

RELM

RELM Testing Center

Danijel Schorlemmer, Matt Gerstenberger,
Dave Jackson, Stefan Wiemer, Yan Kagan,
Lucy Jones, Ned Field

Why a Testing Center?

- Re-run the tests with alternative options (different magnitudes)
- Re-run the tests in case of bugs in the testing procedure
- Document each models code and potential changes to it
- Track the modeler's additional data and deposit it
- 'Certify' all steps of testing

	QS5	TD	QS1	RT
Forecast duration	5 years	1 day	1 year	1 day
Aftershocks	no	yes	yes	yes
Magnitude range	5-9	4-9	5-9	4-9
Region	CA++	CA+	CA++	CA+
Modeler provides	numbers	code	code	code
Revised data	no	yes	yes	no

Test data-consistency and compare each model's performance

Parameter uncertainties

error distributions of location, magnitude, and FM angles

Independence probabilities (Declustering)

aftershock vs. main shock

Magnitude completeness windows

time and magnitude

Resolution independent

location, magnitude, focal mechanisms angles

Analysis of spatial and magnitude-range performance

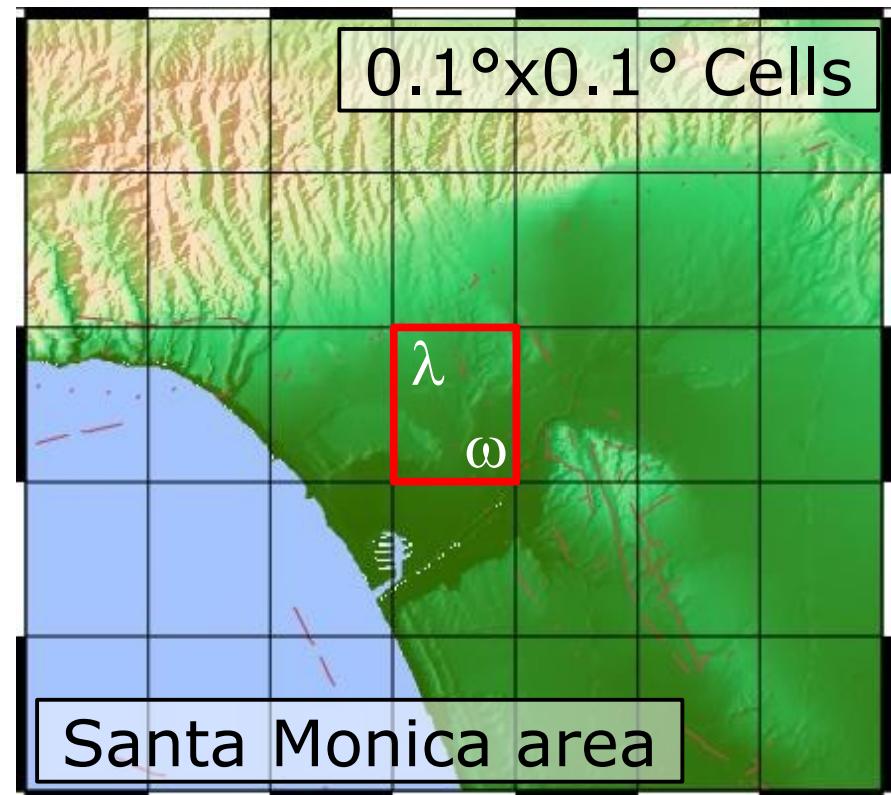
The testing area is separated into cells

A bin defines a volume (cell), magnitude range, and range of focal mechanism angles for which a forecast is issued

In each bin: **Expectation** Γ^M
Observation Γ^M

The proposed default binning:

Lon/Lat	$0.1^\circ \times 0.1^\circ$
Depth	0-50km
Magnitude	0.1
Focal Mech.	None (30°)



Computing the **likelihood** as the Poissonian probability of making an observation given an expectation.

We apply 3 different tests:

L-Test

Examines the consistency of a model with the observation (in the likelihood space)

N-Test

Test if the number of observed events is in the range of the expectation of a model

R-Test

Compares 2 models by its log-likelihood-ratio. It estimates the differences in spatial performance.

