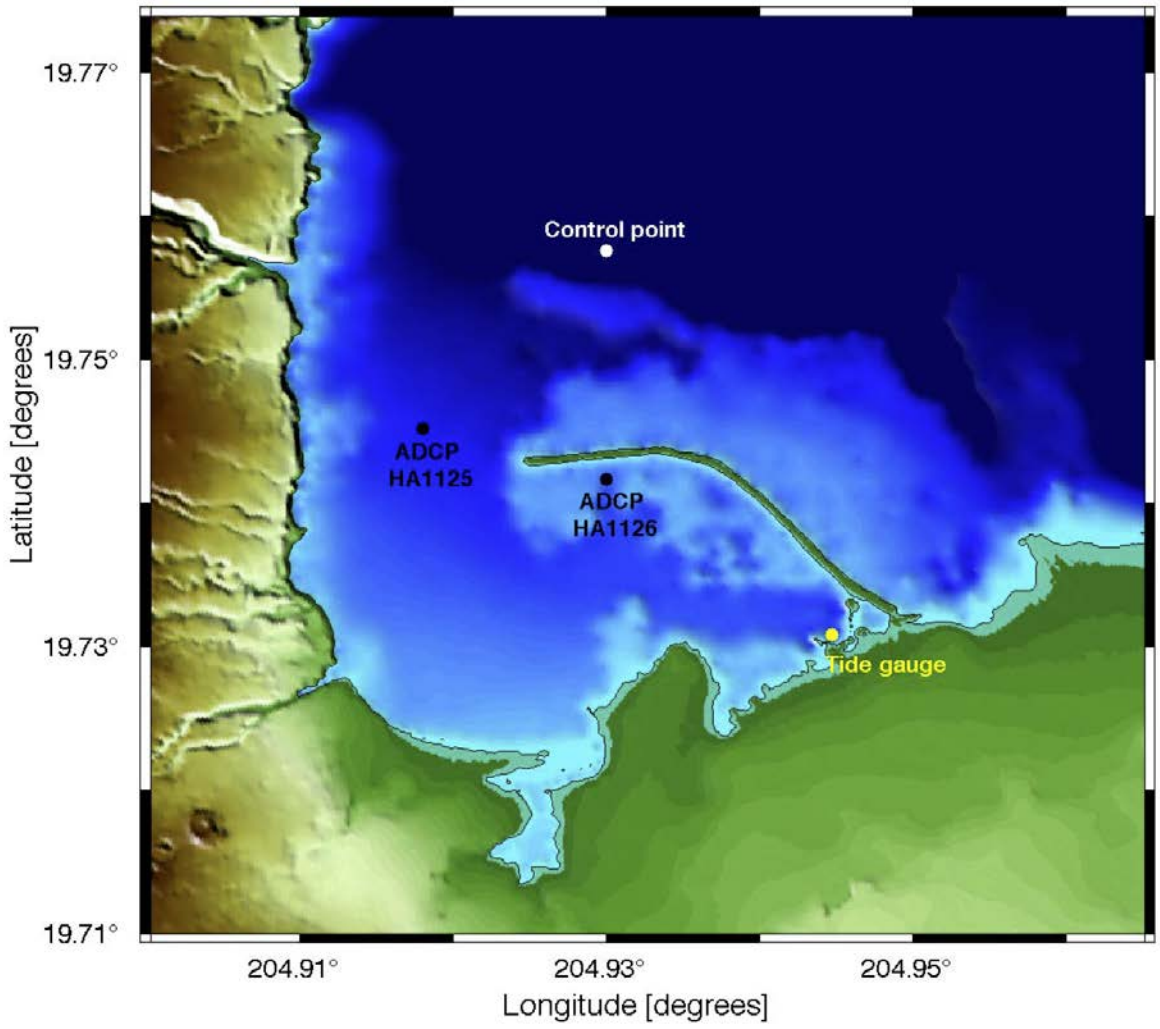


**Figure 2: Bathymetry data from Hilo Harbor, Hawaii. Shown on this figure are the simulation control point (white dot), the two ADCP locations (black dots) and the tidal station (yellow dot).**

For the incident wave, modelers were asked to drive their simulations with an offshore *simulated* free surface elevation time series from Cheung et al. (2013) (at the "control point"). Modelers were permitted to force their simulations in whichever way was convenient (e.g. through upper grid boundary, or with an internal source generator in the northern part of the domain), but should check with their modeled time series to ensure that they are generating the proper offshore wave

condition *at the control point*. Note that the above simplification will lead to a physical mismatch between the simulated and actual data; as the incident wave will vary spatially (albeit weakly) as it approaches the harbor. Again, the primary purpose here is inter-model evaluation.

The first comparison made by each modeler was for the water surface elevation at the tide station. Modelers shifted the simulated and recorded data such that the leading numerical wave arrives at the proper time, and this same time shift was used in the velocity comparisons. The ADCPs are named by NOAA as:

- HA1125, Harbor Entrance (referred to as HA25 hereafter)
- HA1126, Inside Harbor (referred to as HA26 hereafter)

Modelers were requested to provide results for at least three different numerical configurations:

1. Simulation result at ~20 m resolution (2/3 arcsec, de-sample the input bathymetry), using a Manning's n coefficient of 0.025 $s/m^{1/3}$
2. Simulation result at ~10 m (1/3 arcsec) resolution using a Manning's n coefficient of 0.025 $s/m^{1/3}$
3. Simulation result at ~5 m resolution (1/6 arcsec; use bi-linear interpolation), using a Manning's n coefficient of 0.025 $s/m^{1/3}$

Many of the numerical models tested here employed regular Cartesian grids for these simulations, and thus the targeted resolution is straightforward to enforce. For models that use irregular gridding (e.g. finite elements), modelers were requested to utilize meshes with nodal distances no less than the target resolutions.

## Overview of Models Tested

Table 1 presents a list of models that performed the benchmarking tests discussed in the previous section. These models are generally those that are in use for tsunami hazard mapping purposes internationally, although there is a strong concentration of models that have been developed in the USA. As this model set represents those that are used for mapping, the most common type of model found is of the depth-averaged or depth-integrated variety, as such models are practically applicable for the large domains (global scale to $km^2$-scale) necessary for tsunami inundation studies. The depth-integrated models fall into two classes: those solving the non-dispersive Nonlinear Shallow Water (NSW) equations and those solving the weakly dispersive Boussinesq-type equations. In addition to the depth-integrated models, in our group of 13 models we have three models that permit arbitrary flow variability in the vertical. Summary details of each model, such as equations solved and numerical accuracy, can also be found in Table 1. Brief descriptions and relevant references for each of the models are given here:

1) ALASKA GI'-T: The ALASKA GI'-T model stems from the TUNAMI model by Imamura (1995), which solves the NSW equations with a Leap-Frog numerical scheme. The primary modification in the ALASKA GI'-T model is the method used for the wetting/drying (Nicolsky et al., 2011). Currently, the ALASKA GI'-T model is used to predict propagation and runup of hypothetical tsunamis along the Alaska shore (Suleimani et al., 2013). The model is open source and freely distributable.

2) NAMI DANCE: The NSW equations with a bottom friction term are solved using the Leap-Frog numerical scheme (Imamura, 1989; Shuto et al., 1990). The model takes an input tsunami source from either a defined rupture, pre-determined wave form, or time history of water surface fluctuation at a grid boundary and computes propagation, coastal amplification, and inundation (e.g. Sozdinler et al., 2015, Dilmen et al., 2015). Complied executables for the model are freely distributable, but the source code is proprietary.
3) MOST: The MOST model has been developed over several decades with multiple co-authors (for a review see Titov et al., 2016, and references herein). The MOST model provides solutions to the NSW equations, including generation, propagation and inundation onto dry land (e.g. Wei et al., 2008; Gica et al., 2008). The model uses an explicit scheme to discretize the NSW equations, using an algorithm based on the method of fractional steps (Yanenko, 1971; Durran, 1999). Complied executables for the model are freely distributable, but the source code is proprietary.
4) Cliffs: Cliffs is an open-source relative of MOST described immediately above. This algorithm solves the fully nonlinear SWE by applying dimensional splitting to reduce the original 2-D problem to two sequential 1-D problems, and solves each 1-D problem in a characteristics form using an explicit finite-difference scheme. The primary difference between Cliffs and MOST is the method of tracking the land-water interface (Tolkova, 2014; Tolkova, 2016). Cliffs operates in Cartesian or spherical coordinates, in 2D and 1D configurations, allows unlimited one-way nesting and a variety of forcing options. The model is open source and freely distributable.
5) GEOCLAW: The open source GeoClaw software has been actively developed at the University of Washington and by collaborators elsewhere for over 10 years, starting with the work of George (2008) and George & LeVeque (2006). GeoClaw solves the NSW equations using a high-resolution shock-capturing finite volume method based on solving Riemann problems at cell interfaces and applying second-order correction terms with limiters to avoid non-physical oscillations near discontinuities. The spatial gridding scheme utilizes automatic mesh refinement (AMR). These general methods are described in detail in LeVeque (2002). The model is open source and freely distributable.
6) GEOCLAW-AECOM: GEOCLAW-AECOM is a modification of the GEOCLAW software described immediately above. The model equations and solution scheme are unchanged, with the exception that the "AECOM" version uses a constant set of nested grids for spatial refinement, while the original GEOCLAW uses AMR. The code is currently used for the development of probabilistic tsunami inundation maps for the State of California. The model is open source and freely distributable.
7) Tsunami-HySEA: Tsunami-HySEA solves the two-dimensional NSW system using a high-order (second and third order) path-conservative finite volume method. The code has the option of using a number of reconstruction operators and flux-limiters. Tsunami-HySEA has been employed for many tsunami prediction cases, and can include the effects of complex bathymetry and overland flow (e.g. Castro et al., 2005; Macías et al., 2015). The model is open source and freely distributable.
8) pCOULWAVE: pCOULWAVE solves the weakly-dispersive, rotational, and turbulent Boussinesq-type equations of Kim et al. (2009) using a high-order finite volume method. In terms of the dissipation mechanisms, bottom friction is included through a standard drag law, subgrid horizontal mixing is captured through a simple Smagorinsky closure, and vertical mixing is coarsely modeled following Elder (1959). A turbulence backscatter

model is employed in the model, allowing for the initiation of large-scale coherent features and a reverse-energy cascade (Kim and Lynett, 2011). The model is open source and freely distributable.

9) FUNWAVE-TVD: FUNWAVE-TVD is the present stage of evolution of a depth-integrated, fully nonlinear Boussinesq model originally proposed in Wei et al. (1995). The model was extensively rewritten in 2012 using a hybrid finite volume/finite difference TVD scheme in order to take advantage of the stability properties and shock-capturing capabilities of the approach, and to utilize Riemann solvers to improve wetting-drying processes during inundation. The work described here uses a version of the code developed for a Cartesian coordinate system, described by Shi et al. (2012) and employed in, for example, Grilli et al. (2015). The model is open source and freely distributable.
10) BOSZ: The BOSZ model is a tool for the computation of hazardous free surface flow problems ranging from near-field tsunamis to extreme swell waves such as those generated by hurricanes (Roeber and Cheung, 2012; Roeber and Bricker, 2015). The code solves a set of equations similar to those of Wei et al. (1995), using a finite-volume technique with a shock-capturing scheme involving a Riemann solver. The code allows for computations over irregular terrain with adaptive wet/dry mapping. The model source code can be obtained upon request from the developers.
11) NEOWAVE: NEOWAVE builds on the NSW equations with a vertical velocity term to account for weakly-dispersive waves and a momentum conservation scheme to describe flow discontinuities (Yamazaki et al., 2011). The vertical velocity term also facilitates modeling of tsunami generation from kinematic seafloor deformation. Conceptually, the code can use an arbitrary number of vertical levels to resolve vertical kinematics (Bai and Cheung, 2013), although in the simulations presented in this paper, only a single level is used. The staggered finite difference model can accommodate multiple sets of two-way nested grids with increasing resolution from the open ocean to the coast. The model source code can be obtained upon request from the developers.
12) TSUNAMI3D: TSUNAMI3D is a 3D Navier-Stokes (NS) solver based on a computational fluid dynamics (CFD) model originally developed at Los Alamos National Laboratory (LANL) during the 1970s (Hirt and Nichols, 1981). It solves transient fluid flow with free surface boundaries based on the fractional volume of fluid (VOF) method using an Eulerian mesh of rectangular cells. The code has been validated against standard inundation benchmarks (Horrillo et al., 2014) and has been applied for landslide tsunami simulation (e.g. Horrillo et al., 2013). The model is open source and freely distributable.
13) SCHISM: SCHISM (Semi-implicit Cross-scale Hydroscience Integrated System Model; Zhang et al, 2016) is a derivative product of SELFE (Semi-implicit Eulerian-Lagrangian Finite Elements; Zhang and Baptista, 2008) and is a general purpose model for geophysical fluid dynamics utilizing unstructured grids (e.g. Zhang et al. 2011). SCHISM solves 3D Reynolds-averaged Navier-Stokes equations in either hydrostatic or non-hydrostatic form using a hybrid Galerkin Finite-Element Method together with an Eulerian-Lagrangian Method. The model is open source and freely distributable.

