Collaboratory for the Study of Earthquake Predictability (CSEP)

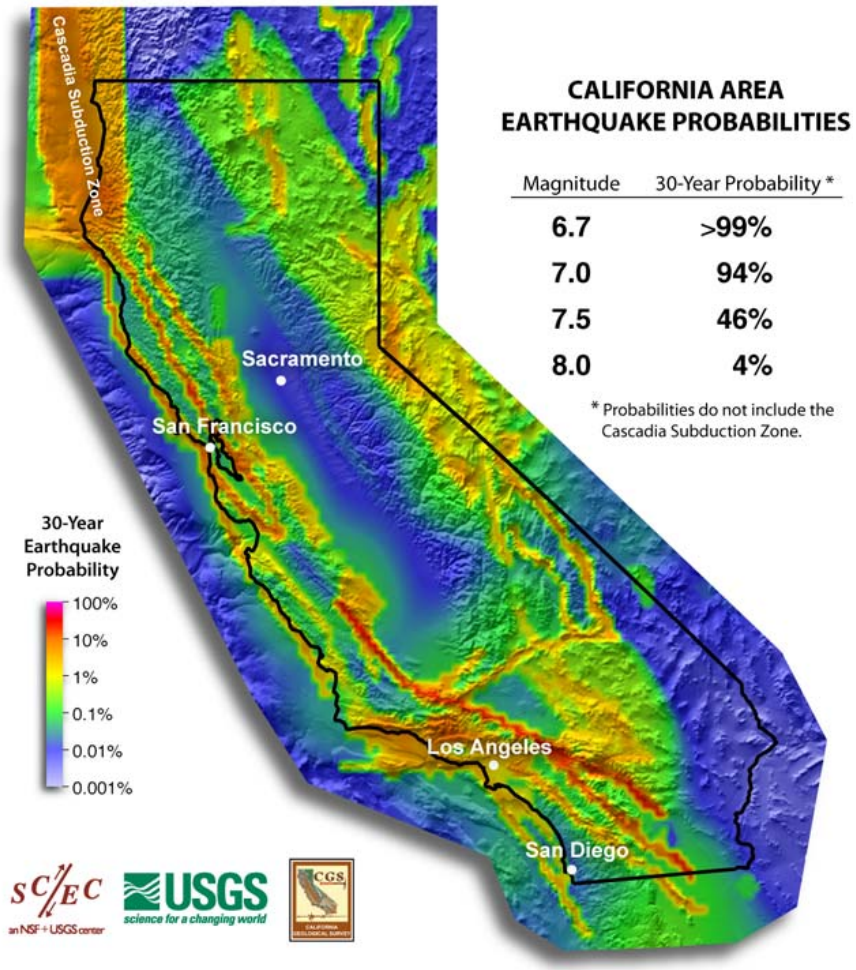
D. Schorlemmer,[1] J. Zechar,[2] M. Gerstenberger,[3]
N. Hirata,[4] K. Nanjo,[4] T. H. Jordan,[1]
and the CSEP Working Group

1 - Southern California Earthquake Center, University of Southern California, Los Angeles, USA
2 - Lamont Doherty Earth Observatory, Columbia University, Palisades, USA
3 - GNS Science, Lower Hutt, New Zealand
4 - Earthquake Research Institute, University of Tokyo, Tokyo, Japan

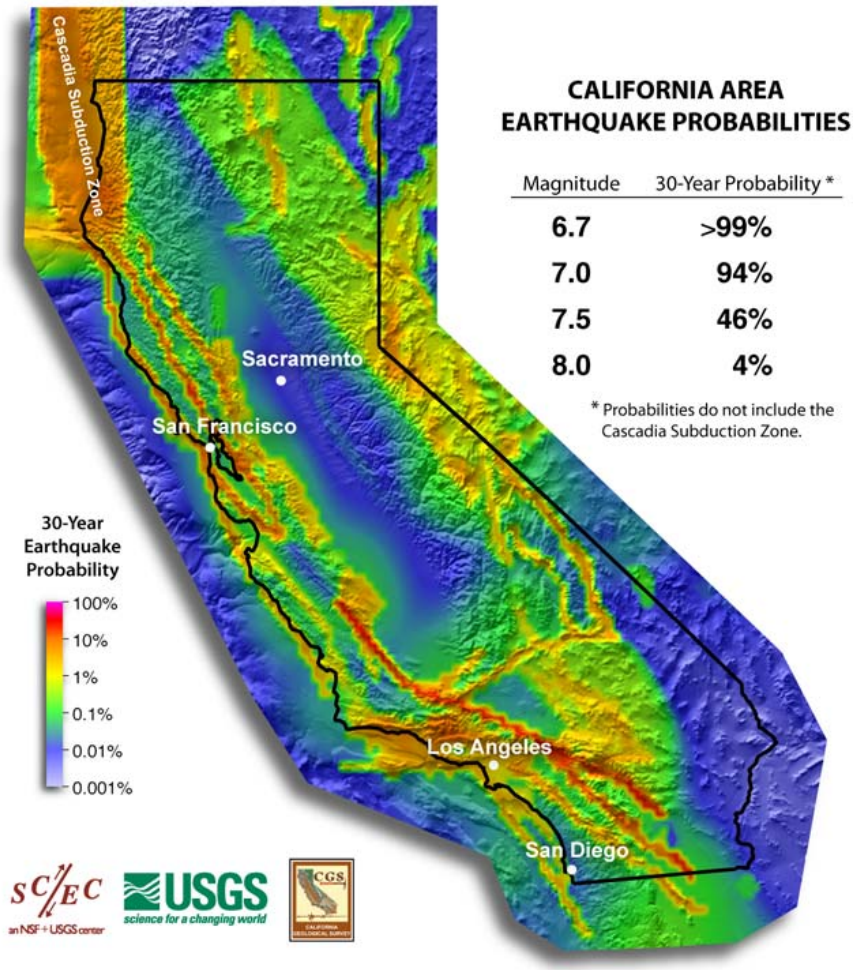
The fundamental principle of science, the definition almost, is this: the sole test of the validity of any idea is experiment.

Richard P. Feynman

UCERF2


Cascadia Subduction Zone
Sacramento
San Francisco
Los Angeles
San Diego
CALIFORNIA AREA
EARTHQUAKE PROBABILITIES
Magnitude
30-Year Probability *
6.7
>99%
7.0
94%
7.5
46%
8.0
4%
* Probabilities do not include the Cascadia Subduction Zone.
30-Year
Earthquake
Probability
100%
10%
1%
0.1%
0.01%
0.001%
SCEC
an NSF+USGS center
USGS
science for a changing world
CGS

UCERF2



This isn't right. This isn't even wrong.

Wolfgang Pauli

Problems in Assessing Predictions

Even in the case where a hypothesis is stated:

- Scientific publications provide insufficient information for independent evaluation
- Data to evaluate prediction experiments are often improperly specified
- Active researchers are constantly tweaking their procedures, which become moving targets
- Difficult to find resources to conduct and evaluate long-term prediction experiments
- Standards are lacking for testing predictions against reference forecasts

What is CSEP?

CSEP promotes rigorous research on earthquake predictability through:

- an open and international collaboratory infrastructure
- rigorous and prospective testing of earthquake forecast models and scientific hypotheses
- a global program in a variety of tectonic environments