In each test we compare observed values with the value obtained from catalogs simulated based on expectations of a model.

RELM

How do we evaluate?

The winning model must:

- Beat all other models in the **R-Test**
- Show consistency with the observation in the **L-/N-Test**

Testing Center

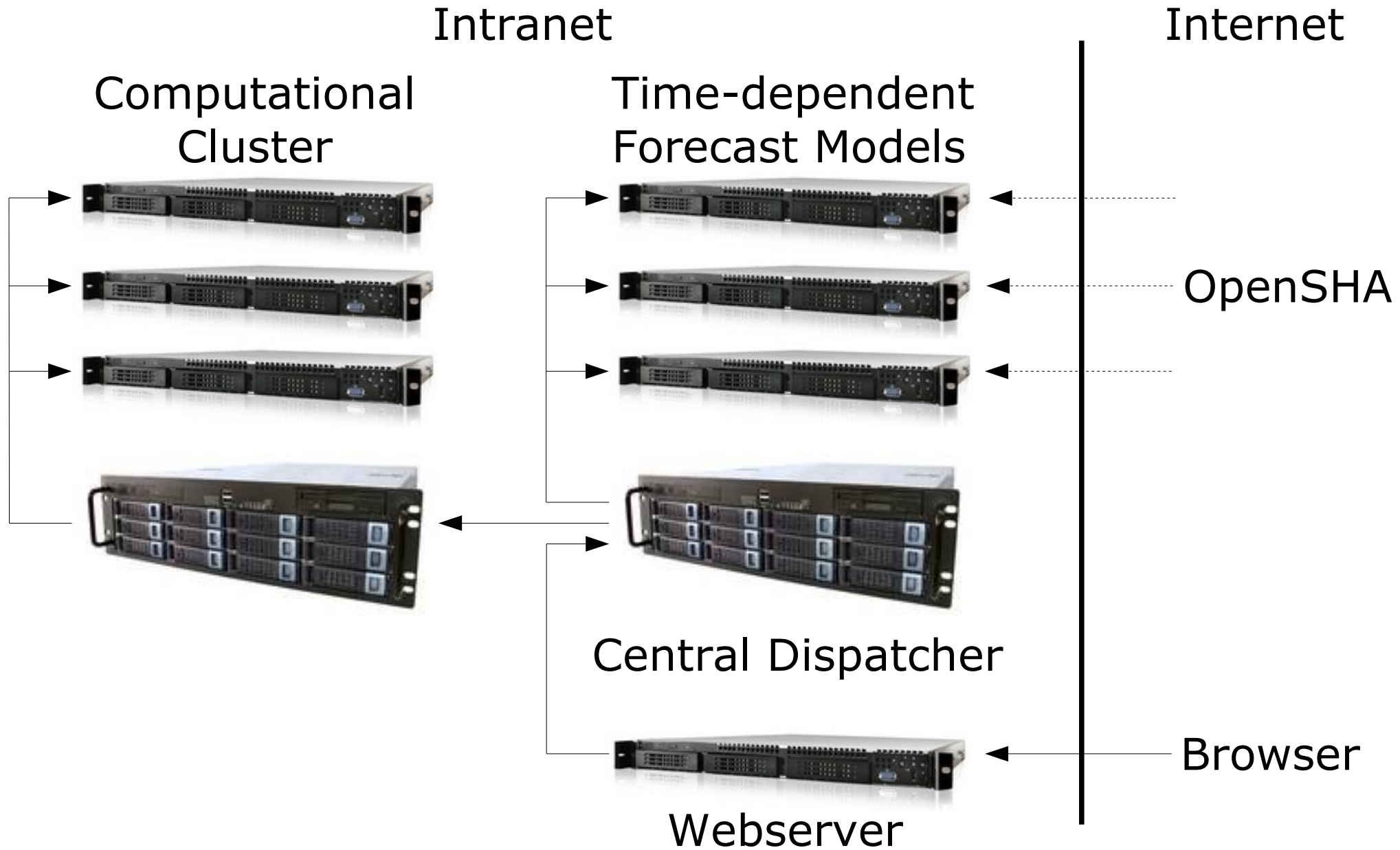
Each time-dependent model gets a computer in the testing center:

