Bivariate Counting Processes for Risk Management

Mathieu Boudreault & **Arthur Charpentier**

Université du Québec à Montréal

charpentier.arthur@uqam.ca

http ://freakonometrics.blog.free.fr/

IFM2, Mathematical Finance Days, May 2012

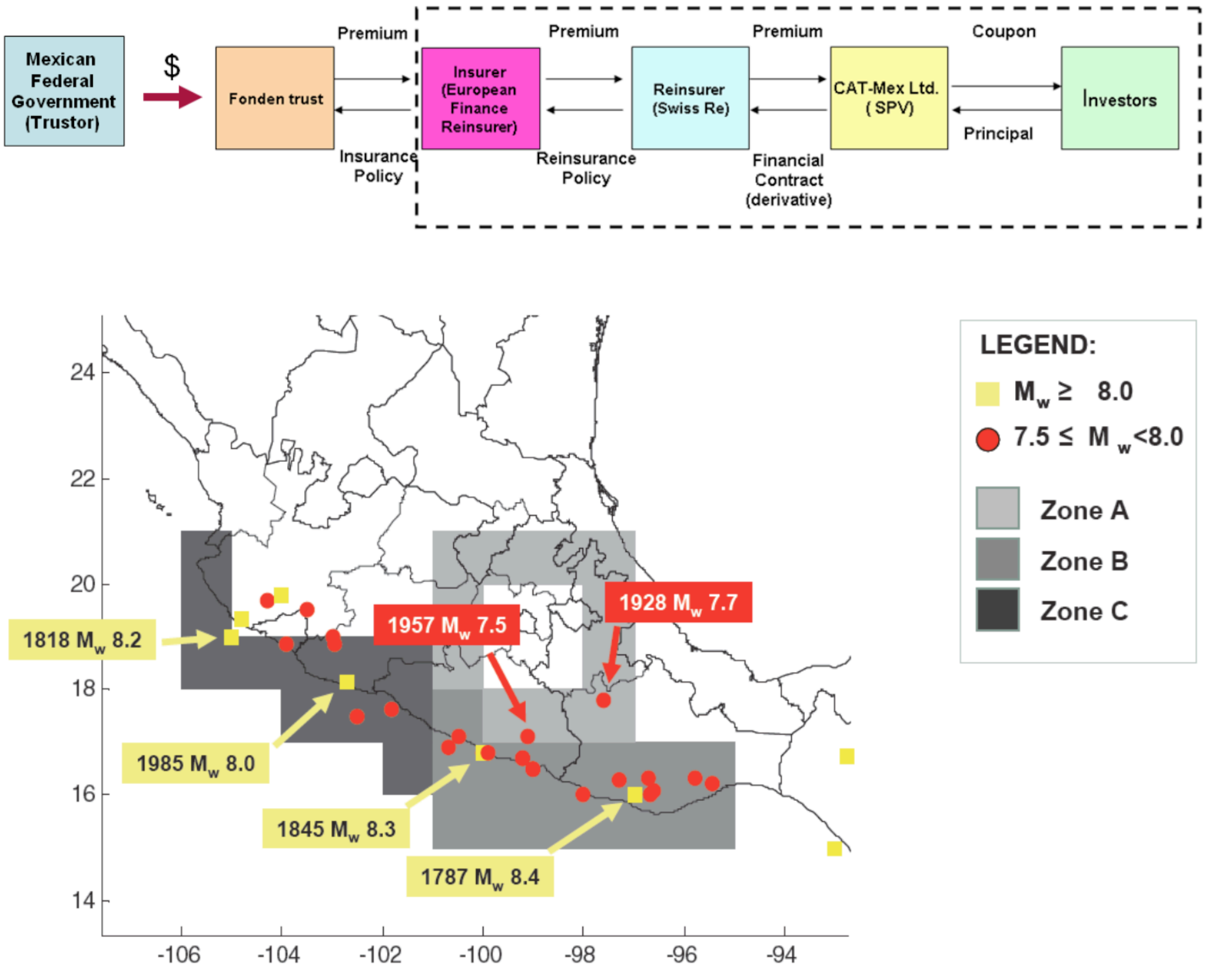


Figure : Mexican catastrophe bond, 2006-2009, via Cabrera (2006)

