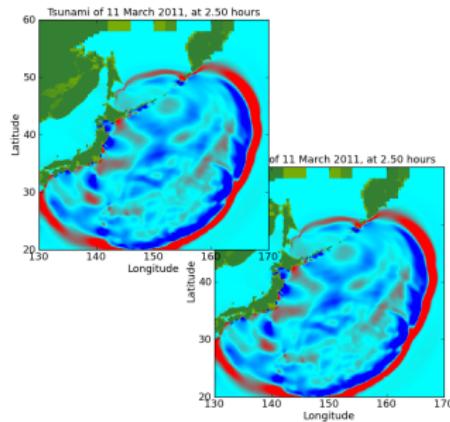


Reproducibility and Open Science

Randy LeVeque

Boeing Professor of Applied Mathematics

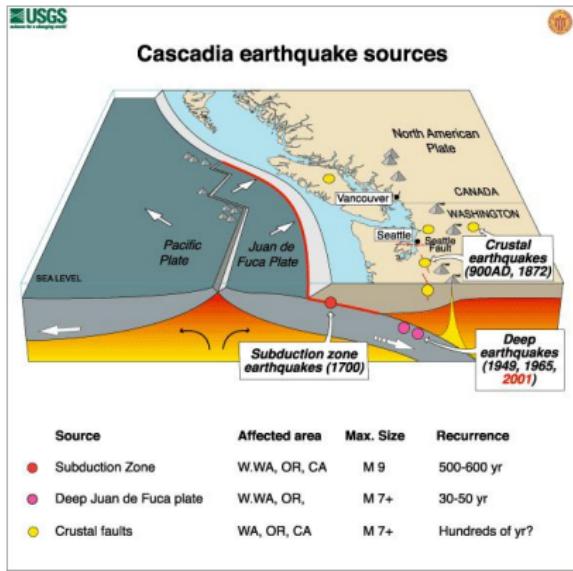
Senior Data Science Fellow, eScience Institute



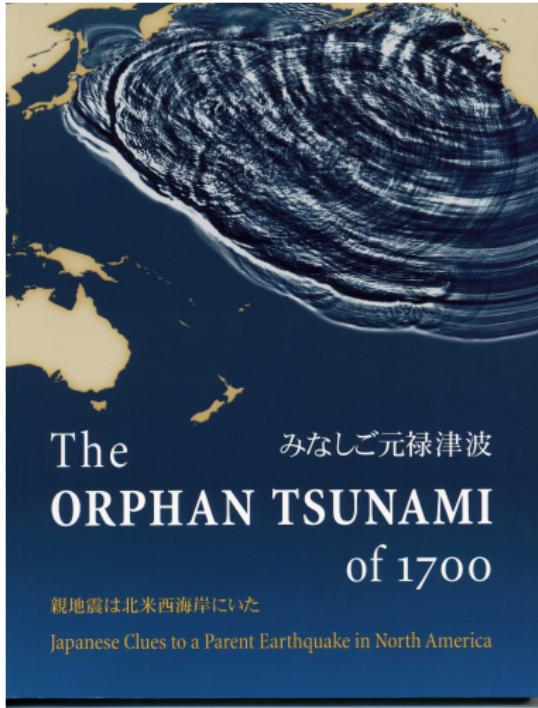
My main research interests

- Numerical methods for wave propagation,
Computational Fluid Dynamics
- Since 1994: Open source Clawpack software
(Shock waves, seismic, volcanic flows, astrophysics,
traumatic brain injury, etc.)
www.clawpack.org
[www.github.com/clawpack](https://github.com/clawpack)
- Since 2003: GeoClaw, geophysical flows and hazards
Tsunami modeling, storm surge, debris flows, etc.
Benchmarking codes,
Hazard assessment projects for Washington State,
Probabilistic Tsunami Hazard Assessment (PTHA)

Cascadia Subduction Zone (CSZ)