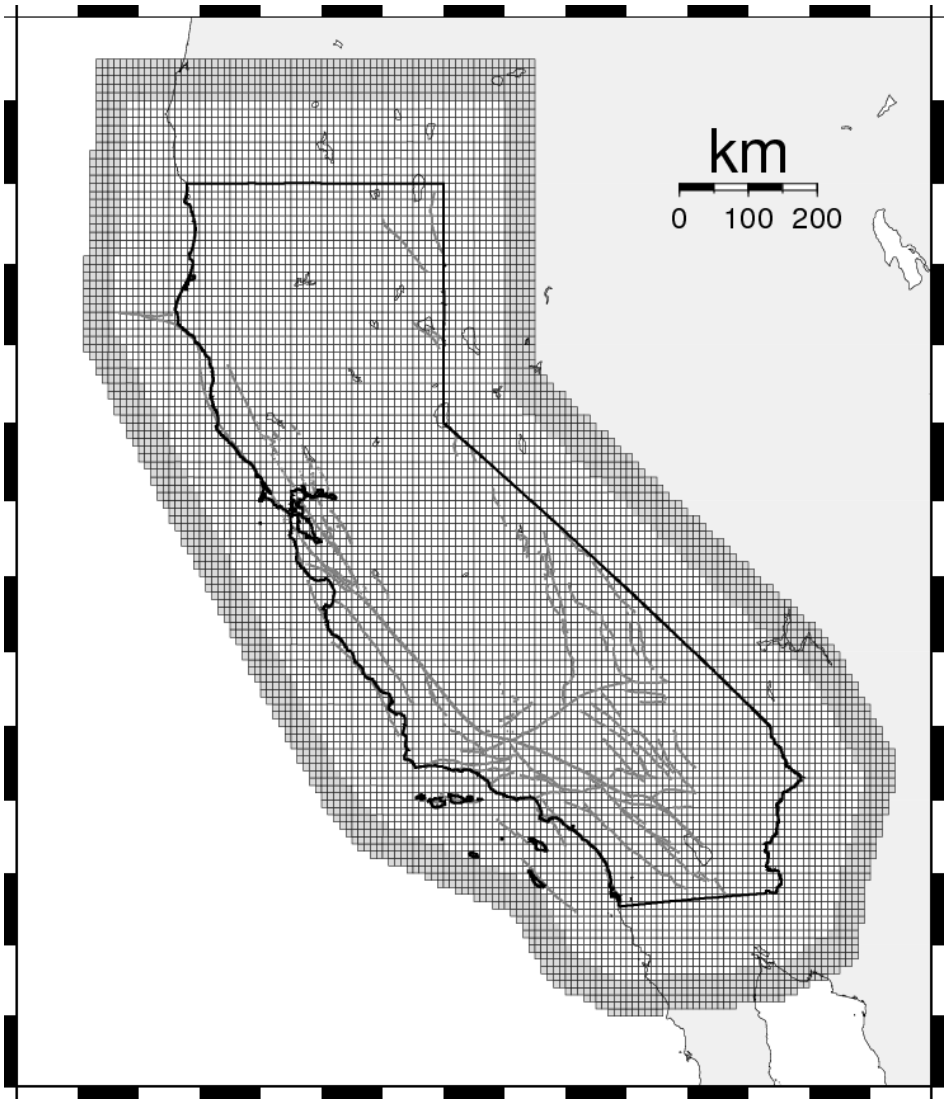
Design of an Experiment

What definitions need to be made for a scientific experiment?

Design of an Experiment

What definitions need to be made for a scientific experiment?

- Testing area



Design of an Experiment

What definitions need to be made for a scientific experiment?

- Testing area
- Exact description of forecast

	5-year	1-year	1-day
Forecast duration	5 years	1 year	1 day
Aftershocks	yes/no	yes/no	yes
Magnitude range	5-9	5-9	4-9
Modeler provides	numbers	code	code

Design of an Experiment

What definitions need to be made for a scientific experiment?

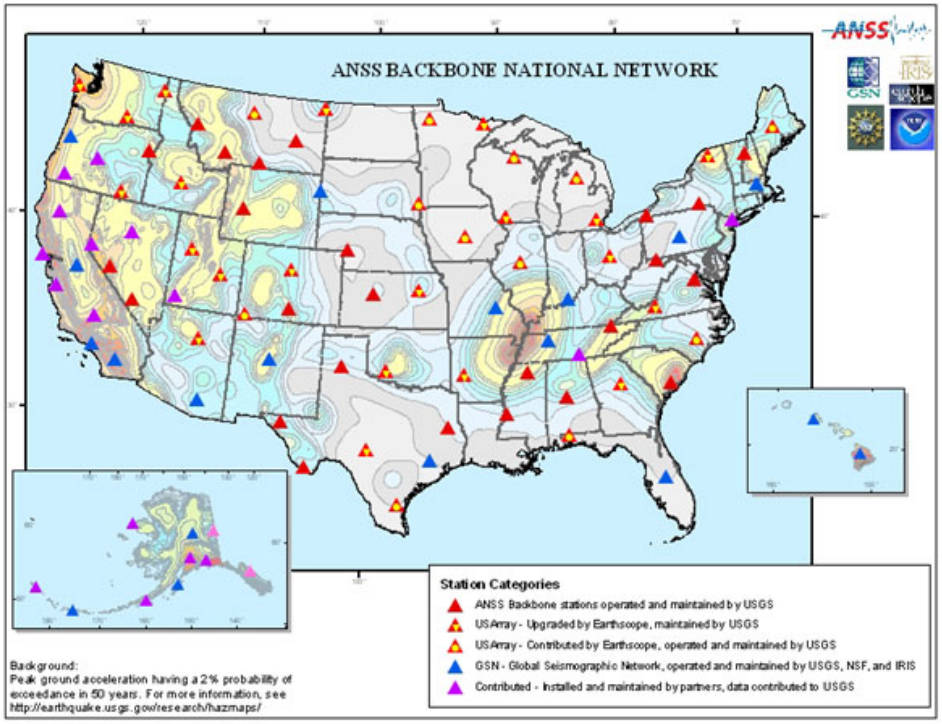
- Testing area
- Exact description of forecast

```
<?xml version='1.0' encoding='UTF-8'?>
<CSEPForecast xmlns='http://www.scec.org/xml-ns/csep/forecast/0.1'>
  <forecastData publicID='smi:org.scec/csep/forecast/1'>
    <modelName>unknown</modelName>
    <version>1.0</version>
    <author>CSEP</author>
    <issueDate>2005-06-18T10:30:00Z</issueDate>
    <forecastStartDate>2008-01-01T00:00:00Z</forecastStartDate>
    <forecastEndDate>2013-01-01T00:00:00Z</forecastEndDate>
    <defaultMagBinDimension>0.1</defaultMagBinDimension>
    <lastMagBinOpen>1</lastMagBinOpen>
    <defaultCellDimension latRange='0.1' lonRange='0.1'/>
    <depthLayer max='30.0' min='0.0'>
      <cell lat='44.95' lon='5.55'>
        <bin m='4.0'>0.0</bin>
        ...
        <bin m='5.0'>0.0</bin>
        <bin m='6.0'>0.0</bin>
        <bin m='7.0'>0.0</bin>
        <bin m='8.0'>0.0</bin>
        <bin m='9.0'>0.0</bin>
      </cell>
      <cell lat='45.05' lon='5.55'>
        ...
      </cell>
      ...
    </depthLayer>
  </forecastData>
</CSEPForecast>
```

Design of an Experiment

What definitions need to be made for a scientific experiment?

- Testing area
- Exact description of forecast
- Exact definition of input data (authorized & calibrated)



Design of an Experiment

What definitions need to be made for a scientific experiment?

- Testing area
- Exact description of forecast
- Exact definition of input data (authorized & calibrated)
- Measure of success
  - N-Test (Consistency)
  - L-Test (Consistency)
  - R-Test (Comparison)

Earthquake Likelihood Model Testing

D. Schorlemmer,[1,2] M. C. Gerstenberger,[3] S. Wiemer,[1] D. D. Jackson,[4] and D. A. Rhoades[5]

INTRODUCTION

The Regional Earthquake Likelihood Models (RELM) project aims to produce and evaluate alternate models of earthquake potential (probability per unit volume, magnitude, and time) for California. Based on differing assumptions, these models are produced to test the validity of their assumptions and to explore which models should be incorporated in seismic hazard and risk evaluation. Tests based on physical and geological criteria are useful but we focus on statistical methods using future earthquake catalog data only. We envision two evaluations: a test of consistency with observed data and a comparison of all pairs of models for relative consistency. Both tests are based on the likelihood method, and both are fully prospective (*i.e.*, the models are not adjusted to fit the test data). To be tested, each model must assign a probability to any possible event within a specified region of space, time, and magnitude. For our tests the models must use a common format: earthquake rates in specified "bins" with location, magnitude, time, and focal mechanism limits.

Seismology cannot yet deterministically predict individual earthquakes; however, it should seek the best possible models for forecasting earthquake occurrence. This paper describes the statistical rules of an experiment to examine and test earthquake forecasts. The primary purposes of the tests described below are to evaluate physical models for earthquakes, assure that source models used in seismic hazard and risk studies are consistent with earthquake data, and provide quantitative measures by which models can be assigned weights in a consensus model or be judged as suitable for particular regions.

In this paper we develop a statistical method for testing earthquake likelihood models. A companion paper (Schorlemmer and Gerstenberger 2007, this issue) discusses the actual implementation of these tests in the framework of the RELM initiative.

Statistical testing of hypotheses is a common task and a wide range of possible testing procedures exist. Jolliffe and Stephenson (2003) present different forecast verifications from atmospheric science, among them likelihood testing of probability forecasts and testing the occurrence of binary events. Testing binary events requires that for each forecasted event, the spatial, temporal and magnitude limits be given. Although major earthquakes can be considered binary events, the models within the RELM project express their forecasts on a spatial grid and in 0.1 magnitude units; thus the models are a distribution of rates over space and magnitude. These forecasts can be tested with likelihood tests.