Motivation : Mexican (earthquake) catastrophe bond

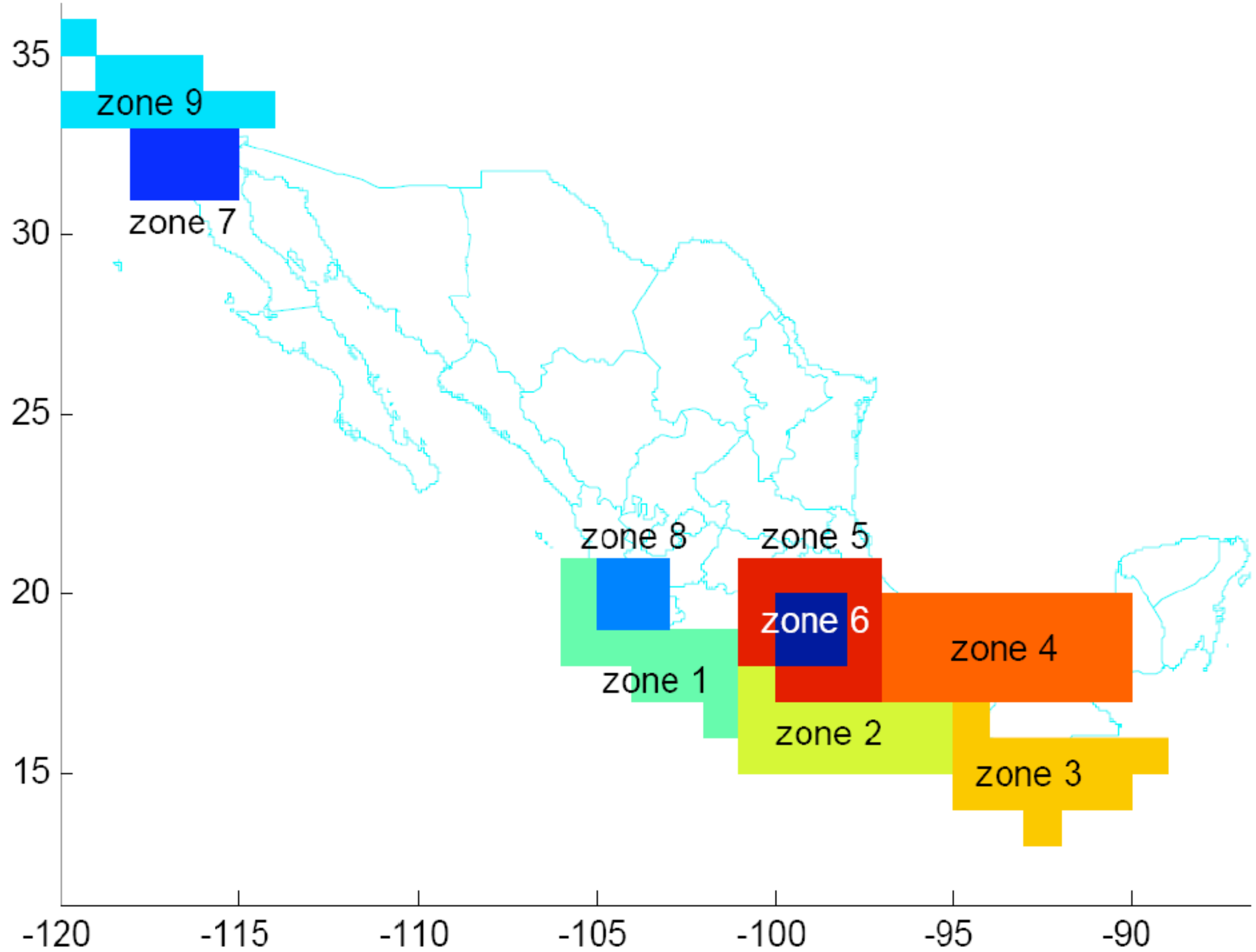
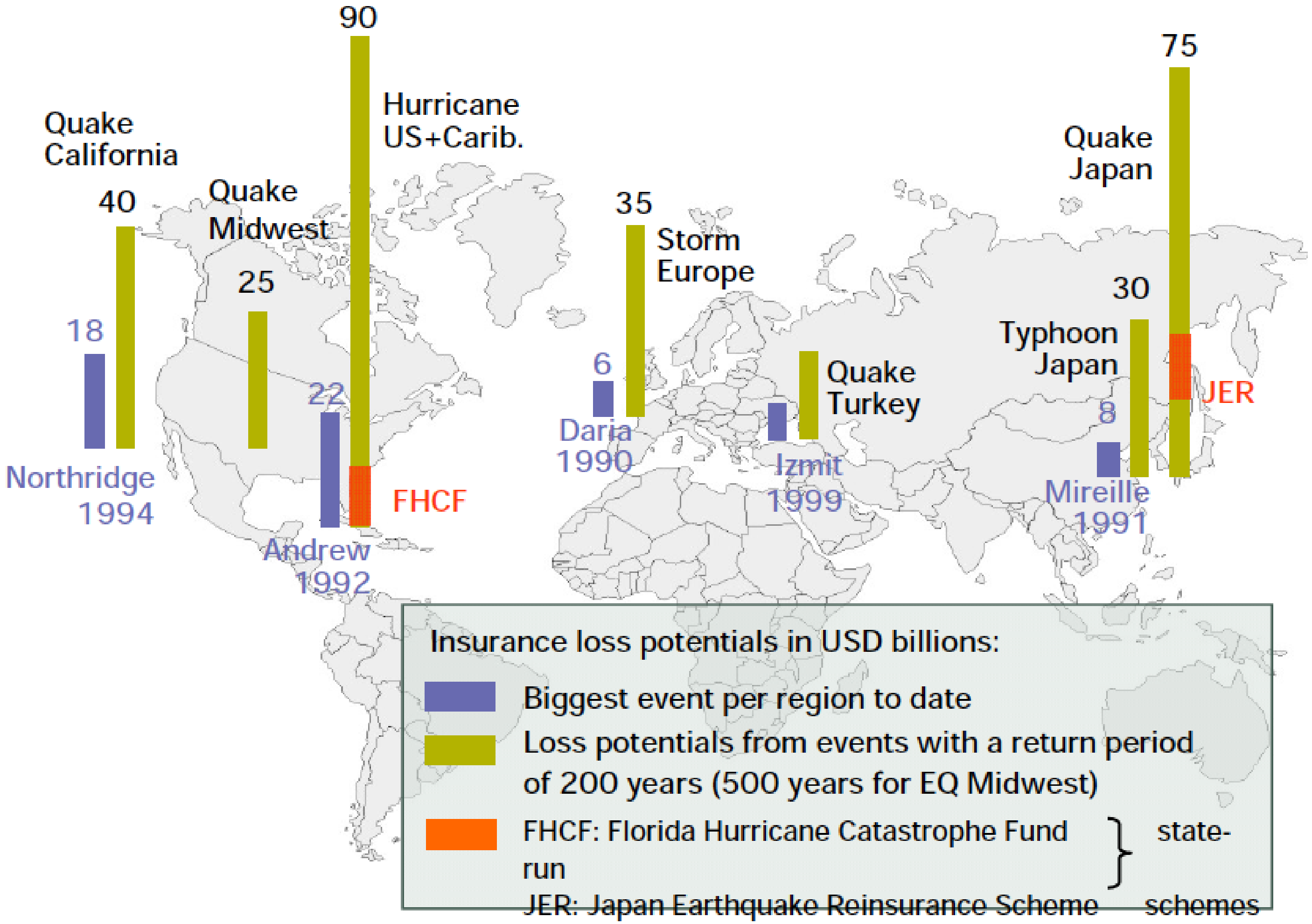