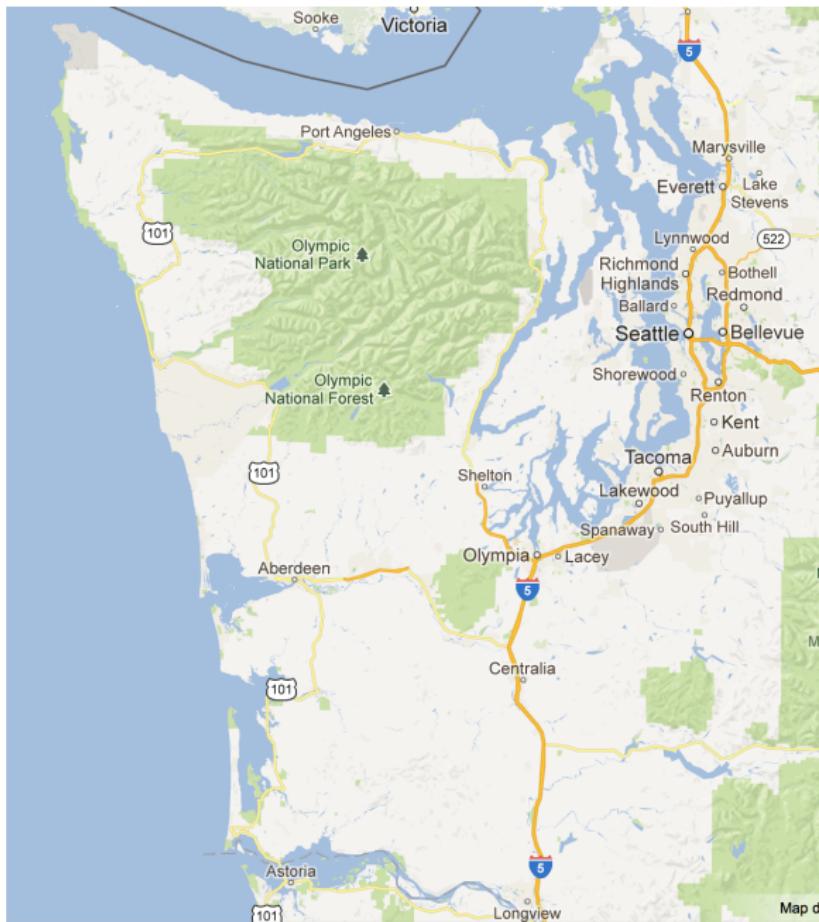
Magnitude 9 event in 1700, mean inter-event time \approx 500 years



by B. Atwater, S. Musumi-Rokkaku, K. Satake, *et. al.*, 2005

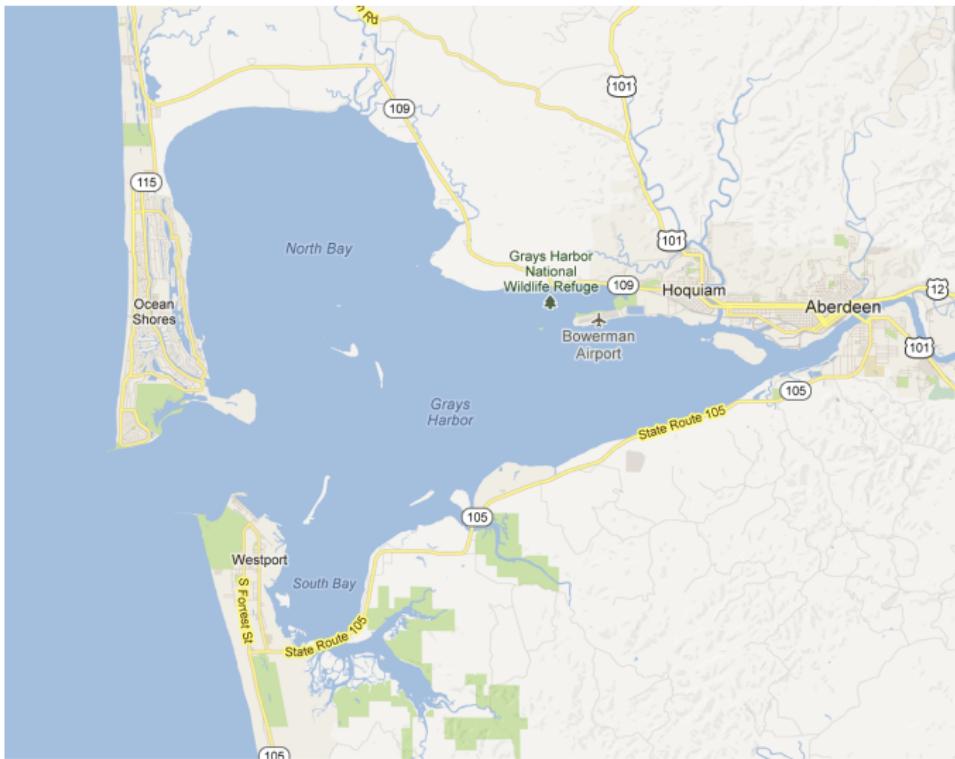
Tsunami of January 26, 1700 known from historical records in Japan,
and recently identified as originating from Cascadia Subduction Zone.