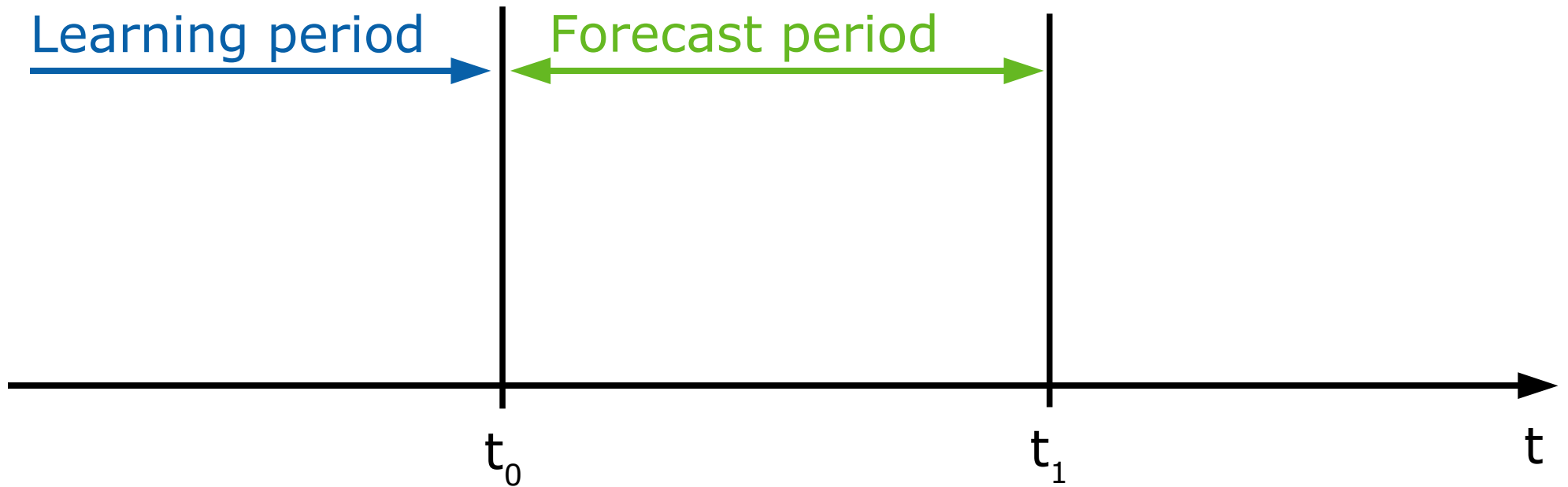
The modeler installs the model's code

The code is deposited in a CVS (Subversion) repository

If the code runs, modeler access to the computer is removed

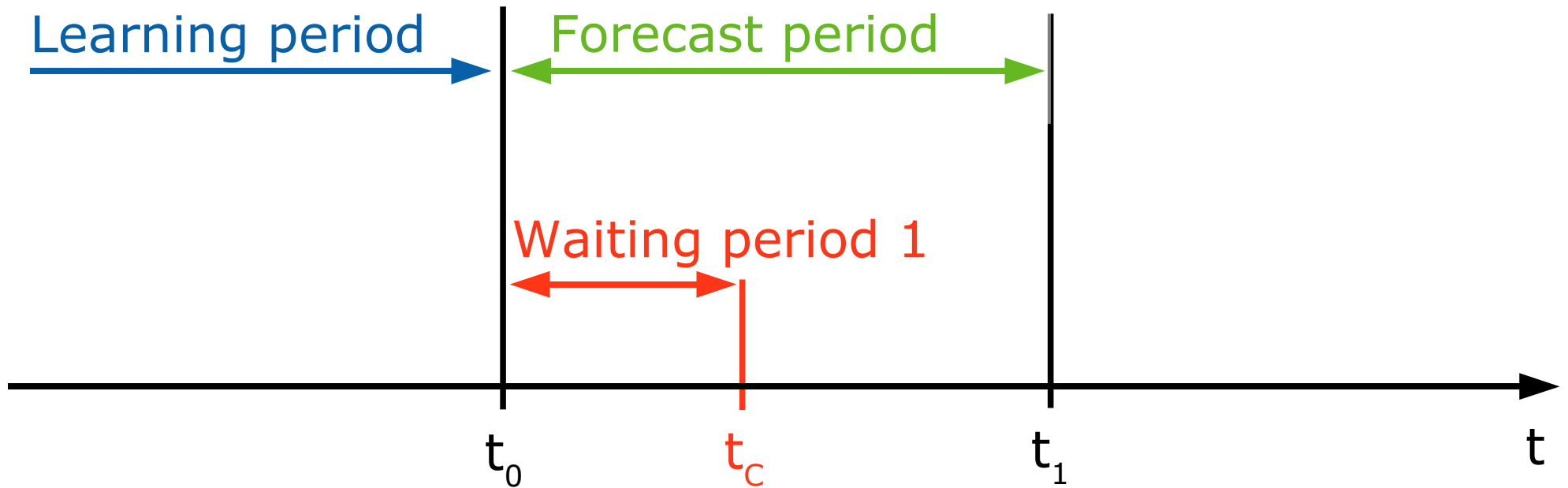
In case the code crashes, the modeler is allowed to fix the problem: With the CVS, we can track the changes and document them