We note that all of the model results discussed in this paper were submitted at the same time, with no post-submission modification or optimization. Furthermore, while the benchmark data was not provided to the modelers, both of the datasets will we discuss in this paper had been previously published, and thus was available beforehand. The comparisons were not blind comparisons.

**Table 1. Summary of models that used for the benchmark tests.**

| Model # | Model Name | Equations Solved [Spatial Dimensions] | Class | Numerical Approach | Numerical Treatment of Convection Terms | Numerical Accuracy of Other Gradient Terms | Numerical Treatment of Time Integration |
|---|---|---|---|---|---|---|---|
| 1 | ALASKA GI'-T | Nonlinear Shallow Water [2D] | I | FD | Upwind (1st-order accurate) | Centered (2nd-order accurate) | Semi-implicit (1st-order accurate) |
| 2 | NAMI DANCE | Nonlinear Shallow Water [2D] | I | FD | Upwind (1st-order accurate) | Centered (2nd-order accurate) | Explicit (2nd-order accurate) |
| 3 | MOST | Nonlinear Shallow Water [2D] | I | FD | Centered (2nd-order accurate) | Centered (2nd-order accurate) | Explicit (1st-order accurate) |
| 4 | Cliffs | Nonlinear Shallow Water [2D] | I | FD | Centered (2nd-order accurate) | Centered (2nd-order accurate) | Explicit (1st-order accurate) |
| 5 | GeoClaw | Nonlinear Shallow Water [2D] | I | FV | Limiter-based (1st-order near shocks, 2nd-order when smooth) | Centered (2nd-order accurate) | Explicit (2nd-order accurate) |
| 6 | GeoClaw - AECOM | Nonlinear Shallow Water [2D] | I | FV | Limiter-based (1st-order near shocks, 2nd-order when smooth) | Centered (2nd-order accurate) | Explicit (2nd-order accurate) |
| 7 | Tsunami-HySEA | Nonlinear Shallow Water [2D] | I | FV | Limiter-based (2nd-order near shocks, 3rd-order when smooth) | Centered (2nd-order accurate) | Explicit (3rd-order accurate) |
| 8 | pCOULWAVE | Highly Nonlinear Boussinesq-type [2D] | II | FV | Limiter-based (2nd-order near shocks, 4th-order when smooth) | Centered (4th-order accurate) | Semi-implicit (4th-order accurate) |
| 9 | FUNWAVE-TVD | Highly Nonlinear Boussinesq-type [2D] | II | FV / FD | Limiter-based (2nd-order near shocks, 5th-order when smooth) | Centered (4th-order accurate) | Explicit (3rd-order accurate) |
| 10 | BOSZ | Weakly Nonlinear Boussinesq-type [2D] | II | FV / FD | Limiter-based (2nd-order near shocks, 5th-order when smooth) | Centered (2nd-order accurate) | Explicit (4th-order accurate) |
| 11 | NEOWAVE | One-Layer, Non-Hydrostatic [2D] | III | FD | Upwind (1st-order near shocks, 2nd-order when smooth) | Centered (2nd-order accurate) | Semi-implicit (2nd-order accurate) |
| 12 | TSUNAMI3D | Navier-Stokes [3D] | III | FD | Upwind (1st-order accurate) | Centered (2nd-order accurate) | Explicit (2nd-order accurate) |
| 13 | SCHISM | Navier-Stokes, Hydrostatic [3D] | III | FE / FV | Limiter-based (1st-order near shocks, 2nd-order when smooth) | Centered (2nd-order accurate) | Semi-implicit (2nd-order accurate) |

**Note that in the "Numerical Approach" column, FD= Finite Difference, FV=Finite Volume, FE=Finite Element**

## Inter-comparison of Model Results

*Benchmark Problem #1: Steady Flow over Submerged Obstacle*

As mentioned in the background section above, modelers were requested to simulate BM#1 using a number of different dissipation (bottom friction) models. The desired outcome from these simulations is the prediction of a vortex street in the lee of the obstacle, with similar shedding frequency and eddy strength as observed in the experiments. It is important to re-iterate that the obstacle in this problem is entirely submerged during the experiment, and therefore the effects of moving shorelines and overland flow play no role here.

In general, simulations with all bottom friction turned off generated a chaotic and irregular vortex street, with very little resemblance to the experimental results. This is a reasonable and expected outcome; the generation of the vortex street and properties of the wake are strongly dependent on the interplay between bottom stress and the inertia of the flow, and poor description of the bottom stress should lead to a poor description of the resulting wake. Thus, the focus of discussion here will be on the numerical results provided with the "optimum" bottom friction coefficient, where here "optimum" is a subjective term with definition left to the discretion of the modeler. Indeed, it is expected that different modelers expended different levels of effort to achieve what they perceived to be the best agreement with the experimental data. Furthermore, the measures to be discussed in this section, used to assess model accuracy, were not provided to the modelers prior to their submission of "optimum" results.

A summary of the model resolutions and dissipation models used by all the modelers is given in Table 2. The minimum spatial resolutions found in the results are typically near 1 cm, with values as high as 2.5 cm. Clearly, the most common submodel used to approximate bottom stress is the Manning Equation. While the recommended Manning's "n", as provided in the original experimental paper, is 0.01 $s/m^{1/3}$, the majority of modelers found that a larger friction coefficient was required. The most common and median "n" value found in the results is 0.015 $s/m^{1/3}$ and employed values ranged from 0.01-0.02 $s/m^{1/3}$. With the expectation that the strength of numerical dissipation is proportional to the grid resolution (i.e. the finer the grid, the smaller the numerical dissipation – of course this is strongly dependent on the numerical scheme), it is possible that, in a statistical sense, relatively small friction factors might be correlated with relatively coarse grid sizes. Put more simply, physical dissipation may play a larger role as numerical dissipation plays a smaller one. Looking at the modeling results that used the Manning Equation for bottom friction, assessment of this potential trend is straightforward. While there is a correlation between grid size and Manning's "n", it is very weak ($R^2$=0.08). Thus, when comparing the different modeling results, variable numerical dissipation due to different grid resolutions is a second-order effect.

Other than the Manning Equation for bottom friction, four modelers used a quadratic friction law. Among those four results, three used a constant friction factor and one used a characteristic roughness height. When using a roughness height approach, the local friction factor is dependent on the local Reynolds number, and is therefore both temporally and spatially variable. For those using a constant friction factor, values ranged from 0.006-0.012; the recommended friction factor as provided in the original experimental papers is 0.006. For the single model that used a roughness height, friction factors ranged from 0.004-0.012 during the simulation.

In addition to dissipation through bottom friction, a few modelers also included dissipation through various horizontal and vertical eddy viscosity models. Due to the limited usage of such approaches within the models tested, and a lack of similar dissipation models used, the sensitivity and effect of such models will not be addressed here. In summary, we have a comprehensive set of numerical models used by the international tsunami community, with various physical and numerical properties. Next, we seek to understand how these various properties are related to model accuracy, and try to identify which of these properties are most relevant to accurate modeling of complex, tsunami-induced coastal currents.

**Table 2. Summary of numerical and physical parameters used for the BM#1 simulations.**

| Model # | Model Name | Class | Spatial Grid Size (m) | Bottom Stress Model | Bottom Stress Parameter | Other Turbulence Closure Models Used |
|---|---|---|---|---|---|---|
| 1 | ALASKA GI'-T | I | 0.01 | Manning Friction Coefficient | n=0.012 s/m$^{1/3}$ | |
| 2 | NAMI DANCE | I | 0.01 | Manning Friction Coefficient | n=0.010 s/m$^{1/3}$ | |
| 3 | MOST | I | 0.01 | Manning Friction Coefficient | on bottom, n=0.010 s/m$^{1/3}$ on island, n=0.017 s/m$^{1/3}$ | |
| 4 | Cliffs | I | 0.025 | Manning Friction Coefficient | n=0.015 s/m$^{1/3}$ | |
| 5 | GeoClaw | I | variable, from 0.01-0.076 | Manning Friction Coefficient | on bottom, n=0.000 s/m$^{1/3}$ on island, n=0.015 s/m$^{1/3}$ | |
| 6 | GeoClaw - AECOM | I | 0.0076 | Manning Friction Coefficient | on bottom, n=0.000 s/m$^{1/3}$ on island, n=0.015 s/m$^{1/3}$ | |
| 7 | Tsunami-HySEA | I | 0.0152 | Quadratic Drag Friction Law | $C_D$=0.006 | |
| 8 | pCOULWAVE | II | 0.015 | Roughness height model to determine $C_D$ based on Moody Diagram, Quadratic Drag Friction Law | $k_S$=0.015 mm ($C_D$ varies from 0.004 to 0.012 during simulation, function of Reynolds Number) | Smagorinsky model for horizontal mixing, Elder's model for vertical mixing, and backscatter model |
| 9 | FUNWAVE-TVD | II | 0.01 | Quadratic Drag Friction Law | $C_D$=0.012 | |
| 10 | BOSZ | II | 0.015 | Manning Friction Coefficient | n=0.020 s/m$^{1/3}$ | |
| 11 | NEOWAVE | III | 0.01 | Manning Friction Coefficient | n=0.010 s/m$^{1/3}$ | |
| 12 | TSUNAMI3D | III | 0.01 in x-y 0.0027 in z | No bottom stress sub-model used, no-slip boundary condition employed, numerically resolved boundary shear | Not applicable | kinematic viscosity = 1x10$^{-6}$m$^2$/s |
| 13 | SCHISM | III | 0.012 | Quadratic Drag Friction Law | $C_D$=0.006 | k-ε turbulence closure scheme |

As a reminder, the experimental data used to compare these models consists of horizontal velocity component time series at two locations; thus there are four separate time series to examine. To understand the ability of each model to recreate the experimental data, two primary measures will be used. First, the magnitude of the fluctuation of each component will be analyzed. To calculate this fluctuation, a zero-crossing technique is employed to first identify each of the individual oscillations in a de-meaned time series. For each oscillation, or segment of the time series between two successive zero-up crossing locations, the total fluctuation (or height) is calculated. Thus, for each time series, we have a set of fluctuations, and a mean fluctuation and a standard deviation of the fluctuations is determined. To interpret these statistical values, the mean fluctuation represents the "best estimate" from the time series while the standard deviation provides a measure of the chaotic nature of the flow.

When performing these calculations on the experimental data, we find that this data has a significant standard deviation, on the order of 10-20% of the mean fluctuation value. This implies that variation in eddy strength during shedding is indeed a physical property of this flow and bathymetry configuration. It is expected that these physical variations are due to small perturbations in the inlet flow profile, irregularities in the bathymetry, and chaotic variations of turbulence in the shear layers, leading to small asymmetries in these shear layers around the obstacle, and finally creating relatively large changes in the wake immediately behind the obstacle. We note this here because the large majority of the models used during this exercise do not include any perturbations in the inlet flow, bathymetrical profile, or small-scale turbulent fluctuations. Therefore any standard deviation of fluctuation found in the modeling results, and indeed the initial development of the vortex street itself, must be driven by numerical errors or gridding asymmetries. Among the models tested here, there was a wide variation in the time required to develop a vortex street, and this is likewise due to differences in numerical errors, gridding, and different turbulence closure approaches. To overcome this physical disconnect, a model might introduce random, small bathymetry perturbations, include a backscatter model, or use a turbulence closure model that requires some initial random seeding of turbulence. Finally, we remark that with geophysical scale simulations using measured bathymetry, the natural variability in bathymetry would likely introduce spatial perturbations in the flow sufficient to initiate vortex street-like instabilities.