Washington Coast

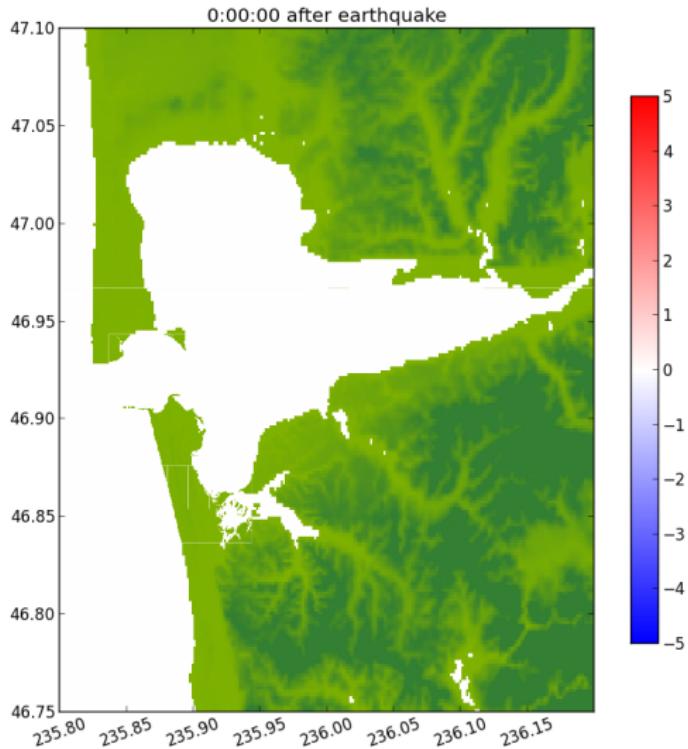


Map by

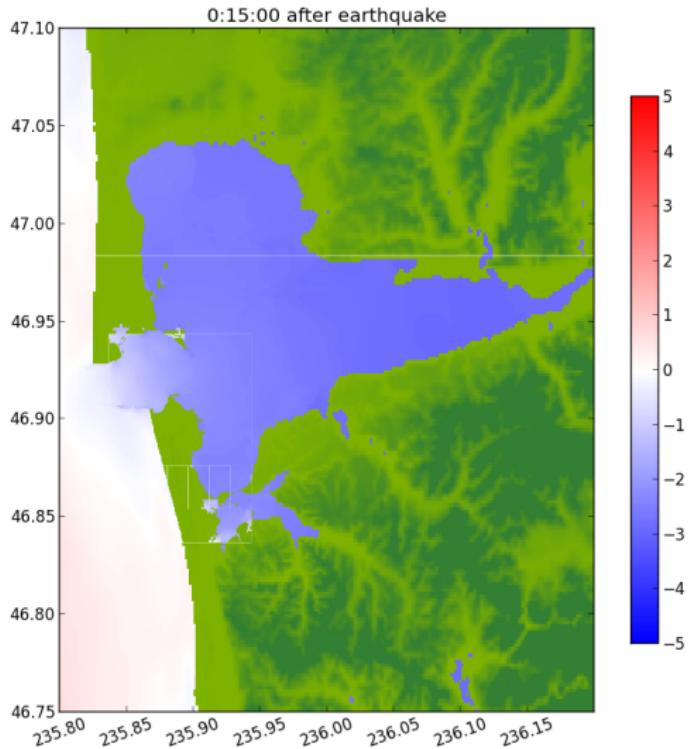
Grays Harbor



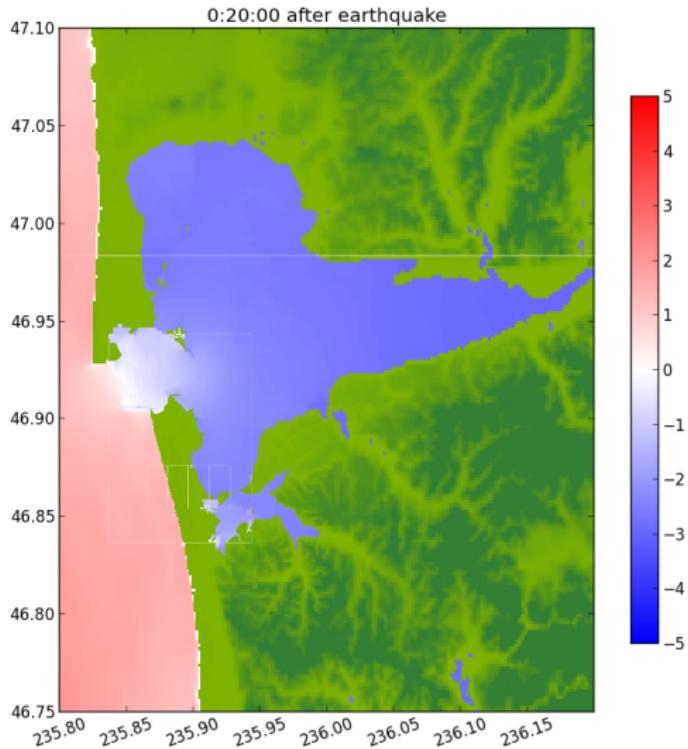
Mw 9.0 Cascadia event hitting Gray's Harbor