Figure : Mexican catastrophe bond, 2006-2009, via Cabrera (2006)


90
75
Quake
California
Hurricane
US+Carib.
Quake
Japan
40
Quake
Midwest
35
Storm
Europe
25
30
18
Typhoon
Japan
6
Quake
Turkey
22
8
JER
Daria
1990
Northridge
1994
FHCF
Izmit
1999
Mireille
1991
Andrew
1992
Insurance loss potentials in USD billions:
Biggest event per region to date
Loss potentials from events with a return period of 200 years (500 years for EQ Midwest)
FHCF: Florida Hurricane Catastrophe Fund
JER: Japan Earthquake Reinsurance Scheme
state-run schemes

Motivation

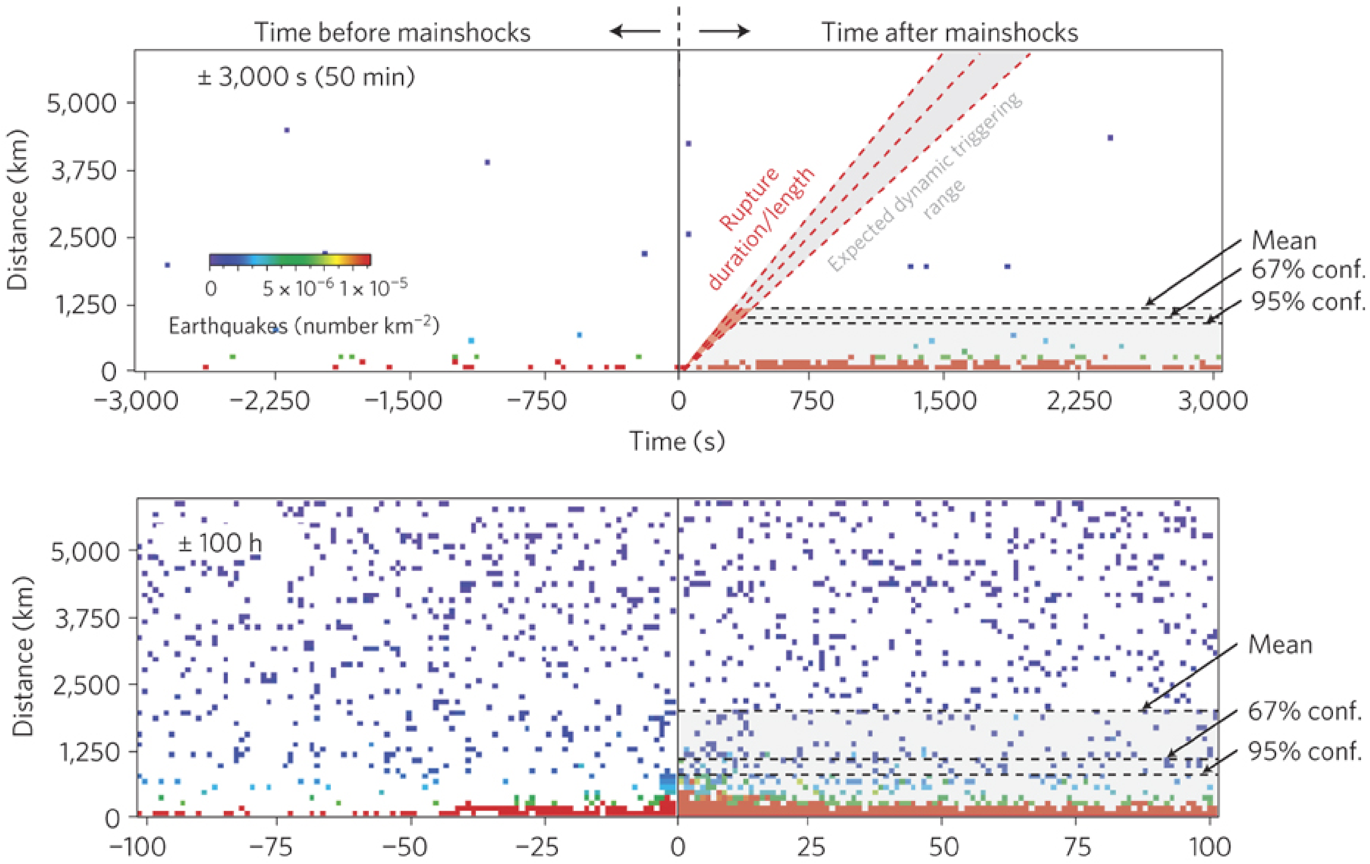
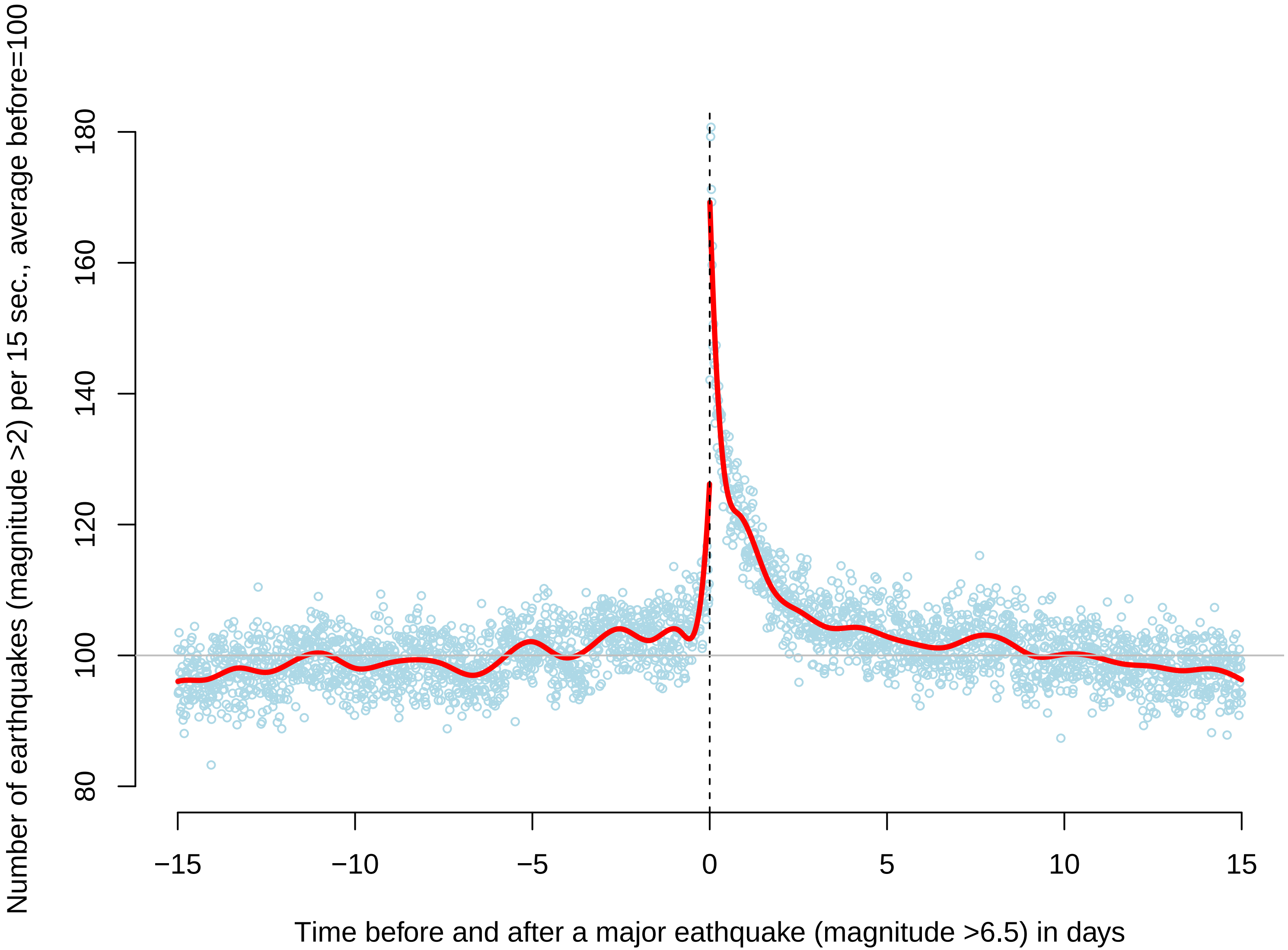