Learning period

Any data not provided by the testing center can be deposited by the modeler. This data gets a time-stamp for reproducing results during re-runs.

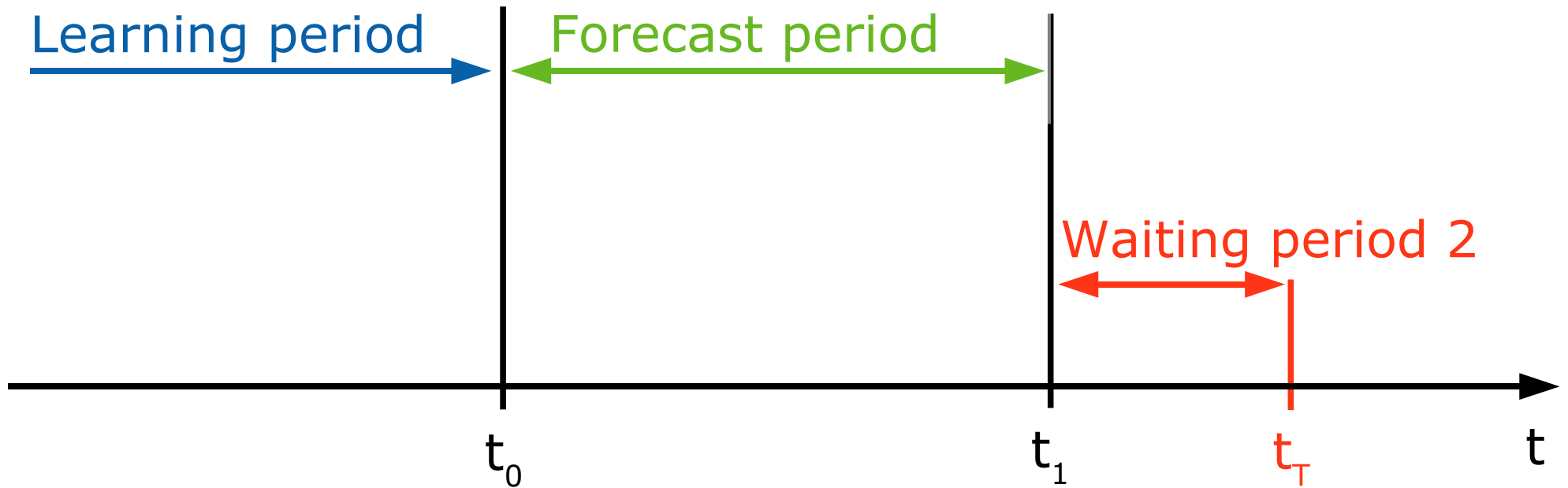


Waiting period 1

Time before the 'authorized' catalog is provided.

t_c

Compute the forecasts.