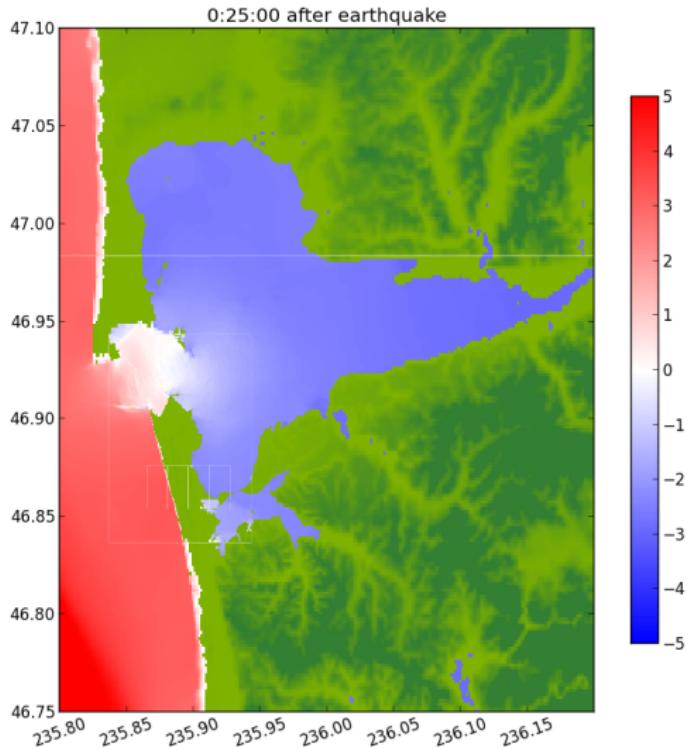
Mw 9.0 Cascadia event hitting Gray's Harbor



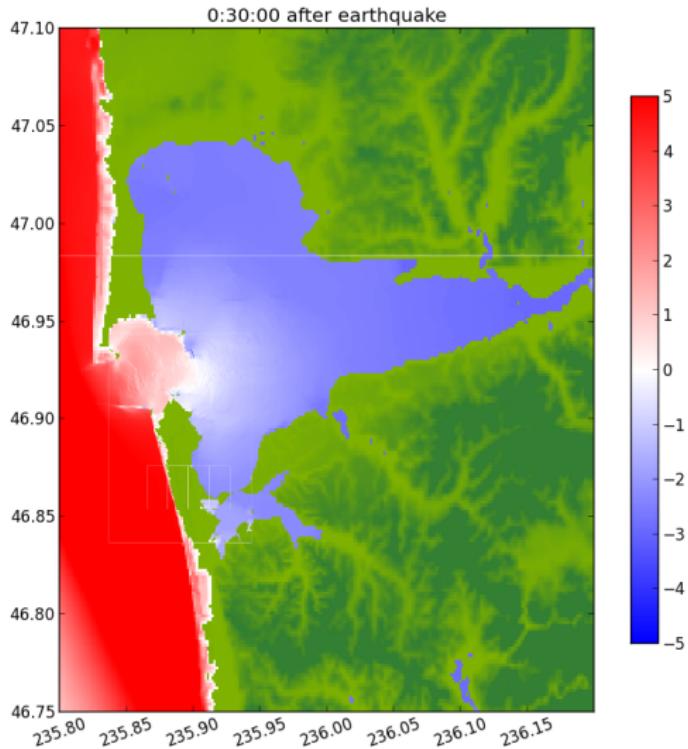
Mw 9.0 Cascadia event hitting Gray's Harbor