Figure 3 shows the modeled velocity component fluctuations scaled by the respective experimental values. While there are a couple large-error outliers for each velocity component comparison, most models provide agreement with the experimental mean fluctuations to within 50% for all components. On the other hand, few models, less than 1/3 of the group, yield an error of 25% for each component, and only half of these models are accurate to within 10%. Also shown in this figure is the standard deviation of the fluctuations, for both the data (given by the horizontal dashed lines) and the models (given by the vertical solid lines). We remark that the standard deviation of the fluctuation is a measure of how much each eddy varies in its properties as compared to the other eddies, and thus requires some physical process that creates a perturbation in the shedding. Looking at the data, it is evident that all velocity components, except the V-component at time series location#1, exhibit standard deviations of 15% of the mean fluctuation. To interpret the modeled accuracy of the standard deviation, the vertical range of the modeled deviation (vertical length of black lines) should be equal to the vertical distance between the data deviations (vertical distance between the dashed lines). While it is noted that no specific measure of standard deviation

accuracy is provided here, it is quite clear that there is little skill in capturing this parameter amongst the models tested. However, this is arguably reasonable, as only two of the models (Model#8 with a backscatter model and Model#13 with a k-ε turbulence closure) contain physics that might permit prediction of such a statistic.

To summarize the data presented in Figure 3, the error for each model, averaged across the four velocity component time series, is shown in Figure 4. Also shown in this figure is the summary standard deviation, likewise averaged across the four time series. In the bottom plot of Figure 4 are shown the same statistical values for the time period of oscillation (i.e. the length of time between two successive zero-crossings). Clearly, there are models which perform substantially better than others in the measure for this benchmark, and model accuracy appears to be related to model complexity, where we consider complexity to be related to both numerical scheme and the physics included in the model equations. To this end, three model classes are defined:

- Class I: Models solving the nonlinear shallow water wave equations, with all types of numerical approaches; the first seven models listed in Table 1: ALASKA GI'-T, NAMI DANCE, MOST, Cliffs, GeoClaw, GeoClaw-AECOM, and Tsunami-HySEA
- Class II: Models solving a set of weakly dispersive equations using an analytical solution to the vertical kinematics, with all types of numerical approaches; Models pCOULWAVE, FUNWAVE-TCD, and BOSZ
- Class III: Models solving a set of equations where the vertical structure of the flow is not specified a-priori and can be hydrostatic or non-hydrostatic with respect to gravity waves, with all types of numerical approaches; Models NEOWAVE, TSUNAMI3D, and SCHISM

The average model accuracy in each of these "Classes" will now be examined. It is emphasized that the conclusions drawn by comparing these groups of models may not pertain to any individual model.

For Class I models tested here, the mean error in predicted fluctuation, averaged across the four time series, is 41%, and the mean standard deviation of the fluctuations is 13%. Interestingly, the Class-I-mean standard deviation of the fluctuation is very close to the observed value; however, none of the models in this group contain the physics to capture this phenomenon, and thus this is a fortuitous numerical error. The error in predicted fluctuation in the Class II models is reduced by 43% as compared to Class I, and therefore there is clearly an accuracy gain when moving from Class I to Class II for this benchmark, within the tested models. Moving to the last group, Class III, there is again a reduction in the error of 22% from Class II and 57% from Class I. This is an indication that for this type of problem, models without an a-priori description of the vertical flow structure (Class III) outperform other models, even those models solved with high-accuracy numerical schemes (e.g. Class II); low-order three-dimensional models are superior to high-order two-dimensional models. However, in the Class-averaged sense, neither Class II nor Class III yields a reasonable prediction of the standard deviation of the fluctuation.

Examining the error in the shedding period, shown in the lower plot in Figure 4, similar conclusions can be drawn. There is a clear decrease in error when moving from Class I to Class II models; however Class II and Class III models perform similarly. Additionally, the errors seen in this shedding period analysis are considerably less than those in the speed analysis, in the relative sense. This conclusion would imply that, for this flow and obstacle configuration, the effect of bottom roughness impacts the strength of the eddies to a greater degree than it impacts

the shedding frequency. We reiterate that while the above inter-Class comparisons are valid, there exist individual model violations of these comparisons, where, for example certain Class I models perform better than Class II or Class III models. Thus, the conclusions drawn here are meant to represent the abilities of the tsunami modeling community as a whole, as all of the models used in the exercise are currently used for operational, planning, and/or research applications.

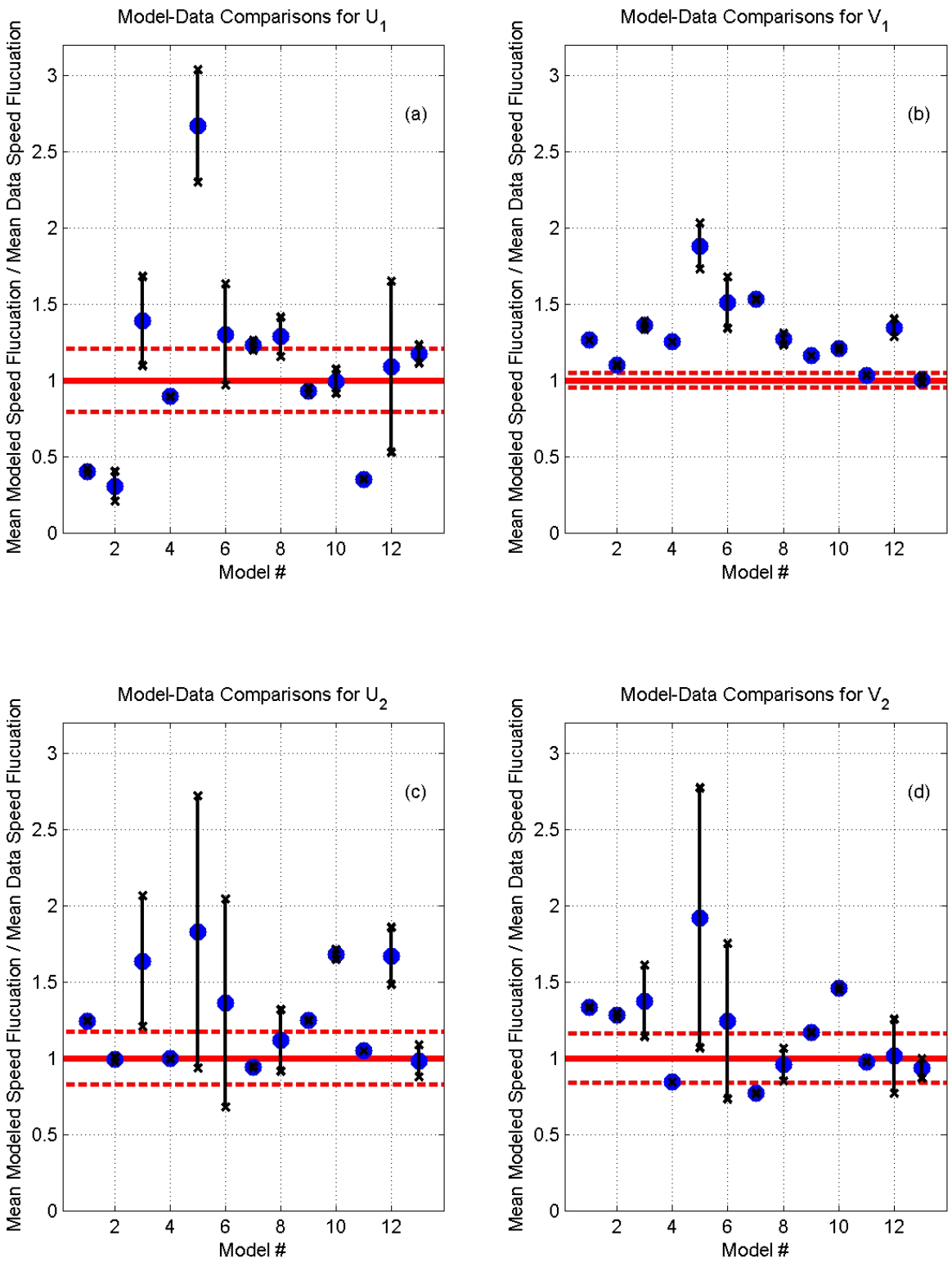


**Figure 3: Data-scaled modeled mean fluctuation (blue dots) and standard deviation of fluctuation (vertical black lines) for each model for U at location #1 (a), V at location#1 (b), U at location #2 (c), V at location#2 (d). The red-dashed horizontal lines show the standard deviation in the data.**

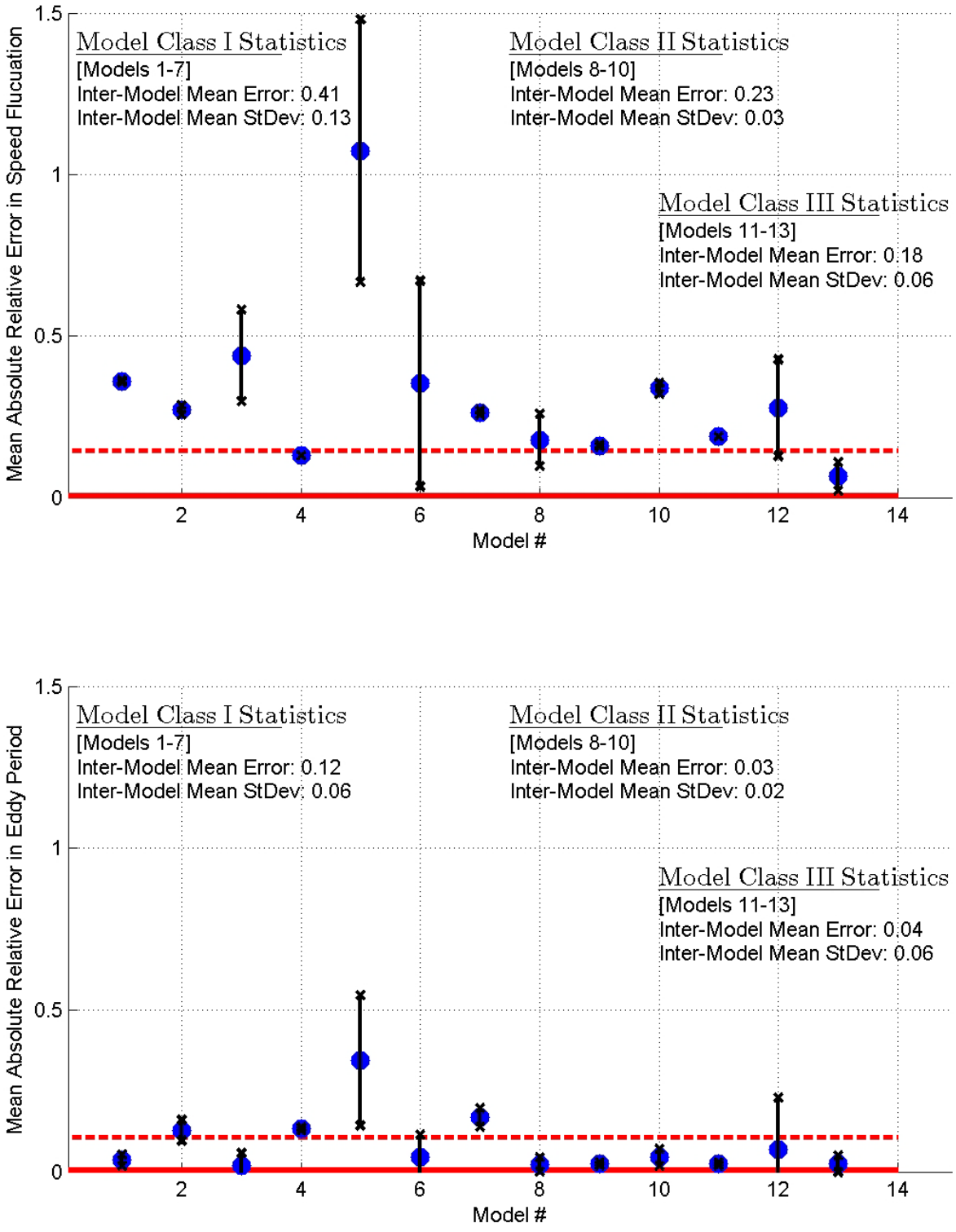


**Figure 4: Error in component-averaged modeled velocity, as a fraction of the experimental value, for the mean fluctuation (top) and period of oscillation (bottom). The data is presented in the same format as Figure 3.**

While using the fluctuation of speed components provides a method to assess a model's ability to predict the total range of speed, it does not necessarily assist in the understanding of whether a model is providing an accurate prediction of the kinetic energy of the flow, which may be more relevant for estimating hydrodynamic forces. To this end, we seek a statistical measure proportional to the square of the velocity. A time-averaged velocity squared, determined for each of the four velocity time series components, is calculated and plotted in Figure 5. What is immediately obvious from these comparisons is that the relative errors here are considerably larger than those found in the previous comparison. At both locations, all models tend to under predict the square of the U-components and all but two models over-predict the V-components. Such a broad and consistent trend in model errors either implies that there is an inconsistency between the experimental and numerical parameters (e.g. incorrect location or different upstream boundary conditions) or a fundamental deficiency in the numerical models examined here. While the former possibility may be statistically justified (due to the large majority of models showing this strong bias), the realization that one of the fully-3D models (Model 13) provides an high accuracy prediction of all four of these component-squared values offers a strong argument to the latter possibility.