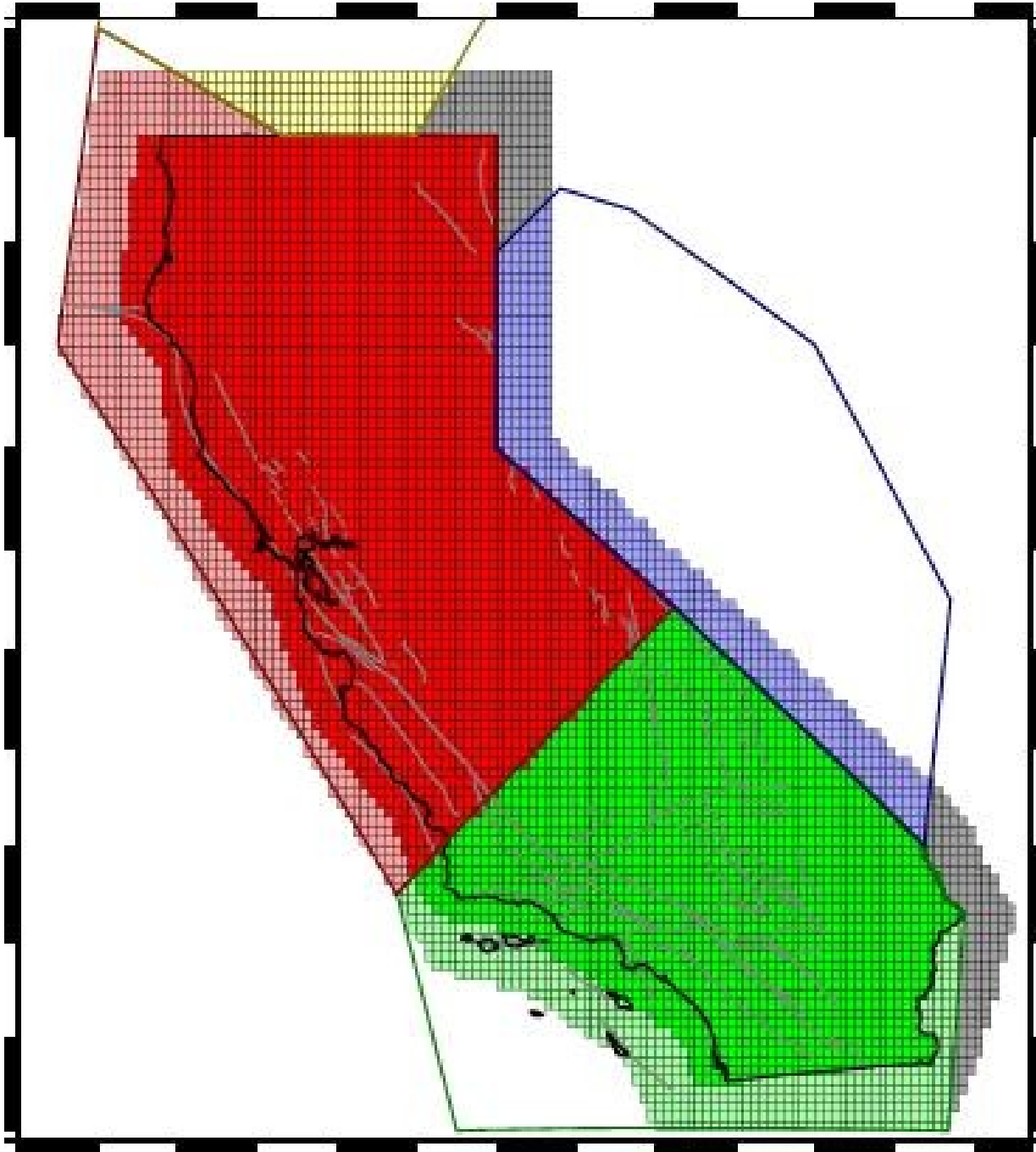


Waiting period 2

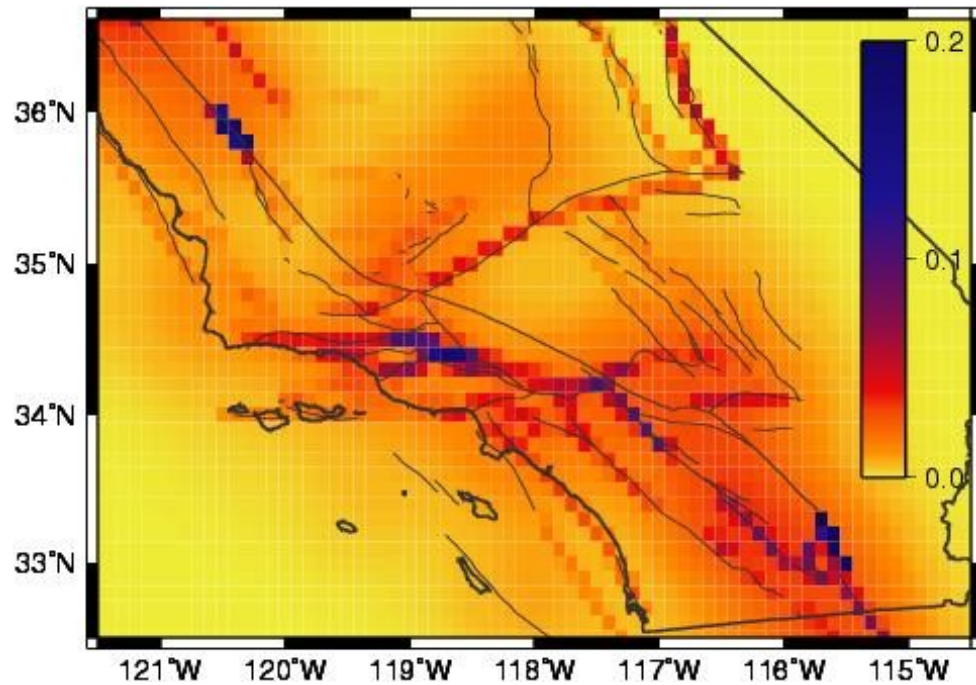
Time before the models can be tested against the 