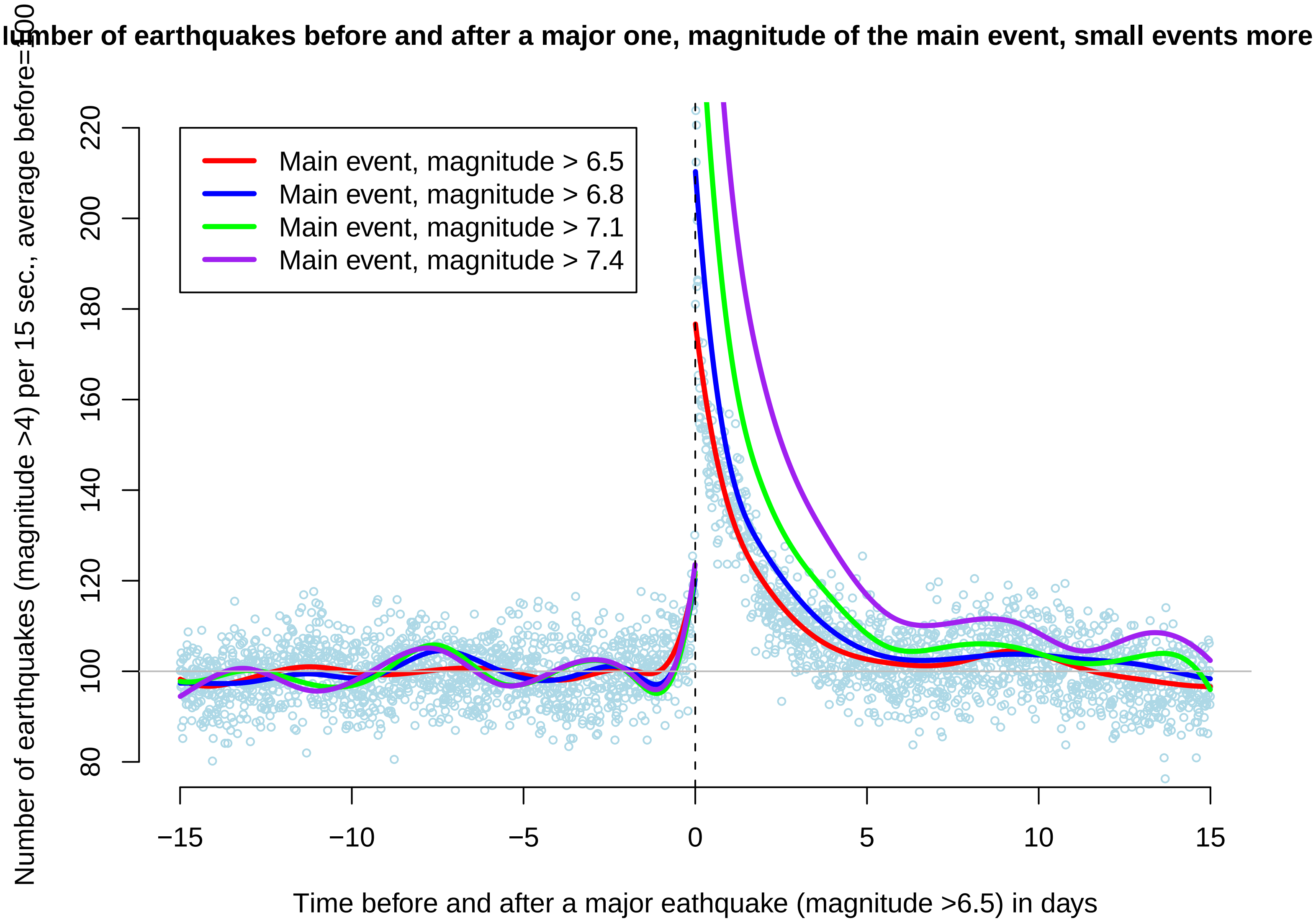
Figure : Time and distance distribution (to 6,000 km) of large (5<M<7) aftershocks from 205 M≥7 mainshocks (in sec. and h.). Parsons & Velasco (2011)