It is reiterated here that the modeled errors in these quantities for most of the models are large, with the exception of the $U_2^2$ component, for which most models are within 30% accuracy (all under-predict). More than half of the models under-predict $U_1^2$ by 75%, 10 of the 13 models over-predict $V_1^2$ by 100% including three that over-predict by 200%, and 7 of the 13 models over-predict $V_2^2$ by 50%. The implications of such errors are particularly significant when one considers on-going attempts to use model results to estimate drag-like loads on marine structures (Suppasri et al., 2013; Keen et al., 2017). In these studies, both the magnitude and direction of the speed is important.

The average of the velocity-squared errors in the four components is provided in Figure 6 for each model. Only two models provide an average error in the component speed squared of less than 50%. The majority of models have an error between 50% and 100%, and there is a high-error outlier with an averaged error of 200%. Clearly, the models included here, with only a couple exceptions, struggle with this comparison, indicating that great care must be taken when trying to utilize model results for drag force estimations. Discussion within the tsunami community is required in order to justify some level of "required accuracy" in order for a specific model to be used in structural loading calculations. For example, setting an accuracy requirement of 100% (or 1.0 in Figure 6) would be inclusive, but a threshold this large and the associated safety-factors (or similar conservatism) that would be needed to reliably use the model output may regularly lead to over-design. On the other hand, specifying an accuracy requirement of 25% might reduce the need to add large safety factors, but would also imply only one of the tested models is acceptable. Since specification of an accuracy requirement based on the number of models that presently met the requirement is illogical, the tsunami community, with input from other engineers who would use the output for design of marine structures, should decide on a target accuracy for BM#1 comparisons. Models should be able to meet this target accuracy, while also demonstrating numerical convergence with the use of a reasonable bottom friction coefficient. Of course, other datasets that directly measure the loading of marine structures due to complex tsunami flows should also be used to validate a model's ability to provide force predictions; however such datasets currently do not exist.

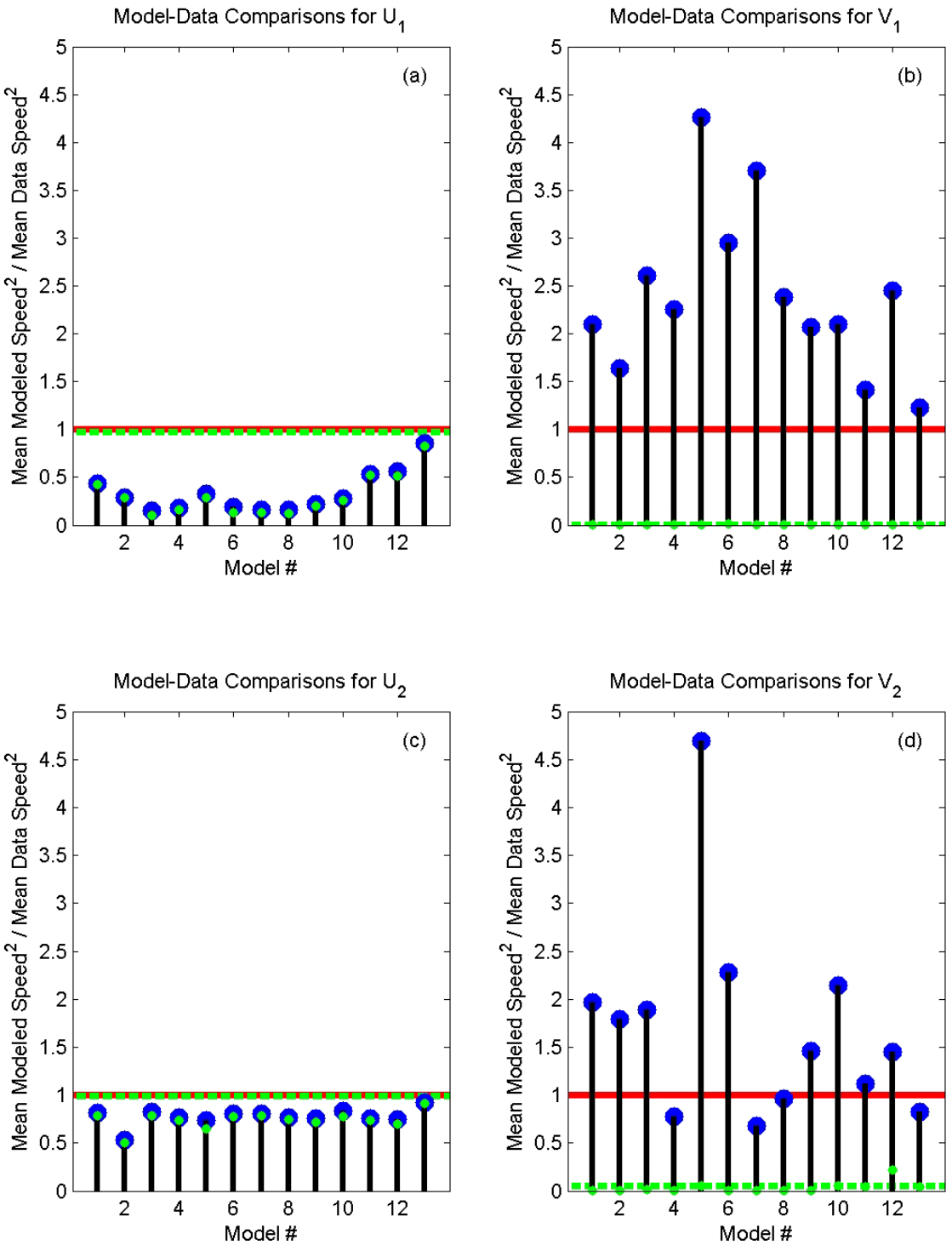


**Figure 5: Data-scaled modeled time-averaged speed squared (blue dots) and the square of the modeled mean speed (smaller green dots) for each model for U at location #1 (a), V at location#1 (b), U at location #2 (c), V at location#2 (d). The red horizontal line in each plot is data line (where, ideally, the blue dots would align), and the green horizontal dashed line is the square of the experimental mean speed (where, ideally, the green dots would align).**

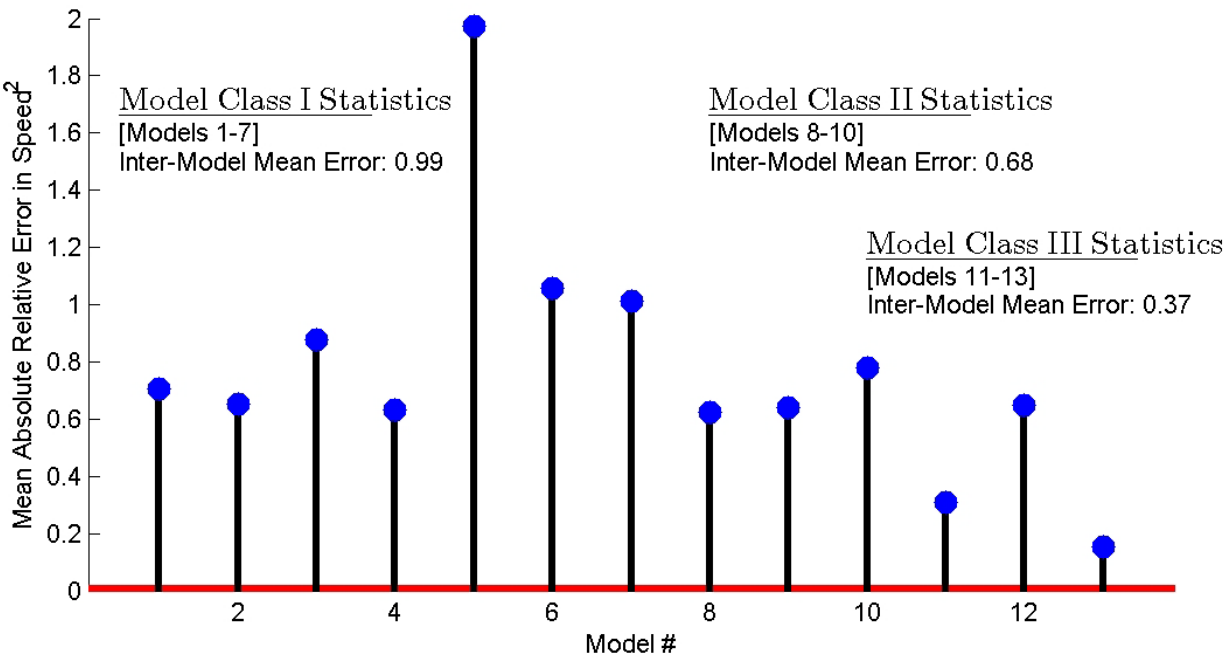


**Figure 6: Error in component-averaged modeled velocity squared, as a fraction of the experimental value.**

*Benchmark Problem #2: Tsunami Currents in Hilo Harbor*

As mentioned in the "Overview of Benchmark Problems" section, the primary purpose of BM#2 is to understand inter-model variability for a field-scale configuration. While the instrumentation observations in Hilo Harbor during the 2011 tsunami make this location unique in its number of closely located measurements, it is still difficult to use this case as a benchmark for demonstrating model accuracy. The reasons for this are two-fold. First, the ADCP time series data is sampled every six minutes, and every data point represents a six-minute average of the vertical profile of the current. The velocity field under a tsunami, particularly inside a harbor, can change quickly; studies suggest that a sample rate of a minute is necessary to resolve nearshore currents (e.g. Lynett et al., 2012) and potentially sub-minute if the flow is affected by eddies. Thus, there exists averaging-driven imprecision and possibly significant aliasing in the Hilo ADCP measurements due to the relatively coarse, discrete sampling; this effect is potentially much more significant, in a relative sense, than errors found in tide gage data and even runup measurements. Secondly, and with the ADCP imprecision in mind, this is a field-data case for which the initial condition and propagation over half of the Pacific Ocean includes uncertainty and error. Combining the far-field uncertainty and relatively large near-field imprecision leads to a situation where quantitative accuracy measures, as would be needed for a rigorous benchmark, become questionable. This is a remarkable statement in light of the fact that Hilo Harbor is, as mentioned, possibly the best instrumented location for tsunami-induced currents, and therefore implies higher resolution sampling of nearshore tsunami-induced currents is a great need.