In general, likelihood tests assume a valid null hypothesis against which a given hypothesis is tested. The outcome is either a rejection of the null hypothesis in favor of the test hypothesis or a nonrejection, meaning the test hypothesis cannot outperform the null hypothesis at a given significance level. Within RELM, there is no accepted null hypothesis and thus the likelihood test needs to be expanded to allow comparable testing of equipollent hypotheses.

To test models against one another, we require that forecasts are expressed in a standard format: the average rate of earthquake occurrence within pre-specified limits of hypocentral latitude, longitude, depth, magnitude, time period, and focal mechanisms. Focal mechanisms should either be described as the inclination of *P*-axis, declination of *P*-axis, and inclination of the *T*-axis, or as strike, dip, and rake angles. Schorlemmer and Gerstenberger (2007, this issue) designed classes of these parameters such that similar models will be tested against each other. These classes make the forecasts comparable between models. Additionally, we are limited to testing only what is precisely defined and consistently reported in earthquake catalogs. Therefore it is currently not possible to test such information as fault rupture length or area, asperity location, etc. Also, to account for data-quality issues, we allow for location and magnitude uncertainties as well as the probability that an event is dependent on another event.

As we mentioned above, only models with comparable forecasts can be tested against each other. Our current tests are designed to examine grid-based models. This requires that any fault-based model be adapted to a grid before testing is possible. While this is a limitation of the testing, it is an inherent difficulty in any such comparative testing. Please refer to appendix B for a statistical evaluation of the application of the Poisson hypothesis to fault-based models.

The testing suite we present consists of three different tests: L-Test, N-Test, and R-Test. These tests are defined similarly to

1. ETH, Zürich, Switzerland
2. Now at University of Southern California, Department of Earth Sciences.
3. U.S. Geological Survey, Pasadena, California, now at GNS Science.
4. University of California—Los Angeles Department Earth and Space Sciences.
5. GNS Science, Lower Hutt, New Zealand.

Design of an Experiment

What definitions need to be made for a scientific experiment?

- Testing area
- Exact description of forecast
- Exact definition of input data (authorized & calibrated)
- Measure of success
- Truly prospective (Zero degrees of freedom)

Design of an Experiment

What definitions need to be made for a scientific experiment?

- Testing area
- Exact description of forecast
- Exact definition of input data (authorized & calibrated)
- Measure of success
- Truly prospective (Zero degrees of freedom)

Goals:

- Reproducibility
- Transparency
- Controlled Environment
- Comparability

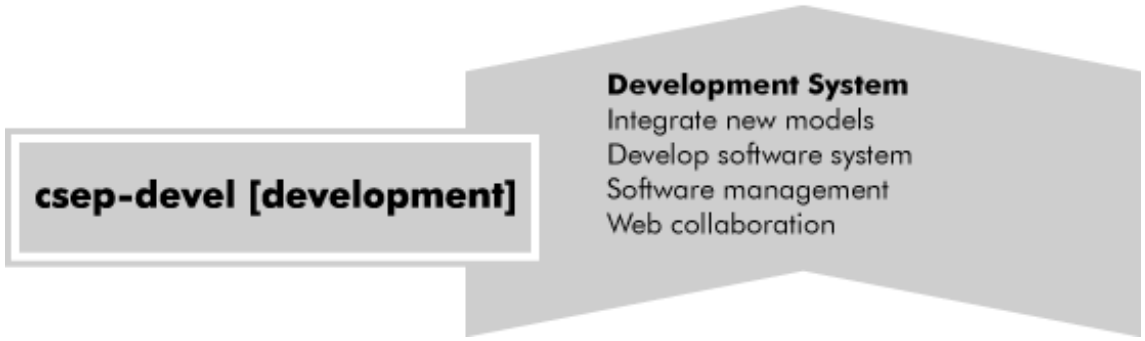
We Need a Testing Center

- Automated processing of earthquake forecast tests
- Storing of all input data, forecast data, and results
- Documenting each models code (and changes)
- Recomputation abilities (alternative options, bugs)
- Truly prospective tests with time delay
- Processing independent of modelers
- Long-term testing
- “Certify” all steps of the testing process

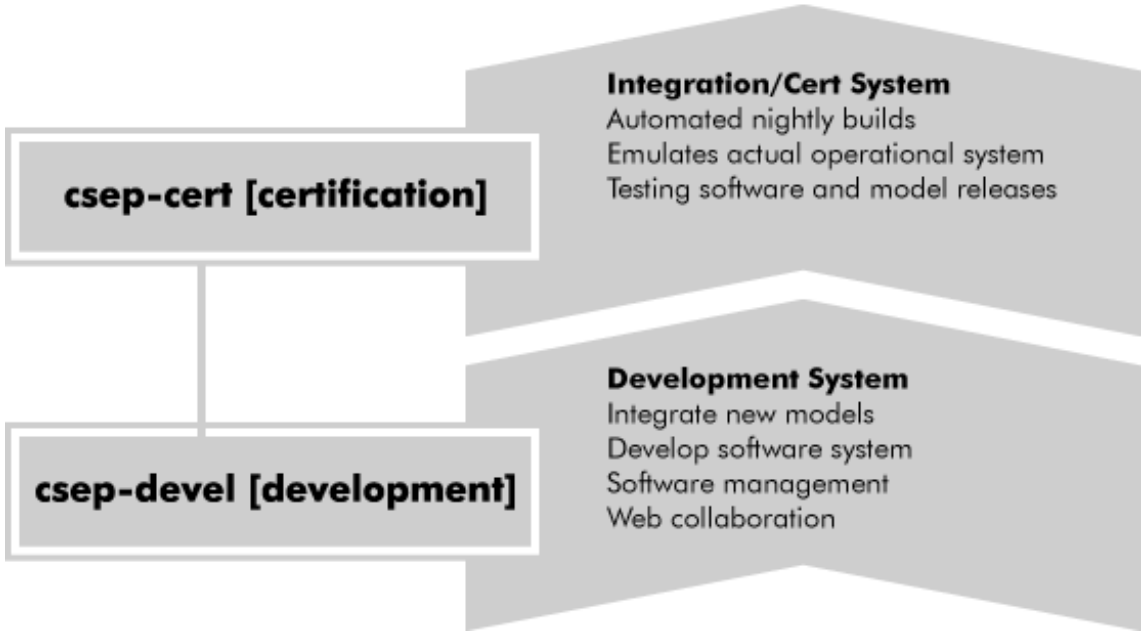
The CSEP Testing Centers

- Multi-tier computer system
- Test-driven software development (Put the system under test)
- Rigorous testing area definitions
- Community standards: rules for the registration and evaluation of scientific prediction experiments
- Communication protocols: procedures for conveying scientific results and their significance to:
  - The scientific community
  - Government agencies
  - The general public