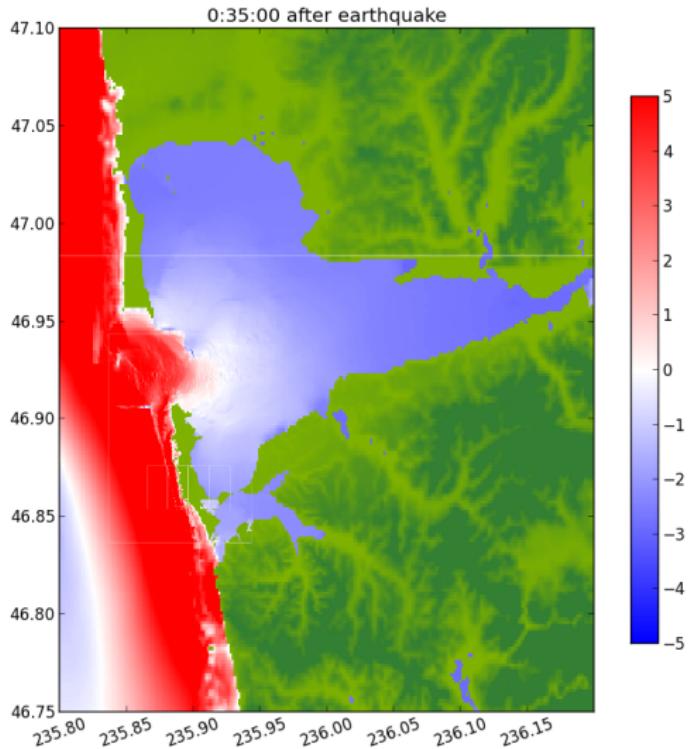
Mw 9.0 Cascadia event hitting Gray's Harbor



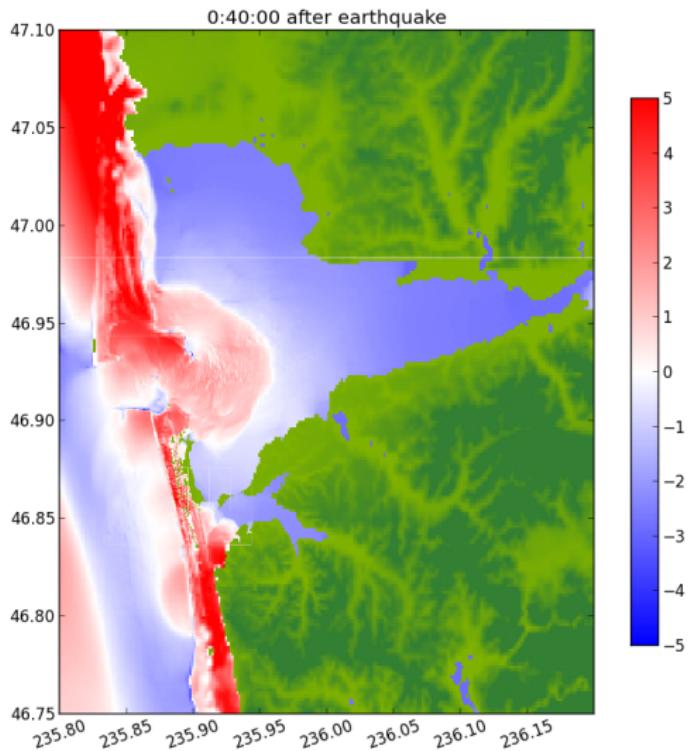
Mw 9.0 Cascadia event hitting Gray's Harbor



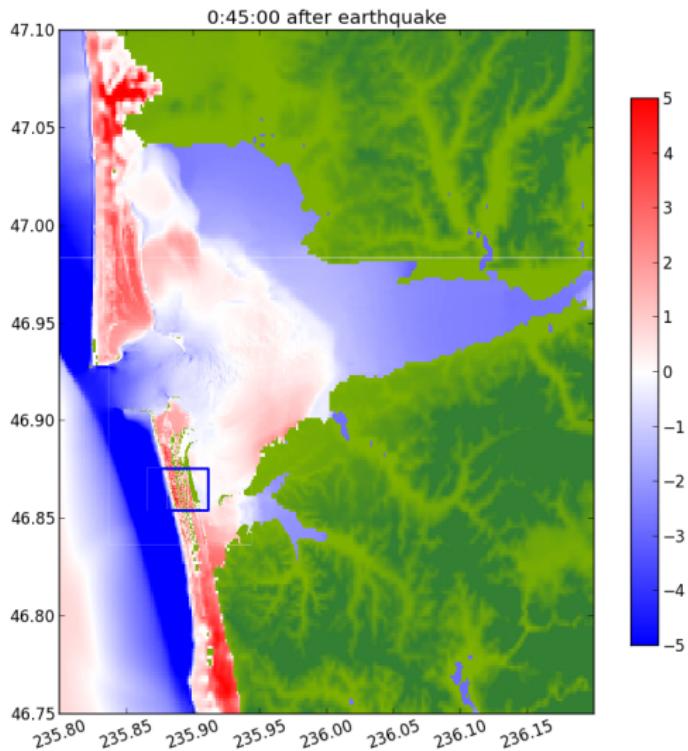
Mw 9.0 Cascadia event hitting Gray's Harbor