The comparisons presented in this section will be divided into two parts: first, analysis of time series measurements at the tide gage and ADCP locations will be presented, followed by an examination of the spatial properties of the model output. Direct inter-model comparisons with time series data is difficult. The reason for this is that, typically, after the first few waves, model differences begin to accumulate, often taking the form of apparent phase errors. For example, two models may predict similar wave shapes, but with arrival time differences of minutes. With existing data and models, it is generally impossible to operationally predict precise arrival times of crests after the leading waves, due to model and data errors. For this reason, the time series comparisons presented here will focus on the envelope, where the envelope is defined as the line that connects the individual crests (or troughs) in the measured or modeled time series. Here, for the tide gage elevation data, a crest is defined as the maximum elevation of a discrete "wave", where a wave is defined as the data between two successive zero-up crossings.

Figure 7 provides a summary of the inter-model and model-data comparisons for the tide gage. Note that for this and all comparisons in this section, no individual models are directly compared with the data; only the inter-model means are shown with data. The top panel of Figure 7 shows each model-predicted ocean surface elevation time series, as well as the mean envelopes of crest and trough elevation. The middle panel again shows these modeled mean envelopes, but also shows the corresponding values from the tide gage. Note that there is a gap in the data starting near 10.3 hours after the earthquake; the gage did not function properly during this time according to the NOAA data record. In the lower panel, the inter-model standard deviation and the mean error in the crest envelope are summarized, as a function of time. Both are provided in relative terms, where the model deviation is scaled by the model-mean envelope, and the model error is scaled by the tide gage envelope. Examining the trend in inter-model standard deviation, we see that for the first four wave crests, the standard deviation remains low, shifting between 20-30% of the mean. After this time, however, the variation grows and fluctuates between 30-80% for the remainder of the time series. Trends in the mean-model error do not exhibit any clear temporal behavior, and are characterized by large (80%) errors both early and late in the examined time series. In a time-averaged sense, both the relative inter-model standard deviation and the mean model error have similar values of 20% during the first hour of the event, and then grow to 40% during the next three hours. These values may represent a precision threshold, and could be used to interpret potential errors in, for example, real-time model results.

In the same format as the elevations shown in Figure 7, Figure 8 provides a summary of the velocity time series comparisons measured at the ADCP locations HA25 (entrance channel) and HA26 (inside breakwater). To match the sampling method of the ADCP's, the numerical time series are filtered using a 6-minute moving average. The individual times series from all the models are given in the top panels in Figure 8, as well as the model-mean envelope from these results. From the individual model time series, we see that there is inter-model phase agreement for the first three or four speed peaks, but after this, phase correlation degrades quickly. This observation is in contrast with the modeled elevation time series shown in Figure 7, where inter-model phase correlation persists for the 4.5 hours of time displayed. The inter-model mean envelopes are shown in the middle plots with the ADCP envelope, and errors and inter-model variations are given in the bottom plots. First, looking at the inter-model variation, it is clear that this measure has a similar behavior with comparable values at the two locations, starting at a relatively low 20-30% for the first hour, and then exhibiting a general trend of increase, but with

values jumping between 20-100%. Throughout the event, the inter-model standard deviation in the speed envelope is between 0.1-0.3 m/s; this is an impressively low value in the context of predicted current speeds, and indicates that an ensemble-based, time-averaged speed envelope may be a stable statistic. However, these locations are minimally impacted by eddies in the simulations, which likely aids the stable inter-model statistics.

From the model error curves, a counter-intuitive result of large errors at HA25 and relatively small errors at HA26 is found. At HA25, there are a number of instances when the model-mean speed envelope predicts a speed greater than twice the measured value, as seen by model errors of 1.0 or greater in Figure 8. At HA26 on the other hand, for much of the compared record, the model error remains below 50% and is quite comparable to the inter-model variation. This is somewhat unexpected as the location of HA26 near to the tip of the harbor breakwater is more likely to be affected by eddies (to be discussed in more detail later), which should make for a much more challenging model-data comparison. However, as mentioned earlier, the model error should not be considered a significant comparison, due to the imprecision in the velocity data. We present the model-data comparisons here primarily to demonstrate the challenges in using existing speed data for model validation.

While time series comparisons are useful for understanding how errors and variations evolve temporally, they do not assist in the understanding of how quickly flow properties change spatially. To examine spatial variability, we use maximum predicted speed surfaces from each of the numerical models. Each surface provides the maximum speed predicted at each grid point throughout the simulation. Figure 9a) gives the inter-model mean maximum speed surface; to create this surface, each of the individual model surfaces is interpolated to the same spatial grid, and then the mean value at each point is determined from the stack of model surfaces. The greatest speeds are found in the area near to the tip of the breakwater, and along the coast where the depths are shallow; in these areas the model mean speeds are in excess of 4 m/s.

With the stack of individual model surfaces, the inter-model standard deviation of maximum speed can also be calculated at each grid point. These values are shown in Figure 9b). The largest standard deviations are also seen near the tip of the breakwater, and are associated with eddies of different strength taking different paths in the different models. The area affected by eddies is very clear with this statistic. The maximum deviations in the eddy area are near 2 m/s, and represent 50-150% of the model-mean speed. This inter-model comparison indicates that it is reasonable to expect large differences between models in areas affected by eddies, and thus velocity predictions from any single model in such locations must be carefully and conservatively interpreted.

There are two additional important observations from the speed deviation in Figure 9b). First, in areas not affected by eddies, the models show excellent convergence with inter-model standard deviations less than 0.2 m/s, often representing 10-20% of the mean speed. Thus, where currents are coupled with the wave (and not de-coupled in the form of eddies), models converge precisely. Second, deviations along the immediate shoreline and in areas of inundation are large. While not the focus of this study, this observation points out that speed predictions for overland flow, among the models tested here, are highly divergent. As model predictions for overland flow speed are being increasingly used for structural loading calculations, additional study is needed to quantify errors and variability in inundation velocities.

The properties of a modeled eddy (e.g. tangential speed, radial speed gradient) are dependent on the magnitude of the shear in the eddy generation area, or separation area. Numerically, this shear magnitude may be limited by the numerical resolution, as relatively coarse resolutions are unable to resolve strong velocity gradients. Thus, there may be a strong connection between the numerical signature of an eddy and the numerical resolution. Figures 9c-e) show the relative inter-model deviations, scaled by the inter-model mean speed, for three different resolutions. Clearly, the local speed deviations grow with decreasing grid size. The reason for this appears to be that, as grid size decreases, models have a tendency to generate eddies with a larger radial velocity gradient. A larger radial velocity gradient equates to rapid spatial changes in fluid speed, and increased sensitivity to the precise path of the eddy. A troublesome conclusion for modeling follows: in eddy areas, many models will increasingly diverge with decreasing grid length, at least through the horizontal resolutions tested here (down to 5 m). A clear example of how this divergence may manifest is shown in Figure 10. This figure provides vertical vorticity snapshots at the same time from the same numerical model, but using two different resolutions. The eddies in the finer resolution snapshot appear smaller with stronger vorticity, in line with the discussion above. Also, the large eddy located in the harbor entrance takes two different paths in the two simulations. In the 10-m resolution simulation, the eddy passes through the HA25 ADCP location, but in the 5-m simulation the eddy stays to the south of the ADCP. This is an example of the trouble with in-situ point measurements of tsunami currents in areas with eddies, and is another argument for the use of ensemble means for speed predictions, based on many realizations of potential eddies.

An alternative to multi-model or multi-realization ensembles is to spatially average the output from a single model, i.e. attempt to average out the local impact of an eddy. Such an approach would have the disadvantage of smoothing peak speeds associated with an eddy, but might provide numerically convergent results with a single model. Figure 11 displays an attempt at spatial averaging. Here, we spatially average across two lengthscales: (1) an “eddy scale” which here is estimated to be 100 m based on simulation output and (2) ten times the “eddy scale” or 1 km. These averaging areas, delimited by boxes with sides equals to the selected length scales, are shown in the top plot of Figure 11. The box-average maximum speed for each of the individual models are plotted for both box sizes. For each model, the box-average for the 20-m and 10-m resolutions are shown, scaled by the box-average from the 5-m resolution result. Thus, if the 10-m (or 20-m) resolution box-average has numerically converged, it would plot at 1.0 along the vertical scale. Clearly, using the “eddy scale” as the averaging lengthscale does not lead to a set of models that demonstrates convergence at 10-m resolution. The reason for this is that small-scale averages are sensitive to the precise path of the eddy, which varies among different models and among different resolutions for the same model. Should for a particular model and a particular box location the eddy exist in the box at one resolution and not another, an especially poor local convergence will result. On the contrary, averaging on the large lengthscale, here 1 km, is not sensitive to variations in eddy’s trajectories, since the larger box encompasses these most or all trajectories. As a result, we see broad convergence across the tested models at 10-m resolution. However, averaging model speed predictions over a 1 $km^2$ area, while suitable for demonstrating a form of individual model convergence, yields output of very limited use for hazard assessment, where local maximum might govern hazardous conditions and damage potential.

Following the analysis above, an ensemble mean of maximum speed is likely to be the a robust and informative modeling product, with significant and decision-impacting benefits over single model, deterministic predictions. The ensemble produced in this paper used many different models; however it is certainly possible to generate a spectrum of realizations with a single numerical model using a distribution of perturbations to initial conditions, bathymetry, bottom roughness, etc. Some research would be needed to specify an appropriate set of perturbations. If an ensemble was available, it is not obvious what the most useful way to present the statistical information would be. Certainly, means and deviations could be provided, but this might require expert-level judgement to use in decision making. An alternative would be a "threshold map", an example of which is shown in Figure 12. This map provides the "chance" that any location might experience a maximum current greater than a set threshold (the threshold is 2 m/s in Figure 12). The "chance" is based on how many of the models in the ensemble predict a speed greater than the threshold. The advantage of such a map is that it provides both a speed magnitude and confidence together, allowing for discussions of hazard informed decisions.

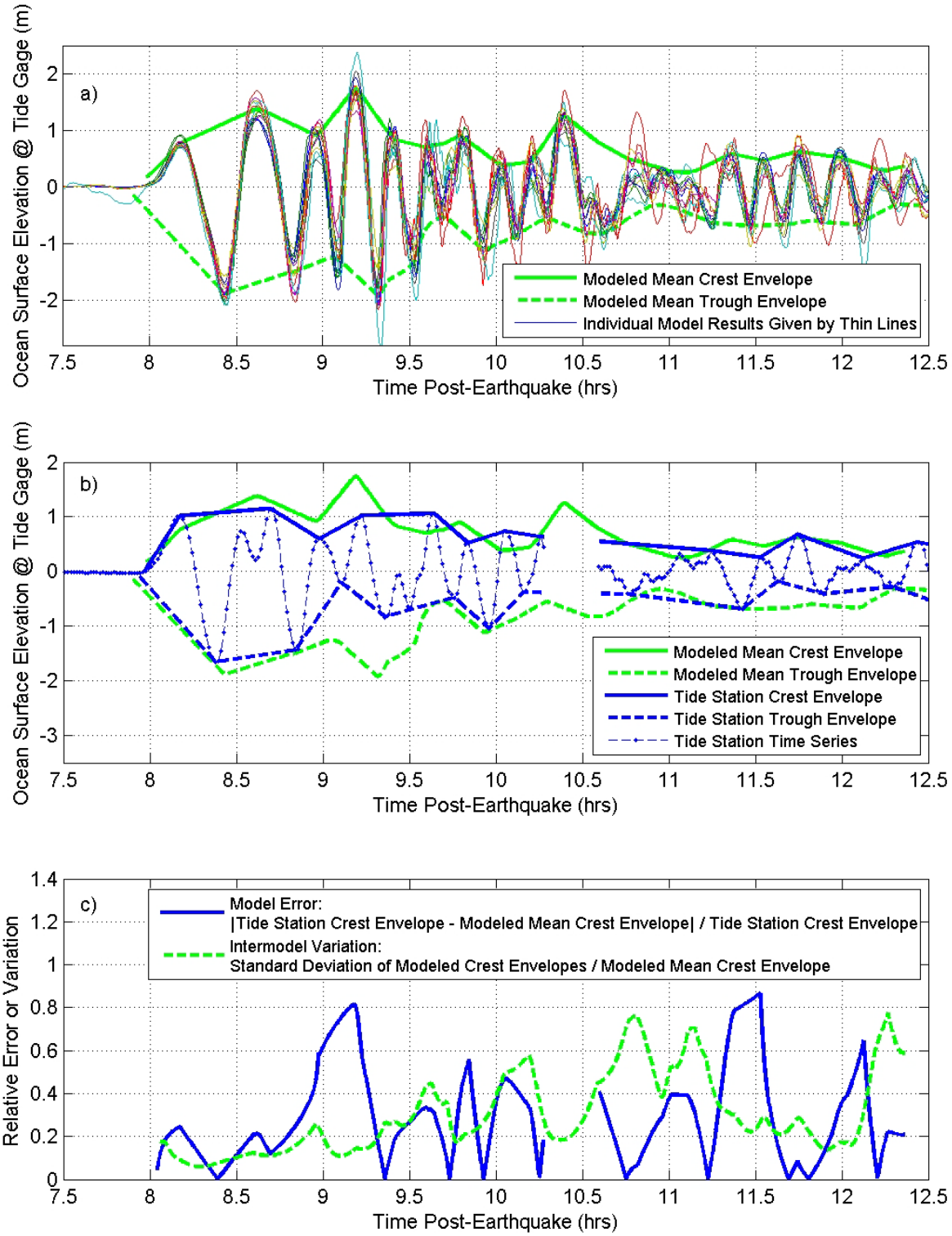


**Figure 7: Inter-model variability and error for ocean surface elevation measured at the tide gage location. Top plot a): Predictions from all models (thin solid lines), inter-model mean crest envelope (thick solid line), and inter-model mean trough envelope (thick dashed line). Middle plot b): Comparison of inter-model envelope to measured tide station data envelope; also shown in the time series from the measured data. Bottom plot c): mean inter-model error (solid line) and intermodal standard deviation (dashed line) for the crest envelope.**