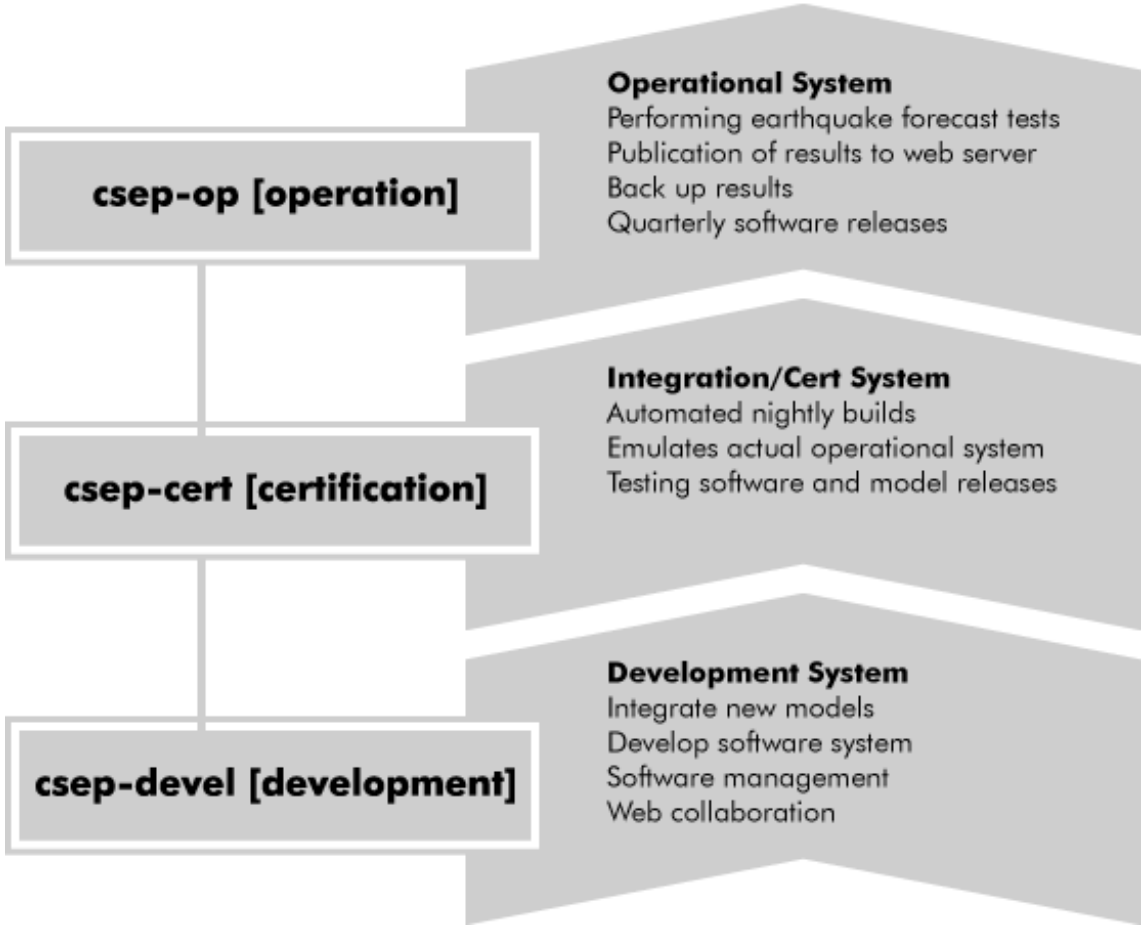
CSEP Computer System



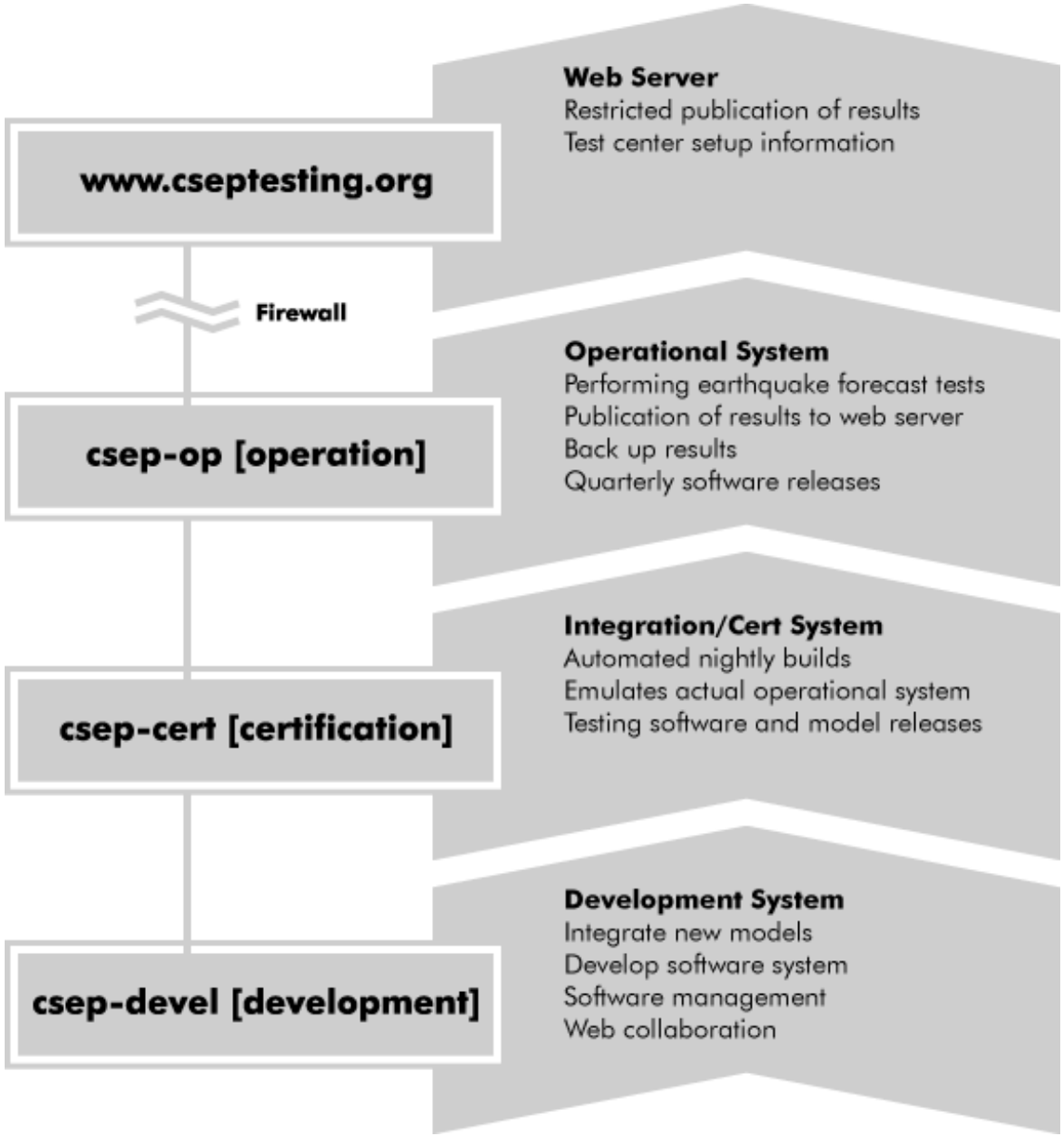
CSEP Computer System



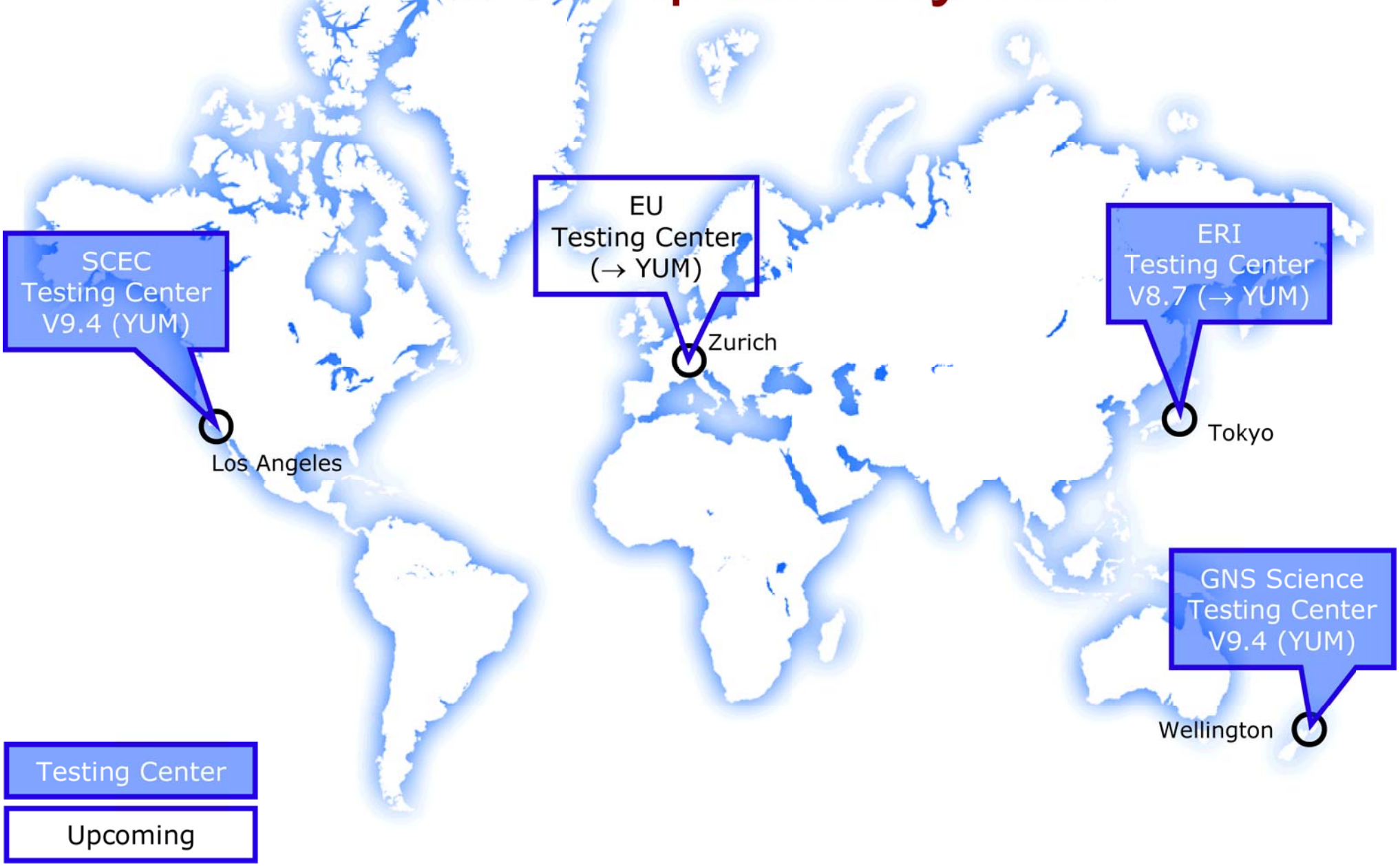
CSEP Computer System



CSEP Computer System