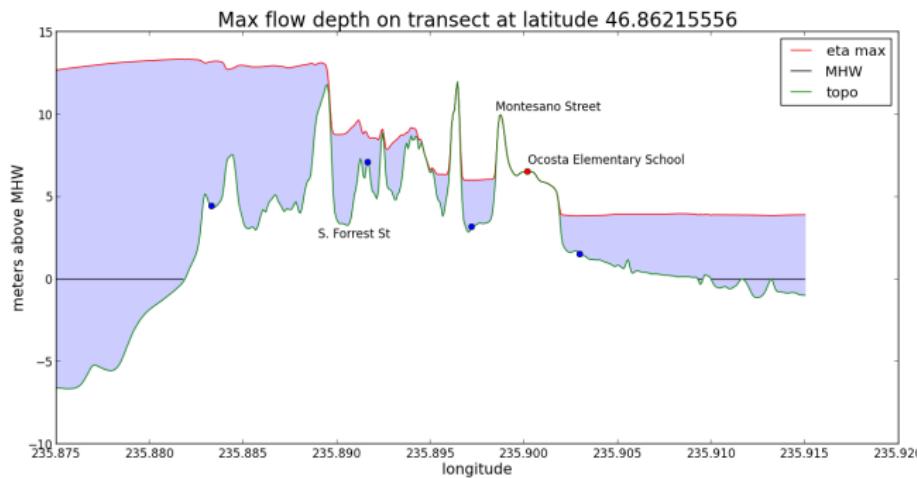
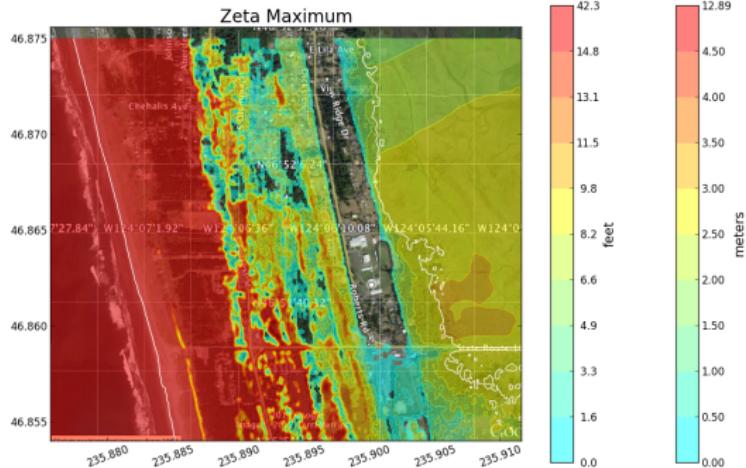


Mw 9.0 Cascadia event hitting Gray's Harbor



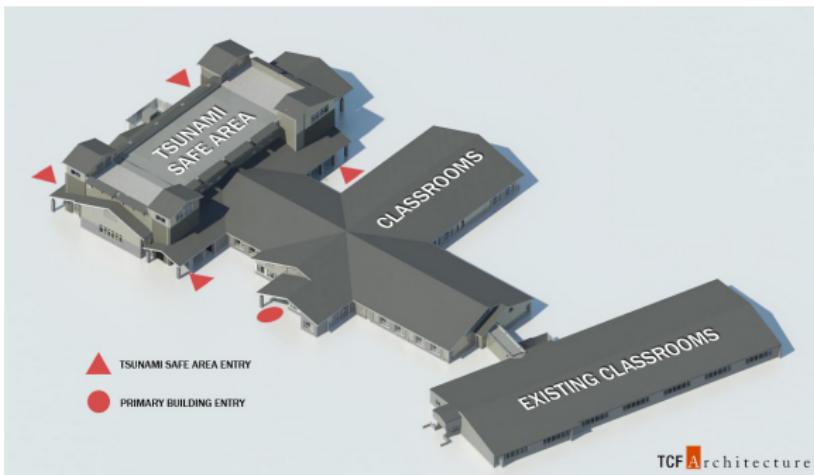
Mw 9.0 Cascadia event hitting Gray's Harbor