'authorized' catalog after the forecast period is finished.

t_T

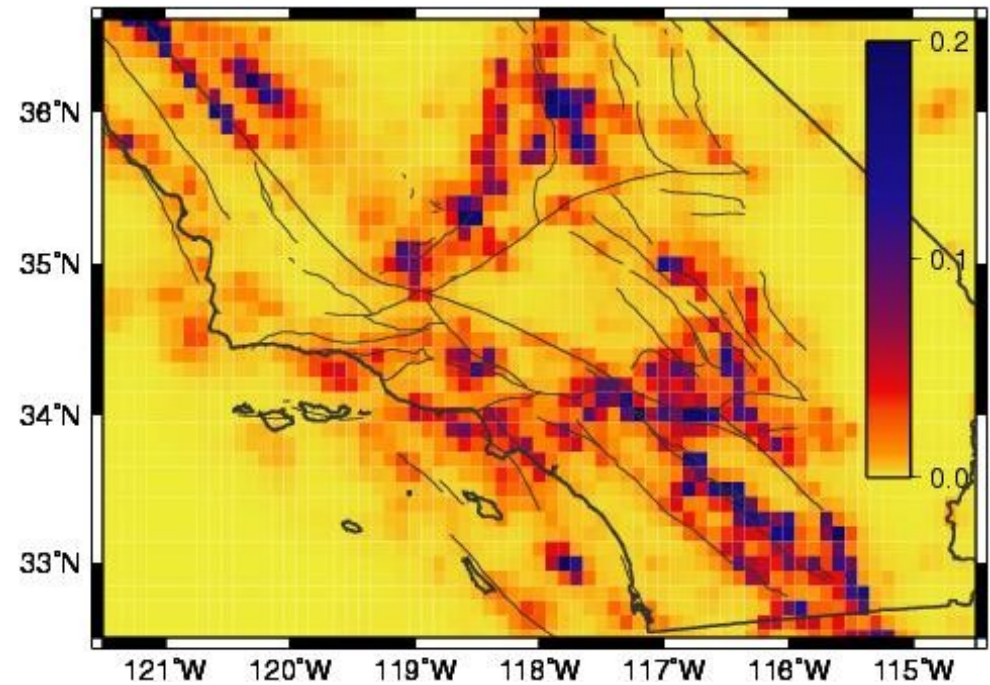
Perform the tests and compute the results.