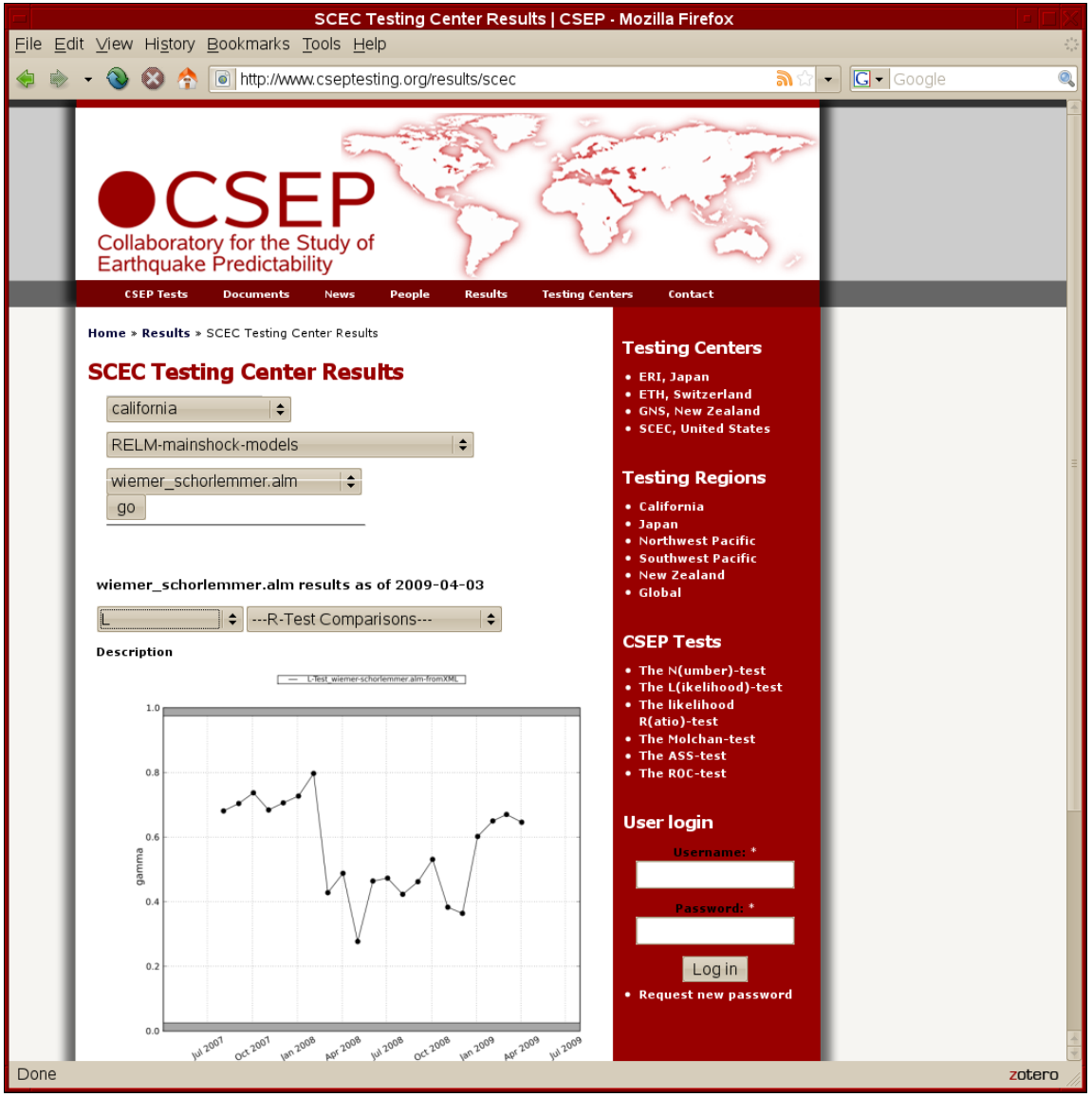


CSEP Computer System



Communication Protocols

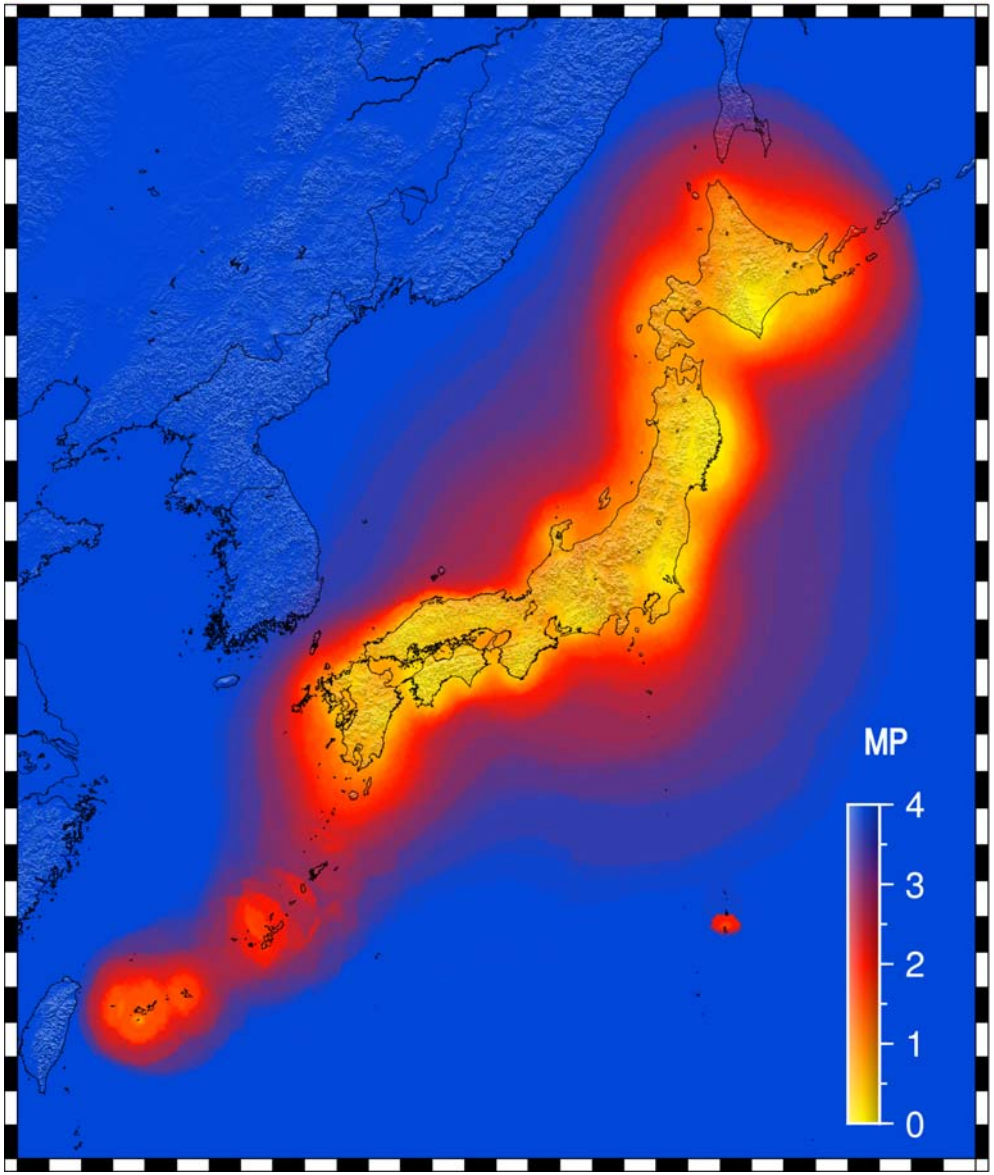
- Result webpages
- News Releases
- Mailing lists
- Weekly minutes posted on website