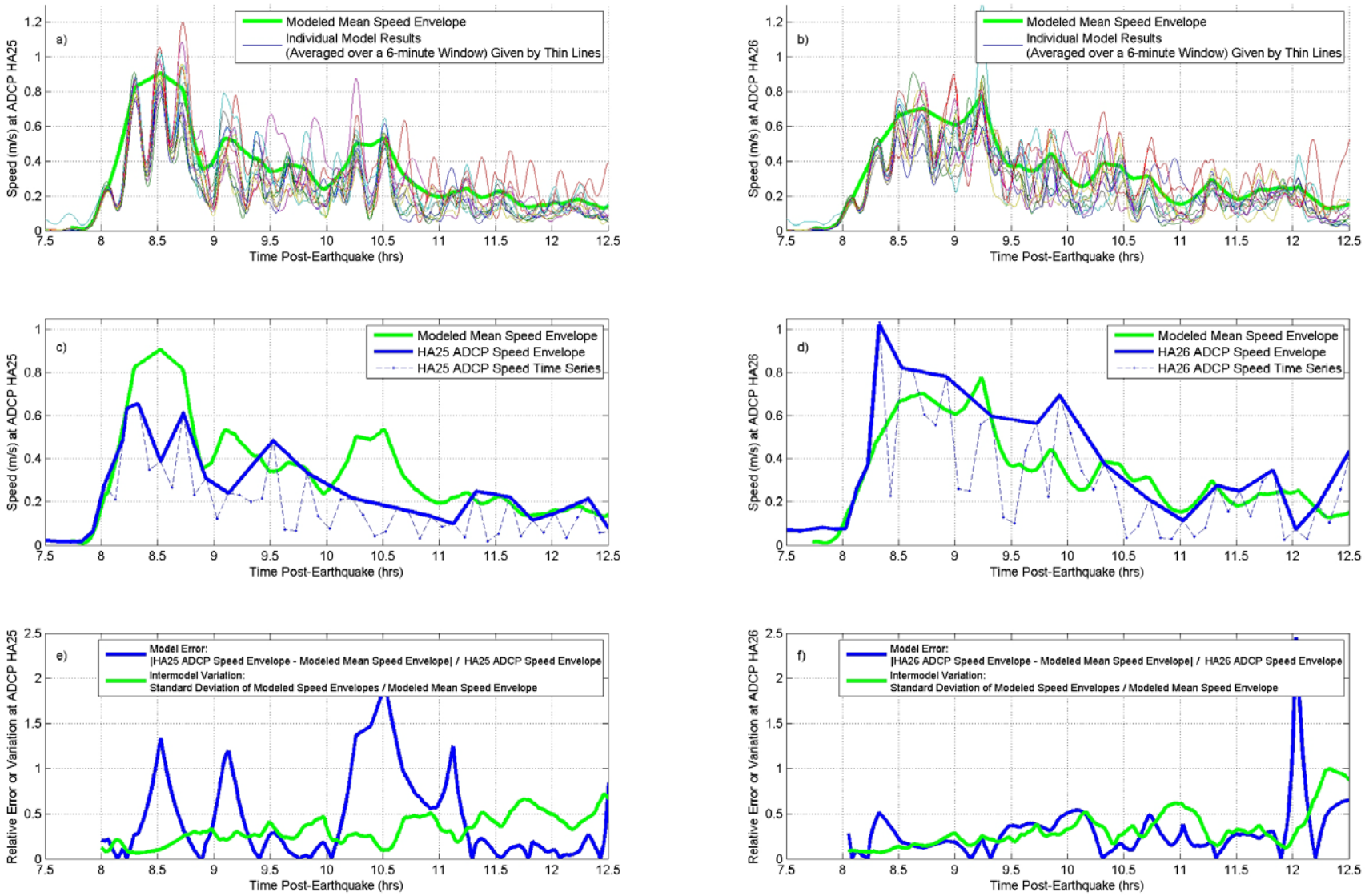


**Figure 8: Inter-model variability and error for ocean current speed measured at the ADPC location HA25 (left column) and HA26 (right column). Top row plots a) and b): Predictions from all models (thin solid lines) and inter-model mean speed envelope (thick solid line). Middle row plots c) and d): Comparison of inter-model envelope to measured ADCP data envelope; also shown in the time series from the measured data. Bottom row plots e) and f): mean inter-model error (blue line) and intermodal standard deviation (green line) for the speed envelope.**

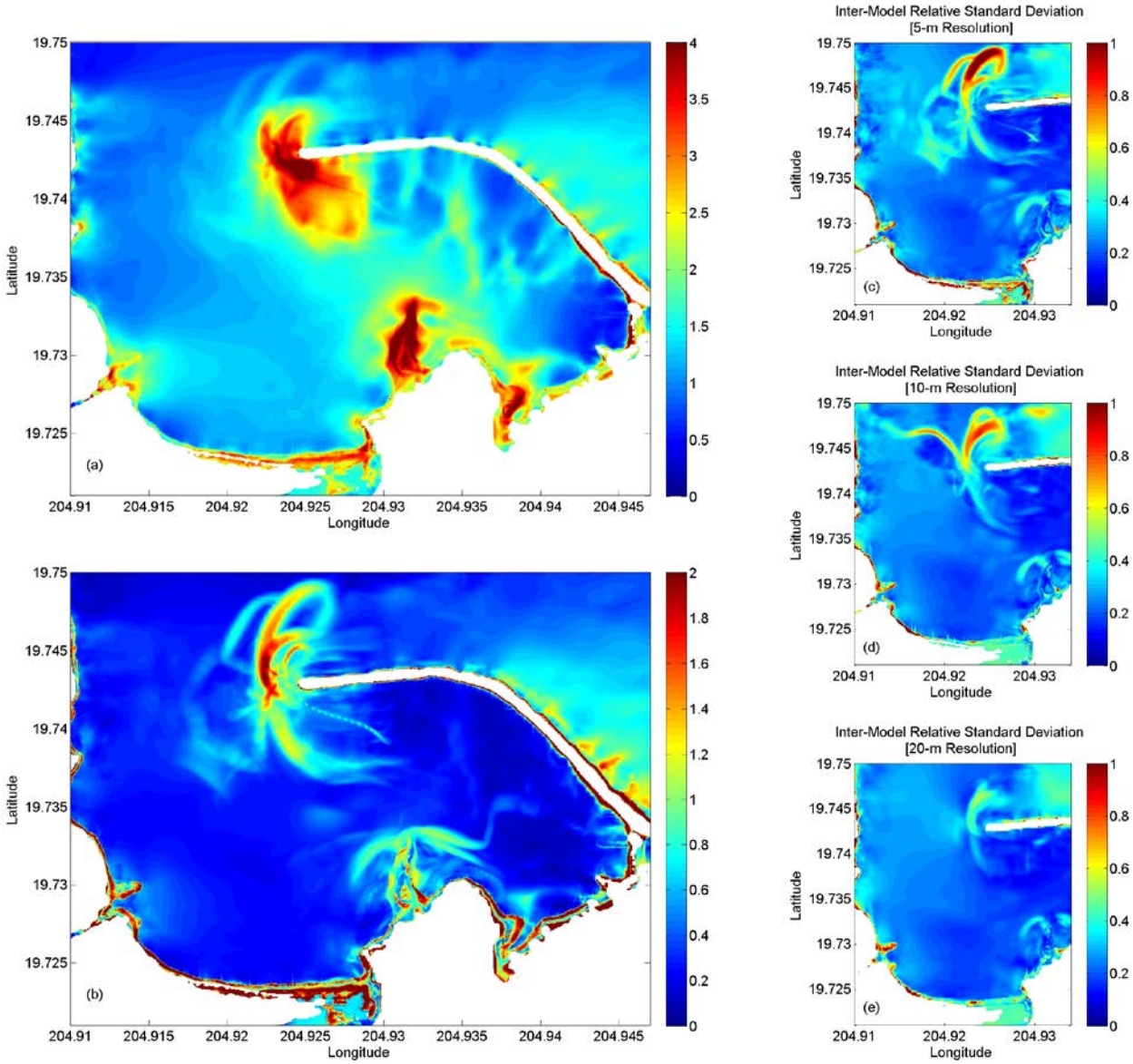


**Figure 9: Summary of inter-model spatial statistics. Top left a): Inter-model mean of predicted maximum speed in m/s as taken from the 5-m resolution runs. Bottom left b): Inter-model standard deviation of predicted maximum speed in m/s as taken from the 5-m resolution runs. Right column: Inter-model standard deviation of predicted maximum speed scaled by model-mean maximum speed as taken from the 5-m resolution runs (c), 10-m resolution runs (d), and 20-m resolution runs (e).**

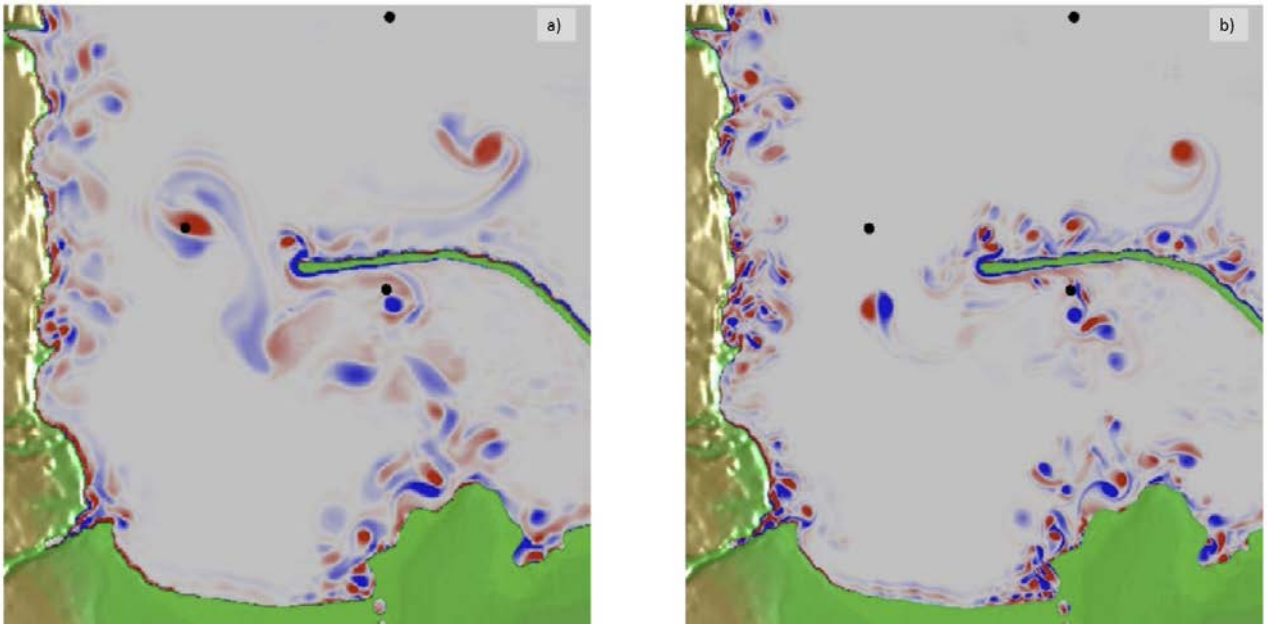


**Figure 10: Snapshots of vertical vorticity at the same simulation time for the same numerical model, with 10-m resolution (a) and 5-m resolution (b). Note that the locations of the ADCP's are shown by the black dots.**