First vertical evacuation structure in US

Ocosta Elementary School, Westport, WA



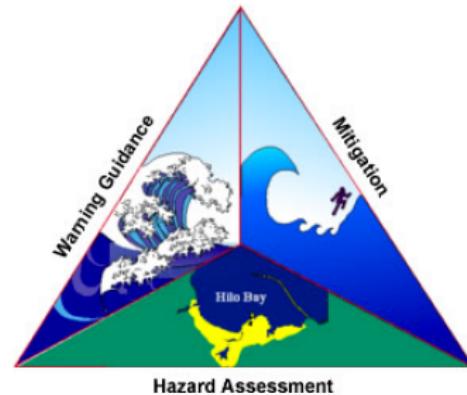
Designed by TCF Architecture of Tacoma, WA, with structural engineering work by Degenkolb Engineers.



National Tsunami Hazard Mitigation Program

NTHMP MMS Tsunami Inundation Model Validation Conference

3-28-2011 to 4-1-2011 Texas A&M Galveston campus



Benchmark data can now be found at
github.com/rjleveque/nthmp-benchmark-problems

What does Reproducible Research mean?

Ability to determine exactly how scientific results were obtained.

- Basis of scientific method.
- Required for confidently building on past results.
- Critical for accountability in engineering analysis / decision making.

What does Reproducible Research mean?

Ability to determine exactly how scientific results were obtained.

- Basis of scientific method.
- Required for confidently building on past results.
- Critical for accountability in engineering analysis / decision making.

Standards and best practices in computational/data science are not yet well codified.

What does Reproducible Research mean?

Ability to determine exactly how scientific results were obtained.

- Basis of scientific method.
- Required for confidently building on past results.
- Critical for accountability in engineering analysis / decision making.

Standards and best practices in computational/data science are not yet well codified.

- **Experimental science:** Lab notebooks, methodology section of publications, etc.

What does Reproducible Research mean?

Ability to determine exactly how scientific results were obtained.

- Basis of scientific method.
- Required for confidently building on past results.
- Critical for accountability in engineering analysis / decision making.

Standards and best practices in computational/data science are not yet well codified.

- **Experimental science:** Lab notebooks, methodology section of publications, etc.
- **Mathematics:** Proofs are required in publications.

Quote from [Reproducible Research: A Cautionary Tale](#)

By David Crotty, March 26, 2014 on [the scholarly kitchen blog](#)

If your experiment consists of running numerical data through an algorithm, then releasing your data and your code allows others to quickly verify that you've done what you've said you've done. But when it comes to other types of research, wet bench experiments or observational work for example, reproduction is not quite so simple.

Quote from [Reproducible Research: A Cautionary Tale](#)

By David Crotty, March 26, 2014 on [the scholarly kitchen blog](#)

If your experiment consists of running numerical data through an algorithm, then releasing your data and your code allows others to quickly verify that you've done what you've said you've done. But when it comes to other types of research, wet bench experiments or observational work for example, reproduction is not quite so simple.

If only it were so easy in computational/data science!

Private reproducibility...

- Use scripts, not GUIs, for data analysis and visualization.
- Use version control / provenance tracking tools.
- Archive code and data used for published results.

Private reproducibility...

Use scripts, not GUIs, for data analysis and visualization.

Use version control / provenance tracking tools.

Archive code and data used for published results.

Why?

- Ability to check results in prior publication,
- Ability to build on your own past research of your own (or students / collaborators).
- Easily modify tables/figures to satisfy referees, etc.

Private reproducibility...

Use scripts, not GUIs, for data analysis and visualization.

Use version control / provenance tracking tools.

Archive code and data used for published results.

Why?

- Ability to check results in prior publication,
- Ability to build on your own past research of your own (or students / collaborators).
- Easily modify tables/figures to satisfy referees, etc.

Auditable Research: Even if code and data are not shared, there should be a permanent record that can be checked.

Analogous to lab notebooks.

Public Reproducibility...

Allowing others to reproduce your results.

(Readers, referees, researchers down the hall...)

Public Reproducibility...

Allowing others to reproduce your results.

(Readers, referees, researchers down the hall...)

Why?

- Verifying scientific integrity of results.
- Aids in understanding ideas, implementing methods
- Increases impact of work.

Compare to Mathematics

Traditional research in Mathematics is reproducible...

A paper containing a new theorem cannot be published without the proof.

Compare to Mathematics

Traditional research in Mathematics is reproducible...

A paper containing a new theorem cannot be published without the proof.

It wasn't always so...

There is no . . . mathematician so expert in his science, as to place entire confidence in any truth immediately upon his discovery of it. . . . Every time he runs over his proofs, his confidence increases; but still more by the approbation of his friends; and is raised to its utmost perfection by the universal assent and applause of the learned world.