20-year expectations for events of magnitude $M \geq 5$



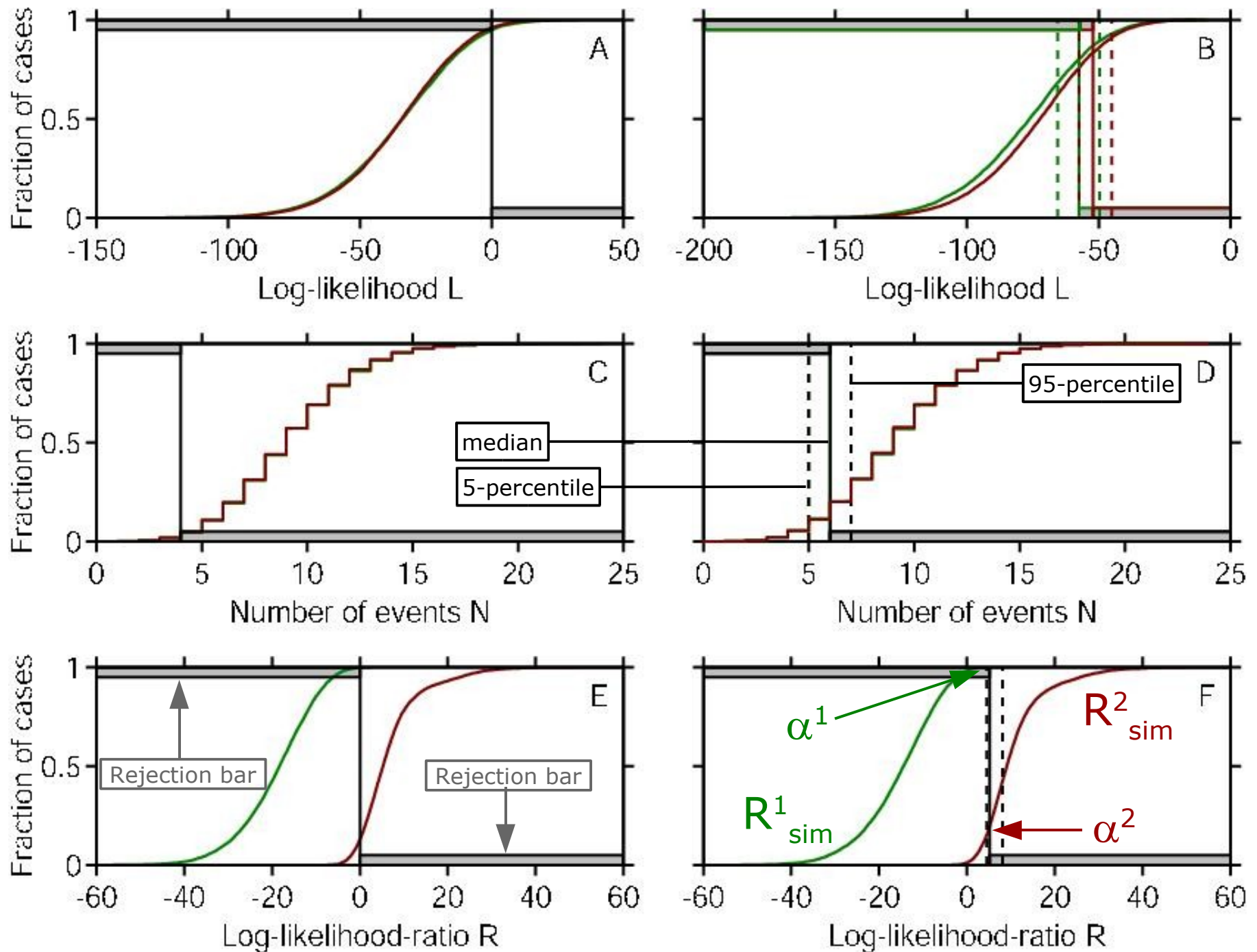
[Frankel, 1996]
(smoothed to grid)



[Helmstetter, submitted]

RELM

Example



Testing Center for grid-based forecast models

- Rigorous test of fully specified (in advance) forecasts
- 5-year test within the RELM project
- Community-accepted testing algorithms
- Setting a new standard of rigorous forecast tests
- Nucleation point for a RELM successor or RELM-like efforts in Europa and worldwide