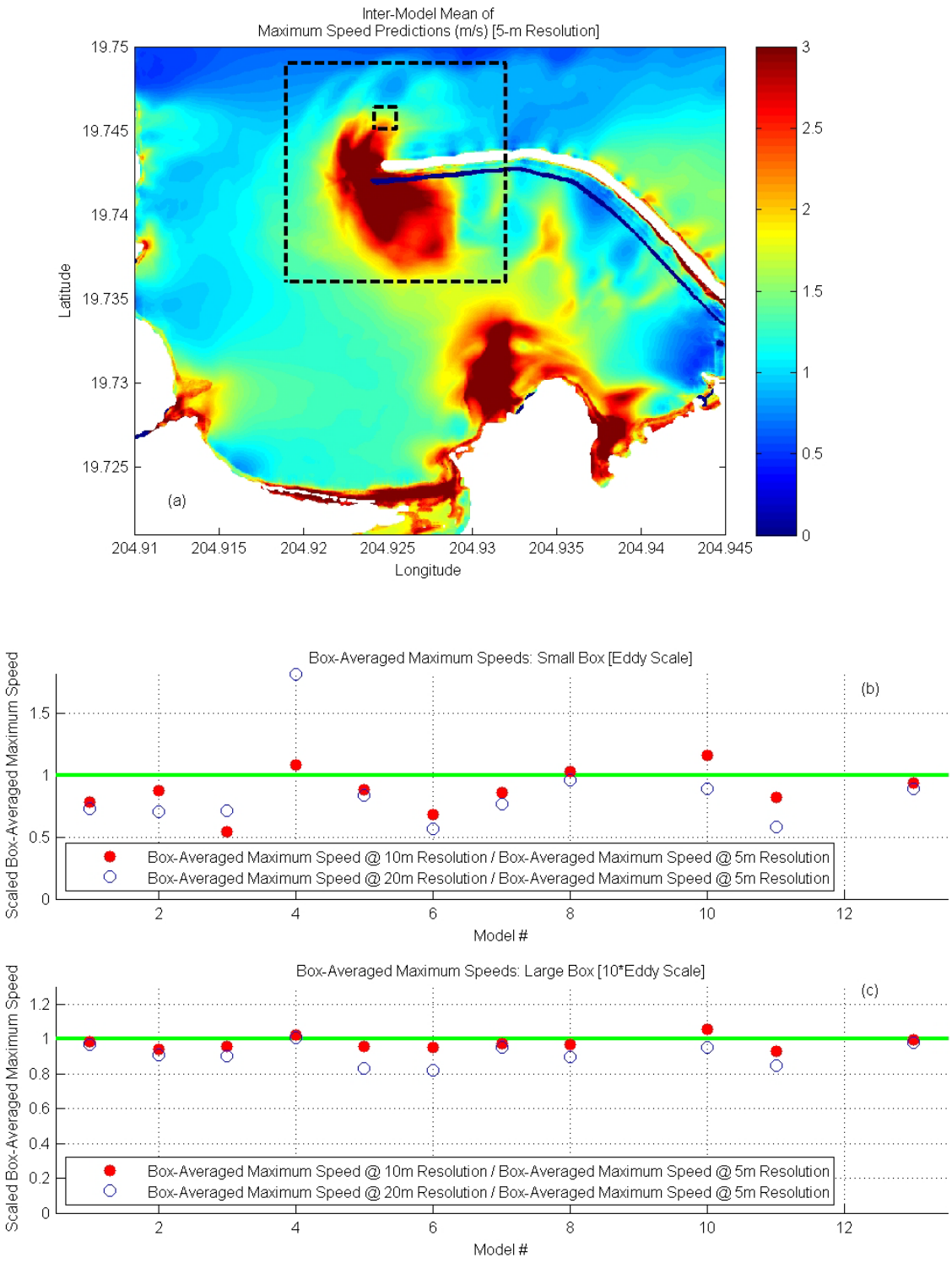


**Figure 11: The effect of spatial-averaging on model convergence, with the top plot (a) indicating the locations of the "small box" and the "large box" as referenced in the two lower subplots. The two lower subplots show the box-averaged maximum speeds for each model at 20- and 10-m resolutions in (b) and (c), scaled by the box-averaged speed from each model's 5-m resolution simulation; middle subplot for the small box, and lower subplot for the large box. Note that two model results are missing; one due to the use of different boundary conditions with different resolutions, and the other due to an inability to perform a 5-m resolution simulation.**

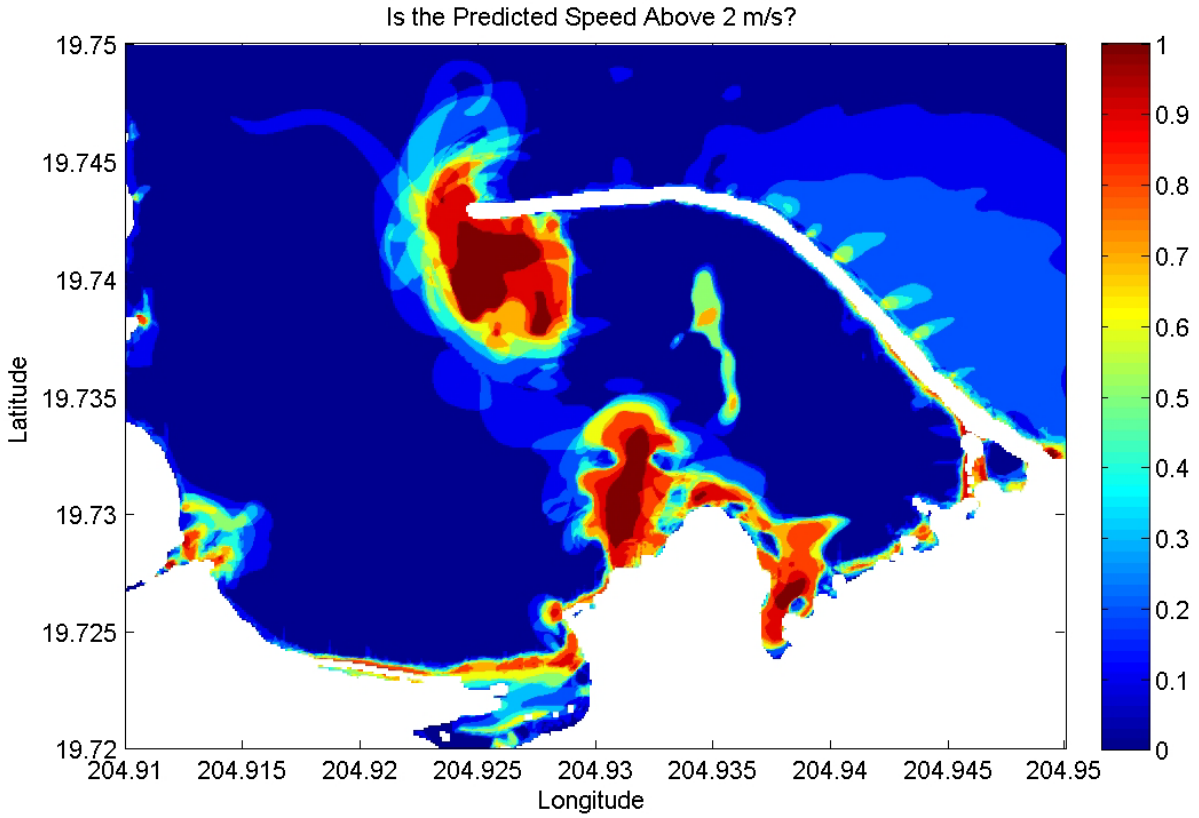


**Figure 12: Example of a current-threshold plot, where the color at each location gives the fraction of models that predicted a maximum current greater than 2 m/s at that particular location.**

## Conclusions

While the datasets used in this benchmarking activity have shortcomings, they do contain the physical processes relevant to a reasonable prediction of nearshore, tsunami-induced currents. For example, BM#1 is a relatively small-scale experiment, with Reynolds numbers below typical geophysical conditions and uniformity in bathymetry and flow conditions that would not be expected in the field. However, similar to the solitary wave runup experiments that have become a validation keystone for inundation models (i.e. Synolakis, 1987), BM#1 represents a canonical and fundamental test case that models should be able to reproduce. An ability to simulate bottom-driven spatial velocity shear and the resulting coherent structures created as flow oscillates is integral for predicting flow complexity in shallow water. The BM#1 analysis provides a range of comparisons for flow properties in the wake of an obstacle, and yields an important conclusion that models without an a-priori description of the vertical flow structure outperform other models, even those models solved with high-accuracy numerical schemes.

While the tsunami community would benefit from a re-examination of the BM#1 experiments, perhaps at a larger scale with state-of-the-art measurement techniques, the existing data represents an ideal case to compare tsunami models. There is strong de-coupling of the currents with the

mean flow, and this de-coupling is driven by bottom shear without the presence of vertical walls and shorelines, which may be numerically troublesome for depth-averaged models. What level of accuracy in BM#1 is "reasonable" would be application-specific. For example, a model used for estimating the currents in a harbor for rough planning purposes need not be held to the same standard as one used to estimate the hydrodynamic loading on a coastal bridge or critical structure. BM#1 represents a starting point to ensure that speed predictions in complex nearshore flows are reasonable. Additional accuracy tests should be added in the future, in both the form of laboratory and field datasets. Of primary importance for a field benchmark dataset is high temporal and spatial resolution, for which existing data is deficient. Additionally, a separate class of overland flow focused accuracy benchmarks must be developed. Indications from this study are that velocity predictions during inundation exhibit very large inter-model variability.

Analysis of BM#2 demonstrated the variability of field-scale currents in the presence of eddies. Model output shows that no single model can provide a converged result in the deterministic sense in eddy-affected areas, as changes in resolution lead to small changes in boundaries and resolvable shear, which then leads to large variability in the evolution of eddies and the predicted maximum currents. Likewise, the chaotic nature of eddies is physically fundamental, possibly making during-event interpretation of point measurements (e.g. ADCP) challenging. The turbulent nature of flows in separation zones and in areas with coherent eddies leads to inter-model differences that are similar in magnitude to the mean flow. The presented analysis leads to a clear argument for the use of ensemble modeling for currents in areas affected by eddies. Ensembles can be generated with multiple models or multiple realizations with a single model. While ensembles are not always available or financially practical, the current practice of using single (or a select few) realizations with a single model must at least be accompanied by careful and conservative expert interpretation and identification of areas where potential eddies could form and move in the nearshore region.

**Acknowledgements**

The workshop was funded through a grant from the National Tsunami Hazard Mitigation Program (NTHMP). The lead author's efforts in writing this paper were supported by the NTHMP, the National Science Foundation, and the California Geologic Survey. Data for the benchmarks at the workshop was provided by Dan Cox and Jose Borrero. The authors acknowledge the workshop participants Randall LeVeque, Loyce Adams, Frank Gonzalez, Bill Knight, Taro Arikawa, Ram Sampath, and Steven Prescott for their discussion during and following the workshop. In addition, we thank Randall LeVeque, Loyce Adams, and Frank Gonzalez for their contribution of numerical results to this paper.