— David Hume, 1739

Compare to Mathematics

Many arguments against publishing code might be applied to proofs in an alternate universe...

“Top Ten Reasons To Not Share Your Code (and why you should anyway)”, **SIAM News, April, 2013**

- The proof is too ugly to show anyone else.
- I didn’t work out all the details.
- I didn’t actually prove the theorem—my student did.
- Giving the proof to my competitors would be unfair to me.
- The proof is valuable intellectual property.
- Etc.

What does Reproducible Research mean?

What does it mean that others can reproduce your results?

What does Reproducible Research mean?

What does it mean that others can reproduce your results?

Possible answers...

- Download the code and type `make plots`,
see identical plots appear.

What does Reproducible Research mean?

What does it mean that others can reproduce your results?

Possible answers...

- Download the code and type `make plots`, see identical plots appear.
- Be able to implement the algorithm from description in paper and other archived sources, and get essentially the same results.

What does Reproducible Research mean?

What does it mean that others can reproduce your results?

Possible answers...

- Download the code and type `make plots`, see identical plots appear.
- Be able to implement the algorithm from description in paper and other archived sources, and get essentially the same results.
- Various things in between.

What does Reproducible Research mean?

What does it mean that others can reproduce your results?

Possible answers...

- Download the code and type `make plots`, see identical plots appear.
- Be able to implement the algorithm from description in paper and other archived sources, and get essentially the same results.
- Various things in between.

Terms such as **replicable** or **repeatable** are sometimes used in addition to **reproducible**.

Tools to facilitate reproducibility

- Version control systems (VCS)
CVS, Subversion (server-client),
Git, Mercurial, Bazaar, etc. (distributed).

Tools to facilitate reproducibility

- Version control systems (VCS)
CVS, Subversion (server-client),
Git, Mercurial, Bazaar, etc. (distributed).
- Public hosting sites for VCS repositories
Github, Bitbucket, Google code, sourceforge, etc.
Collaboration on open source projects,
Archiving code used for publications.

Tools to facilitate reproducibility

- Version control systems (VCS)
CVS, Subversion (server-client),
Git, Mercurial, Bazaar, etc. (distributed).
- Public hosting sites for VCS repositories
Github, Bitbucket, Google code, sourceforge, etc.

Collaboration on open source projects,
Archiving code used for publications.
- Other archives with stable URLs, DOIs
Institutional or public data repositories,
journal supplementary materials,
Figshare, Zenodo, etc.

Tools to facilitate reproducibility

- Literate Programming tools
CWEB, Doxygen, Sphinx, Sweave, etc.

Tools to facilitate reproducibility

- Literate Programming tools

CWEB, Doxygen, Sphinx, Sweave, etc.

- Notebooks / Publishing tools

Mathematica, Maple, Matlab,

Jupyter, Sage, knitr, RStudio, etc.

Tools to facilitate reproducibility

- Virtualization

Package code along with complete environment
(OS, compilers, graphics tools, etc.)

VirtualBox, VMware, Docker, etc.

Tools to facilitate reproducibility

- **Virtualization**

Package code along with complete environment
(OS, compilers, graphics tools, etc.)

VirtualBox, VMware, Docker, etc.

- **Cloud computing**

Amazon EC2, Windows Azure, Google Cloud, etc.

Tools to facilitate reproducibility

- **Virtualization**

Package code along with complete environment
(OS, compilers, graphics tools, etc.)

VirtualBox, VMware, Docker, etc.

- **Cloud computing**

Amazon EC2, Windows Azure, Google Cloud, etc.

- **Web platforms for running code**

RunMyCode.org, wakari.io,

SageMathCloud: cloud.sagemath.com

binder: mybinder.org

Some examples from my work...

Jupyter notebooks + Github + binder:

https://github.com/rjleveque/ptha_tutorial

Publication + preprint + code + notebooks,
using arXiv, Zenodo, Github

faculty.washington.edu/rjl/pubs/KLslip/index.html

Some links ...

- **2012 ICERM Workshop on Reproducibility**
Many links on the [wiki](#) and in [Final report](#)
- **2011 UBC Workshop with videos online**
- reproducibleresearch.net
- **Tutorial: Workflows for reproducible research in computational neuroscience** by Andrew Davison
- **10 Simple Rules for the Care and Feeding of Scientific Data**
by A. Goodman, A. Pepe, A. W. Blocker, et al.
- **Best Practices for Scientific Computing**
by G. Wilson, D. A. Aruliah, C. T. Brown, et al.