Testing Regions

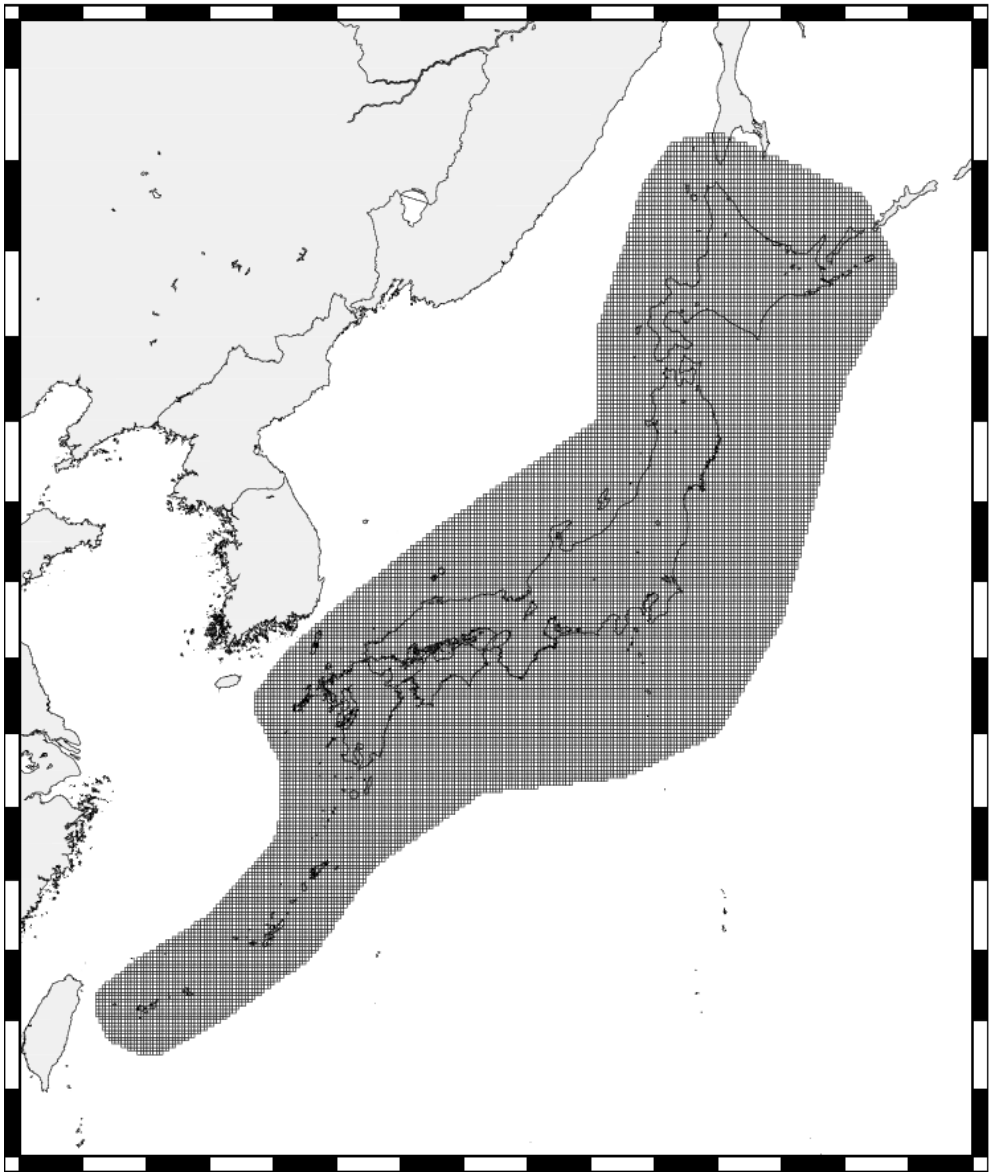
- Characterized and calibrated regions
- Need a local sponsor/agency
- Provide a low-latency earthquake catalog
- Earthquake catalog is authorized as independent data source

CSEP Testing Region Japan



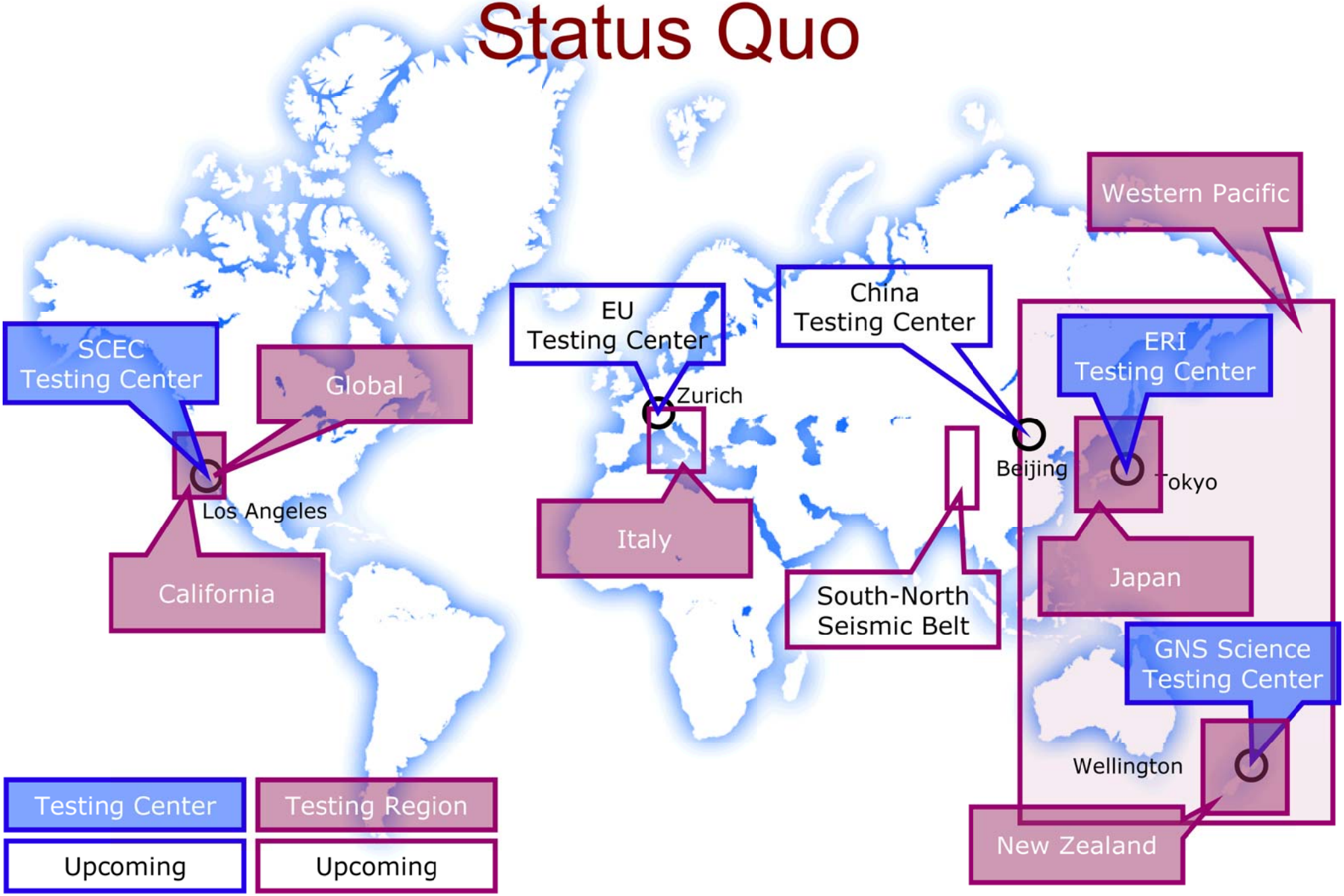
1.1.2008
Depth=30km
P=0.999

CSEP Testing Region Japan



Complete for M ≥ 3.7

Status Quo


Western Pacific
China
Testing Center
EU
Testing Center
SCEC
Testing Center
Global
ERI
Testing Center
Zurich
Beijing
Tokyo
Los Angeles
Italy
Japan
California
South-North
Seismic Belt
GNS Science
Testing Center
Wellington
New Zealand
Testing Center
Testing Region
Upcoming
Upcoming