*References*

1. Arcos, M.E.M., LeVeque, R.J. (2015). Validating Velocities in the GeoClaw Tsunami Model using Observations Near Hawaii from the 2011 Tohoku Tsunami. Pure Appl. Geophys., 172 (3–4), doi:10.1007/s00024-014-0980-y

2. Bai, Y., Cheung, K.F. (2013). Dispersion and nonlinearity of multi-layer non-hydrostatic free surface flow. J. Fluid Mech., 726, 226-260.
3. Borrero, J., Bell, R., Csato, C., DeLange, W., Greer, D., Goring, D., Pickett, V. and Power, W. (2012) Observations, Effects and Real Time Assessment of the March 11, 2011 Tohoku-oki Tsunami in New Zealand, Pure and Applied Geophysics,
4. Borrero J., Lynett P., and Kalligeris N. (2015) "Tsunami currents in ports." Proc. Royal Society of London A. A373:20140372, doi: 10.1098/rsta.2014.0372.
5. Castro, M., A. Ferreiro, J.A. García, J.M. González, J. Macías, C. Parés, M.E. Vázquez (2005). On the numerical treatment of wet/dry fronts in shallow flows: Applications to one-layer and two-layer systems. Math. Comp. Model. 42 (3-4): 419-439.
6. Cheung KF, Bai Y., Yamazaki Y. (2013) Surges around the Hawaiian Islands from the 2011 Tohoku tsunami. J Geophys Res 118:5703-5719, DOI 10.1002/jgrc.20413.
7. Chock, G. Y. (2016). Design for tsunami loads and effects in the ASCE 7-16 standard. Journal of Structural Engineering, 04016093.
8. Chock, G., Yu, G., Thio, H., and Lynett, P. (2016) "Target Structural Reliability Analysis for Tsunami Hydrodynamic Loads of the ASCE 7 Standard." Journal of Structural Engineering, doi: 10.1061/(ASCE)ST.1943-541X.0001499.
9. Dilmen, D. I., Kemec, S., Yalciner, A. C., Düzgün, S., & Zaytsev, A. (2015). Development of a tsunami inundation map in detecting tsunami risk in Gulf of Fethiye, Turkey. Pure and Applied Geophysics, 172, 921–929. DOI 10.1007/s00024-014-0936-2
10. Durran, D. R., Numerical Methods for Wave Equations in Geophysical Fluid Dynamics, Springer-Verlag, New York, Berlin. (1999).
11. Elder, J.W., 1959. The dispersion of marked fluid in turbulent shear flow. Journal of Fluid Mechanics 5, 544560
12. George, D. L. & LeVeque, R. J. Finite volume methods and adaptive refinement for global tsunami propagation and local inundation. Science of Tsunami Hazards 24, 319–328 (2006).

13. George, D. L. Augmented Riemann solvers for the shallow water equations over variable topography with steady states and inundation. J. Comput. Phys. 227, 3089–3113 (2008).
14. Gica, E., Spillane, M., Titov, V.V., Chamberlin, C. and Newman J.C., “Development of the forecast propagation database for NOAA's Short-term Inundation Forecast for Tsunamis (SIFT). NOAA Tech. Memo. OAR PMEL-139, 89 pp. (2008).
15. Grilli, S. T., O'Reilly, C., Harris, J. C., Tajalli Bakhsh, T., Tehranirad, B., Banihashemi, S., Kirby, J. T., Baxter, C. D. P., Eggeling, T., Ma, G. and Shi, F., 2015, "Modeling of SMF tsunami hazard along the upper U. S. East Coast: Detailed impact around Ocean City, MD", Nat. Hazards, 76, 705-746, doi: 10.1007/s11069-014-1522-8.
16. Horrillo, J., Grilli, S.T., Nicolsky, D., Roeber, V. and Zhang J. (2014). Performance Benchmarking Tsunami Models for NTHMP’s Inundation Mapping Activities. Pure Appl. Geophys. Published online: DOI 10.1007/s00024-014-0891-y
17. Horrillo, J., Wood A., Kim, G.-B., A. Parambath A., (2013). A simplified 3-D /Navier–Stokes numerical model for landslide tsunami: Application to the Gulf of Mexico. J. Geophys. Res. Oceans, 118, 6934–6950, doi:10.1002/2012JC008689.
18. Imamura, F. (1989). Tsunami Numerical Simulation with the staggered leap-frog scheme (Numerical code of TUNAMI-N1), School of Civil Engineering, Asian Inst. Tech. and Disaster Control Research Center, Tohoku University
19. Keen, A. S., Lynett, P. J., Eskijian, M. L., Ayca, A., & Wilson, R. (2017). Monte Carlo–Based Approach to Estimating Fragility Curves of Floating Docks for Small Craft Marinas. Journal of Waterway, Port, Coastal, and Ocean Engineering, 04017004.
20. Kim, D.-H. and Lynett, P. (2011) "Turbulent Mixing and Scalar Transport in Shallow and Wavy Flows." Physics of Fluids, v. 23 (1), doi:10.1063/1.3531716
21. Kim, D.-H., Lynett, P., and Socolofsky, S. (2009) "A Depth-Integrated Model for Weakly Dispersive, Turbulent, and Rotational Fluid Flows." Ocean Modelling, v. 27 (3-4), p. 198-214

22. LeVeque, R. J. Finite Volume Methods for Hyperbolic Problems. (Cambridge University Press, 2002).
23. Lloyd, P.M., Stansby, P.K., (1997a). Shallow water flow around model conical island of small slope. I: Submerged. Journal of Hydraulic Engineering, ASCE 123 (12), 1068-1077.
24. Lloyd, P.M., Stansby, P.K., (1997b). Shallow water flow around model conical island of small slope. II: Submerged. Journal of Hydraulic Engineering, ASCE 123 (12), 1057-1067.
25. Lynett, P. J., J. Borrero, R. Weiss, S. Son, D. Greer, and W. Renteria (2012), Observations and modeling of tsunami-induced currents in ports and harbors, Earth Planet. Sci. Lett., 327–328, 68–74.
26. Lynett, P. J., J. Borrero, S. Son, R. Wilson, and K. Miller (2014), Assessment of the tsunami-induced current hazard, Geophys. Res. Lett, 41, 2048–2055, doi:10.1002/2013GL058680.
27. Macías, J., Vázquez, J.T., Fernández-Salas, L.M., González-Vida, J.M., Bárcenas, P., Castro, M.J., Díaz-del-Río, V., Alonso, B. (2015). The Al-Boraní submarine landslide and associated tsunami. A modelling approach. Marine Geology, 361:79-95. DOI: 10.1016/j.margeo.2014.12.006.
28. Nicolsky, D. J., Suleimani, E., and Hansen, R., 2011, Pure and Applied Geophysics, Validation and verification of a numerical model for tsunami propagation and runup, 168:1199–1222, doi 10.1007/s00024-010-0231-9.
29. Okal, E. A., Fritz, H. M., Raad, P. E., Synolakis, C., Al-Shijbi, Y., & Al-Saifi, M. (2006). Oman field survey after the December 2004 Indian Ocean tsunami. Earthquake Spectra, 22(S3), 203-218.
30. Park, H., Cox, D., Lynett, P., Wiebe, D., and Shin, S. (2013) "Tsunami Inundation Modeling in Constructed Environments: A Physical and Numerical Comparison of Free-Surface Elevation, Velocity, and Momentum Flux." Coastal Engineering, v. 79, pp. 9-21, doi: 10.1016/j.coastaleng.2013.04.002.
31. Roeber, V. and Bricker, J.D. (2015). Destructive tsunami-like wave generated by surf beat over a coral reef during Typhoon Haiyan. Nature Communications 6:7854doi: 10.1038/ncomms 8854.
32. Roeber, V. and Cheung, K.F. (2012). Boussinesq-type model for energetic breaking waves in fringing reef environment. Coastal Engineering, 70, 1-20.

33. Shi, F., Kirby, J. T., Harris, J. C., Geiman, J. D. and Grilli, S. T., 2012, "A high-order adaptive time-stepping TVD solver for Boussinesq modeling of breaking waves and coastal inundation", Ocean Modelling, 43-44, 36-51.
34. Shuto, N., Goto, C., & Imamura, F. (1990). Numerical simulation as a means of warning for near field tsunamis. Coastal Engineering in Japan. 33, No: 2, 173–193.
35. Sozdinler, C. O., Yalciner, A. C., Zaytsev, A., Suppasri, A., & Imamura, F. (2015b). Investigation of hydrodynamic parameters and the effects of breakwaters during the 2011 Great East Japan Tsunami in Kamaishi Bay. Pure and Applied Geophysics, 2015 Springer Basel DOI 10.1007/s00024-015-1051-8
36. Sugawara, D., Goto, K., & Jaffe, B. E. (2014). Numerical models of tsunami sediment transport—Current understanding and future directions. Marine Geology, 352, 295-320.
37. Suleimani, E.N., Nicolsky, D.J., and Koehler, R.D., 2013, Tsunami inundation maps of Sitka, Alaska: Alaska Division of Geological & Geophysical Surveys Report of Investigation 2013-3, 76 p., 1 sheet, scale 1:250,000. doi:10.14509/26671
38. Suppasri, A., A. Muhari, T. Futami, F. Imamura, and N. Shuto. 2013. “Loss Functions for Small Marine Vessels Based on Survey Data and Numerical Simulation of the 2011 Great East Japan Tsunami.” Journal of Waterway, Port, Coastal, and Ocean Engineering 140 (5): 04014018. doi:10.1061/(ASCE)WW.1943-5460.0000244.
39. Synolakis, C. E. (1987). The runup of solitary waves. Journal of Fluid Mechanics, 185, 523-545.
40. Synolakis, C. E., Bernard, E. N., Titov, V. V., Kânoğlu, U., & González, F. I. (2008). Validation and verification of tsunami numerical models. Pure and Applied Geophysics, 165(11-12), 2197-2228.
41. Titov, V., U.Kanoglu, C. Synolakis (2016), Development of MOST for Real-Time Tsunami Forecasting, J. Waterway, Port, Coastal, Ocean Eng., DOI: 10.1061/(ASCE) WW.1943-5460.0000357
42. Tokimatsu, K., Ishida, M., & Inoue, S. (2016). Tsunami-Induced Overturning of Buildings in Onagawa during 2011 Tohoku Earthquake. Earthquake Spectra.

43. Tolkova, E., Land - Water Boundary Treatment for a Tsunami Model with Dimensional Splitting. Pure and Applied Geophysics, 171(9), 2289-2314 (2014). doi: 10.1007/s00024-014-0825-8.
44. Tolkova, E. Cliffs Benchmarking, 2016. arXiv:1601.06486 (http://arxiv.org/abs/1601.06486)
45. Wei, G., Kirby, J.T., Grilli, S.T., Subramanya, R., 1995. A fully nonlinear Boussinesq model for surface waves: Part I. Highly nonlinear unsteady waves. J. Fluid Mech. 294, 71–92.
46. Wei, Y., Bernard, E., Tang, L., Weiss, R., Titov, V., Moore, C., Spillane, M., Hopkins, M. and Kânoğlu, U., Real-time experimental forecast of the Peruvian tsunami of August 2007 for U.S. coastlines. Geophys. Res. Lett., 35, L04609, doi: 10.1029/2007GL032250 (2008) .
47. Yamazaki, Y., K. F. Cheung, and Z. Kowalik (2011), Depth-integrated, non-hydrostatic model with grid nesting for tsunami generation, propagation, and run-up, Int. J. Numer. Meth. Fluids, 67(12), 2081–2107.
48. Yanenko, N. N. “The Method of Fractional Steps.” (Translated from Russian by M. Holt), Springer, New York, Berlin, Heidelberg. (1971).
49. Yeh, H. H. J., Robertson, I., & Preuss, J. (2005). Development of design guidelines for structures that serve as tsunami vertical evacuation sites (Vol. 4). Washington State Department of Natural Resources, Division of Geology and Earth Resources
50. Zhang, Y., and Baptista, A.M. (2008) An efficient and robust tsunami model on unstructured grids. Part I: inundation benchmarks, Pure and Applied Geophysics, 165(11), 2229-2248.
51. Zhang, Y., Witter, R.C. and Priest, G.R. (2011) Tsunami–tide interaction in 1964 Prince William Sound tsunami, Ocean Modelling, 40(3–4), 246-259.
52. Zhang, Y., Ye, F., Stanev, E.V., Grashorn, S. (2016) Seamless cross-scale modeling with SCHISM, Ocean Modelling, 102, 64-81.