Models Under Test at SCEC

California: RELM Models

Bird & Liu
SHIFT main shock + aftershock model

Ebel et al.
5-yr main shock+aftershock model
5-yr main shock model

Helmstetter, Kagan, Jackson
HKJ 2005 long-term main shock model
HKJ 2005 long-term main shock + aftershock model

Holliday et al.
Pattern Informatics

Kagan et al.
5-yr main shock model
5-yr main shock + aftershock model

Shen, Jackson, and Kagan
Geodetic main shock model
Geodetic main shock + aftershock model

Ward
combo81
geodetic81
geodetic85
geologic81
seismic81
simulation

WG 2002
National Hazard Model

Wiemer & Schorlemmer
Asperity Likelihood Model

California: 5-year Models

Zechar
Triple_S (alarm-based)
Triple_S (rate-based)

Kagan & Jackson
KJSS

California: 1-day Models:

Gerstenberger et al.
STEP

Zhuang et al.
ETAS
ETAS with optimization

Kagan & Jackson
KJSS

California: 3-month Models

Rhoades
EEPAS (5 versions)
PPE (2 versions)

Shebalin et al.
EAST

Western Pacific: 1-year Models

Zechar
Triple_S (alarm-based)
Triple_S (rate-based)

Kagan & Jackson
KJSS

Lombardi & Marzocchi
DBM

Western Pacific: 1-day Models

Kagan & Jackson
KJSS

Global: 1-year Models

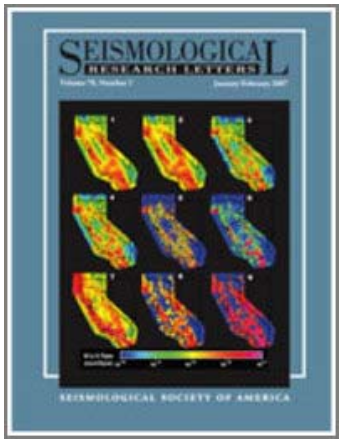
Zechar
Triple_S (alarm-based)
Triple_S (rate-based)

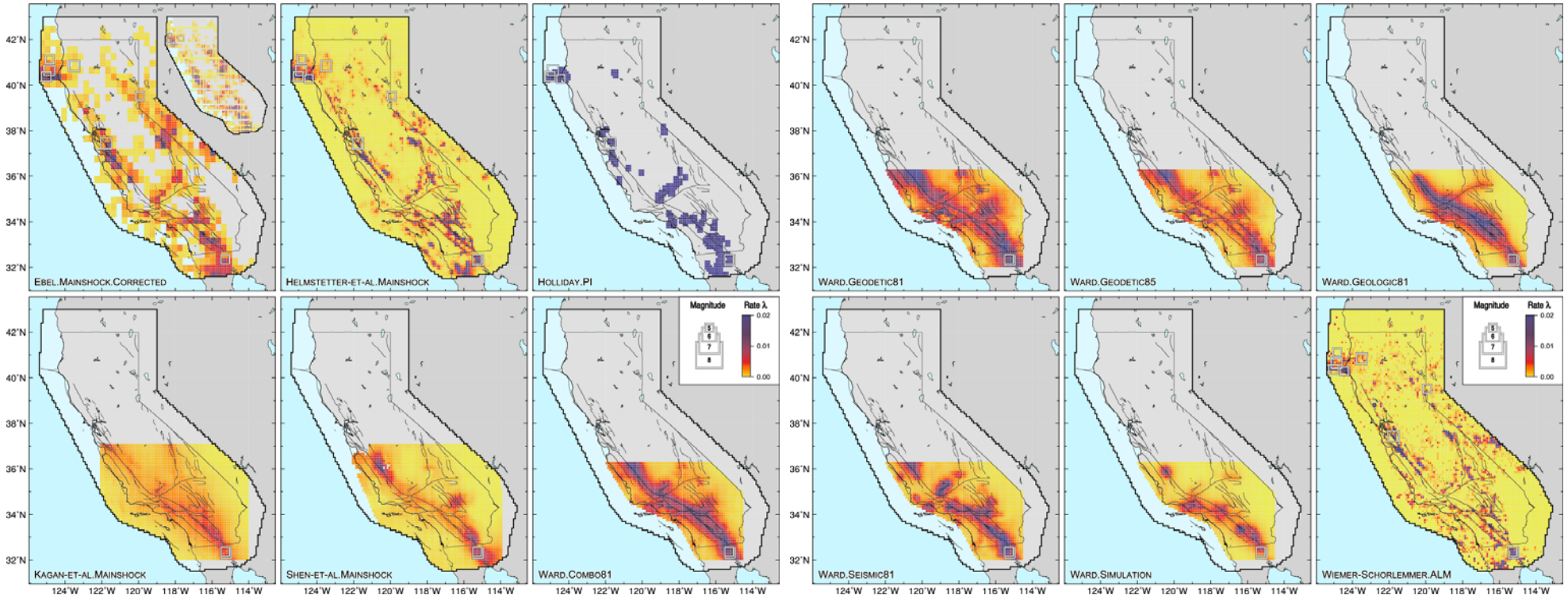
Lombardi & Marzocchi
DBM

44 Models

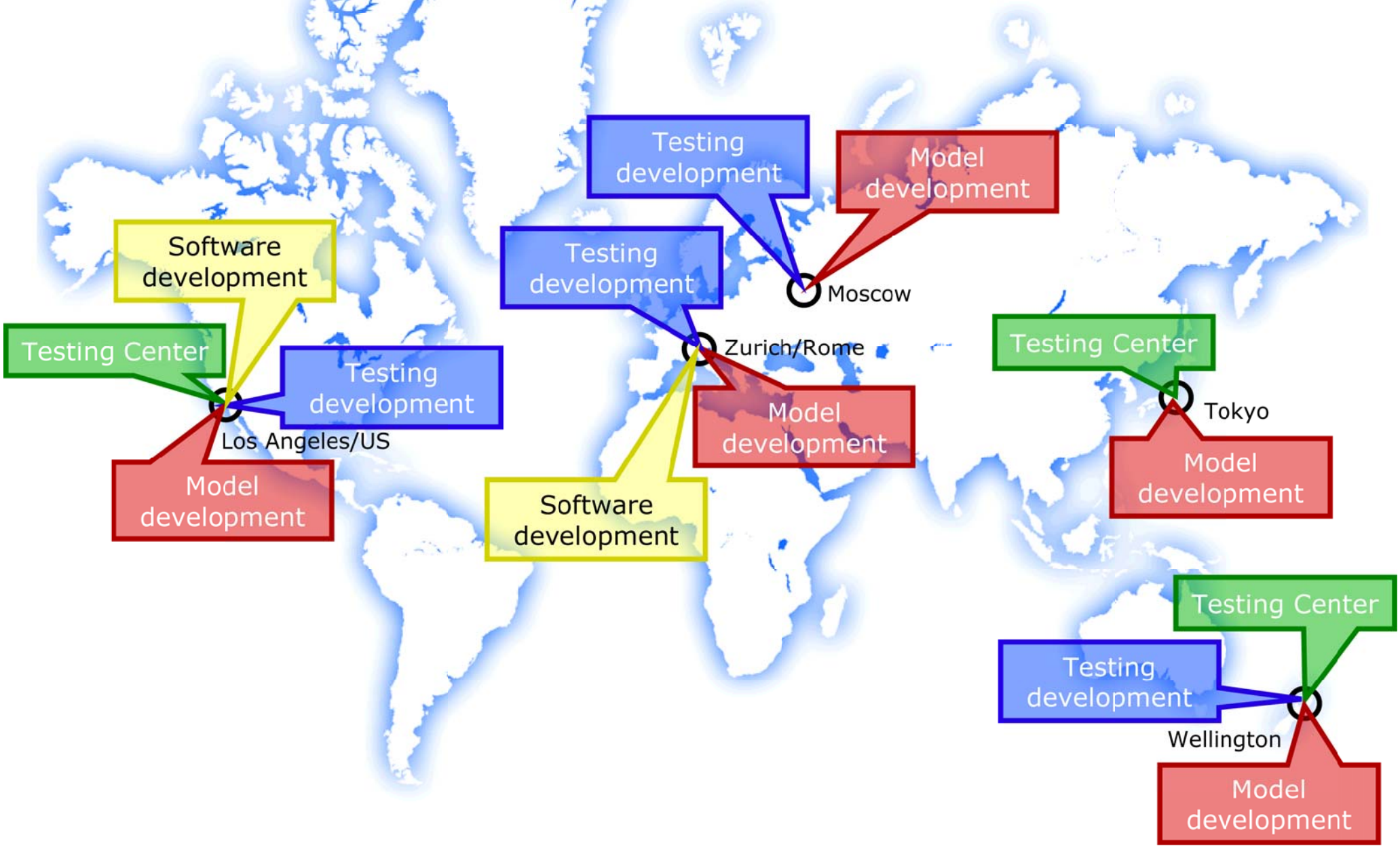
Testing Center (California)

- 19 5-year models have been submitted to the Testing Center
- Special volume of SRL was published





Global Collaboration



Summary

- CSEP expanded to key areas
- CSEP is globally established with 3(5) testing centers
- Blueprint for testing
  - Earthquake early warning
  - Earthquake source inversions
  - …
- CSEP is recognized for rigorous earthquake prediction research
  - Full characterization of regions and experiments
  - Standardization of procedures, formats, software, and experiments

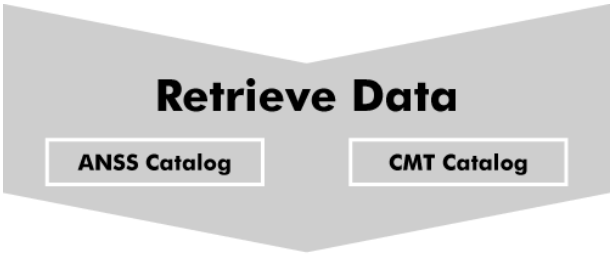
Thank You!

If you're doing an experiment, you should report everything that you think might make it invalid — not only what you think is right about it... Details that could throw doubt on your interpretation must be given, if you know them.

Richard P. Feynman

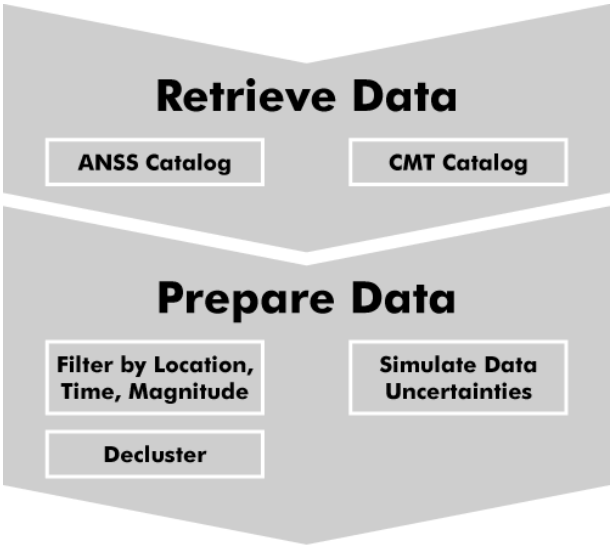
CSEP Software

- Retrive data on a daily basis



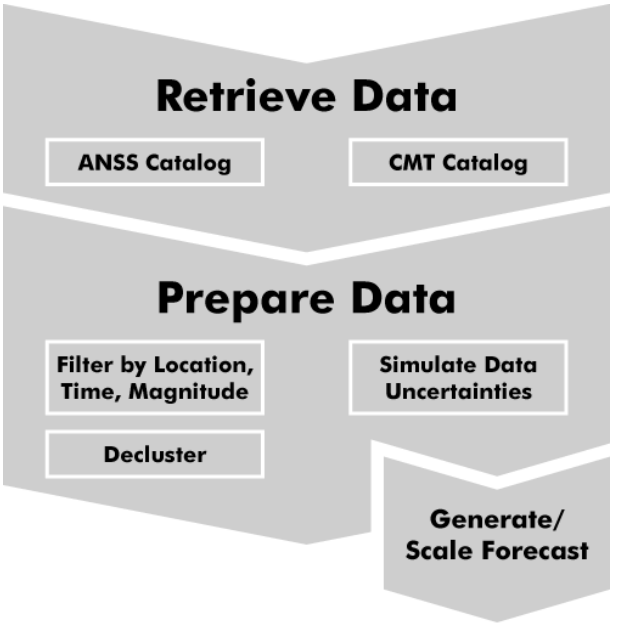
CSEP Software

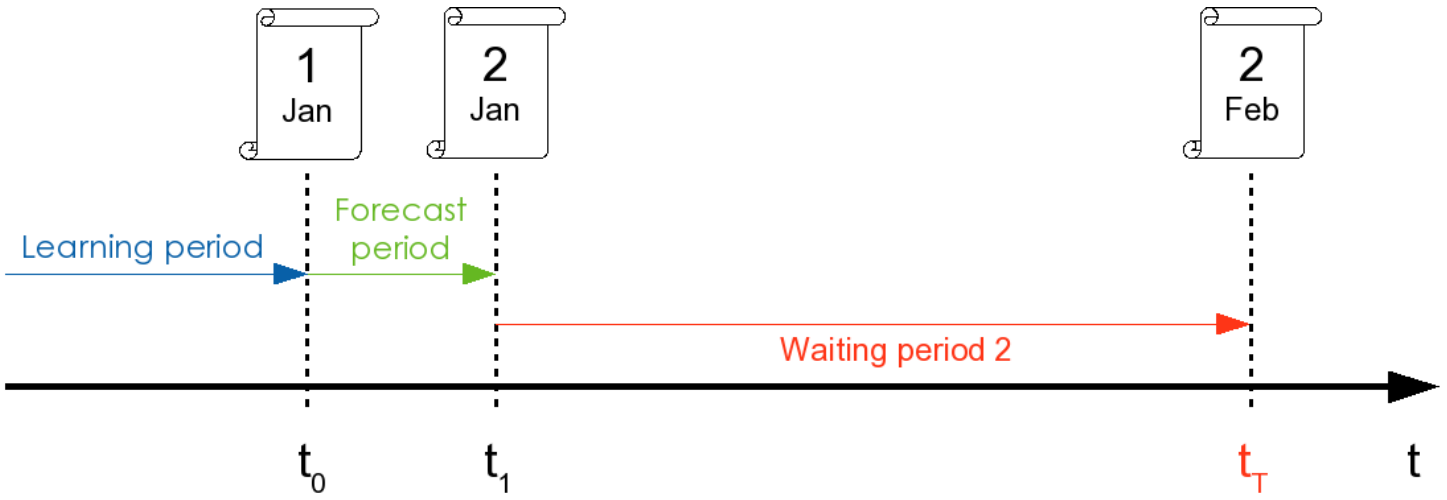
- Retrive data on a daily basis
- Prepare data sets for
  - Models
  - Testing



CSEP Software

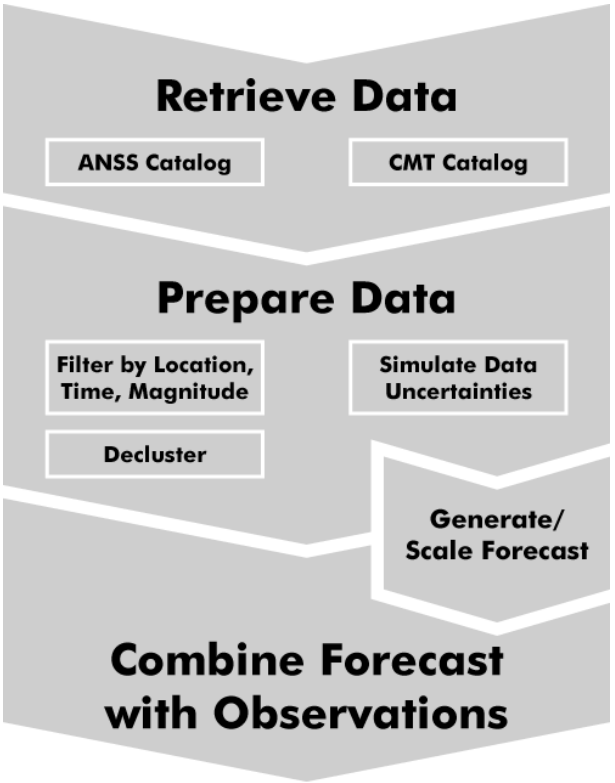
- Retrive data on a daily basis
- Prepare data sets for
  - Models
  - Testing





CSEP Software

- Retrive data on a daily basis
- Prepare data sets for
  - Models
  - Testing
- Prepare for testing



CSEP Software

- Retrive data on a daily basis
- Prepare data sets for
  - Models
  - Testing
- Prepare for testing
- More tests are being developed

CSEP Software

- Retrive data on a daily basis
- Prepare data sets for
  - Models
  - Testing
- Prepare for testing
- More tests are being developed