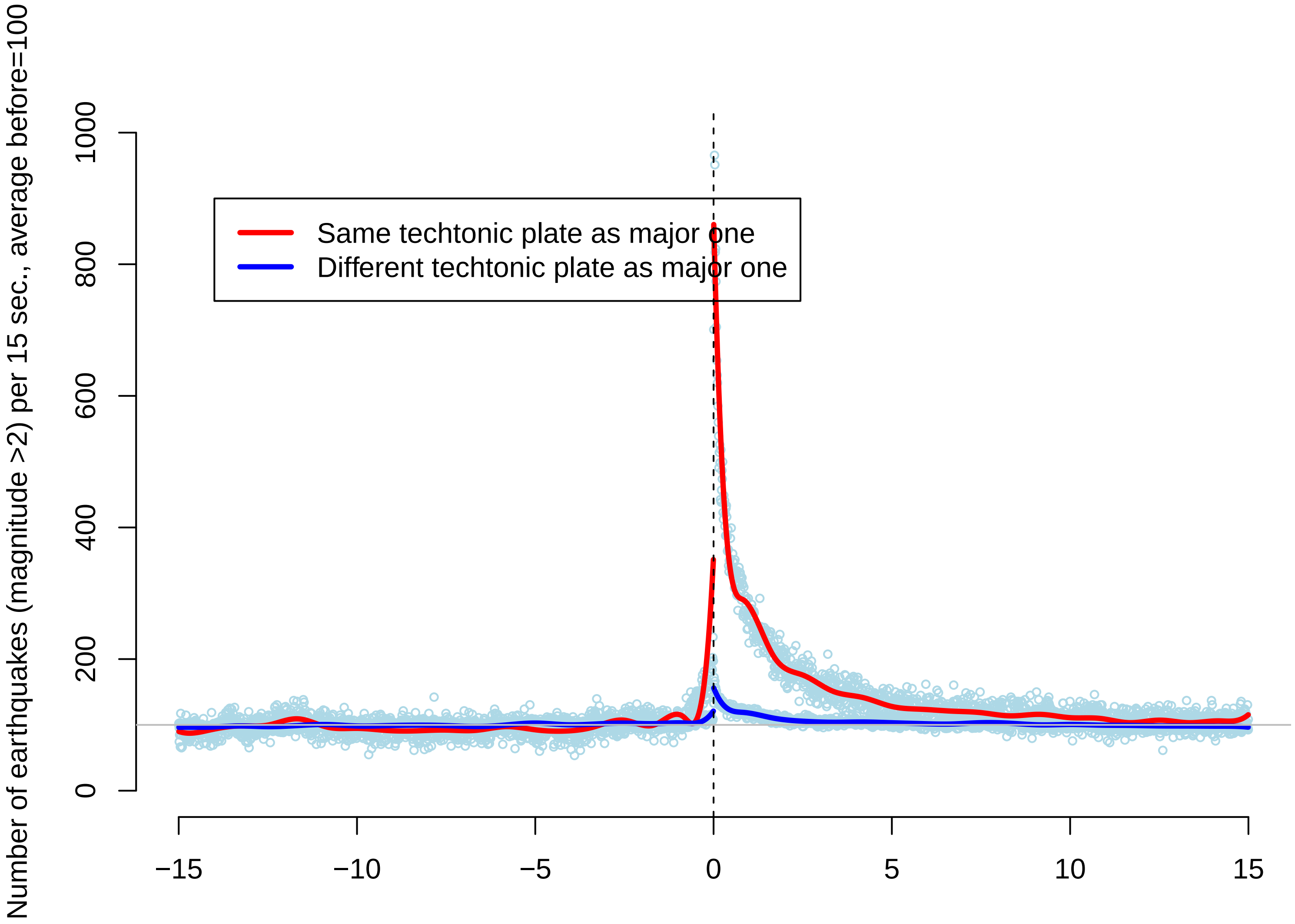
Motivation

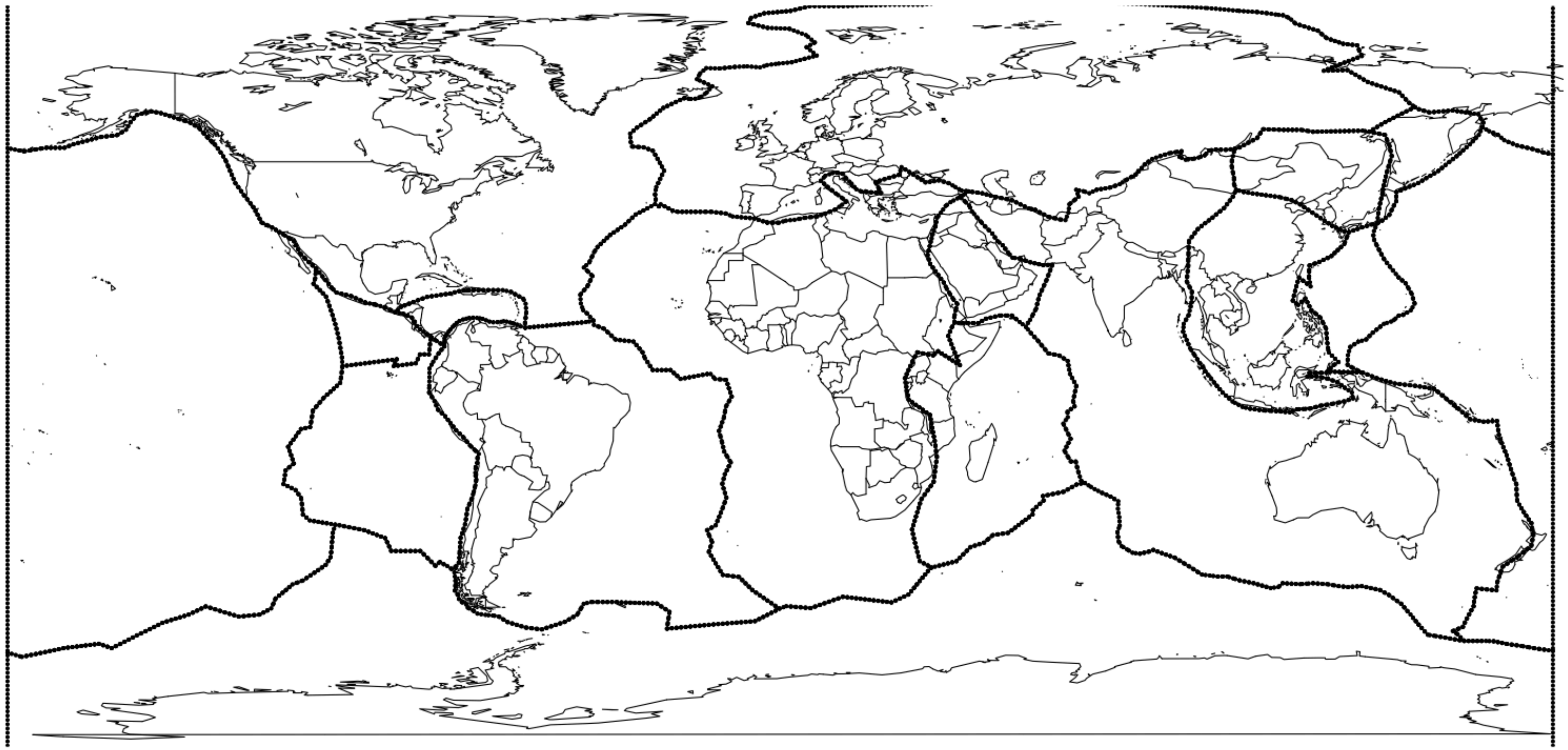
"*Large earthquakes are known to trigger earthquakes elsewhere. Damaging large aftershocks occur close to the mainshock and microearthquakes are triggered by passing seismic waves at significant distances from the mainshock. It is unclear, however, whether bigger, more damaging earthquakes are routinely triggered at distances far from the mainshock, heightening the global seismic hazard after every large earthquake. Here we assemble a catalogue of all possible earthquakes greater than M5 that might have been triggered by every M7 or larger mainshock during the past 30 years.* [...] *We observe a significant increase in the rate of seismic activity at distances confined to within two to three rupture lengths of the mainshock. Thus, we conclude that the regional hazard of larger earthquakes is increased after a mainshock, but the global hazard is not.*" Parsons & Velasco (2011)

Figure : Number of earthquakes (magnitude exceeding 2.0, per 15 sec.) following a large earthquake (of magnitude 6.5), normalized so that the expected number of earthquakes before and after is 100.


Number of earthquakes (magnitude >2) per 15 sec., average before=100
80
100
120
140
160
180
−15
−10
−5
0
5
10
15
Time before and after a major eathquake (magnitude >6.5) in days

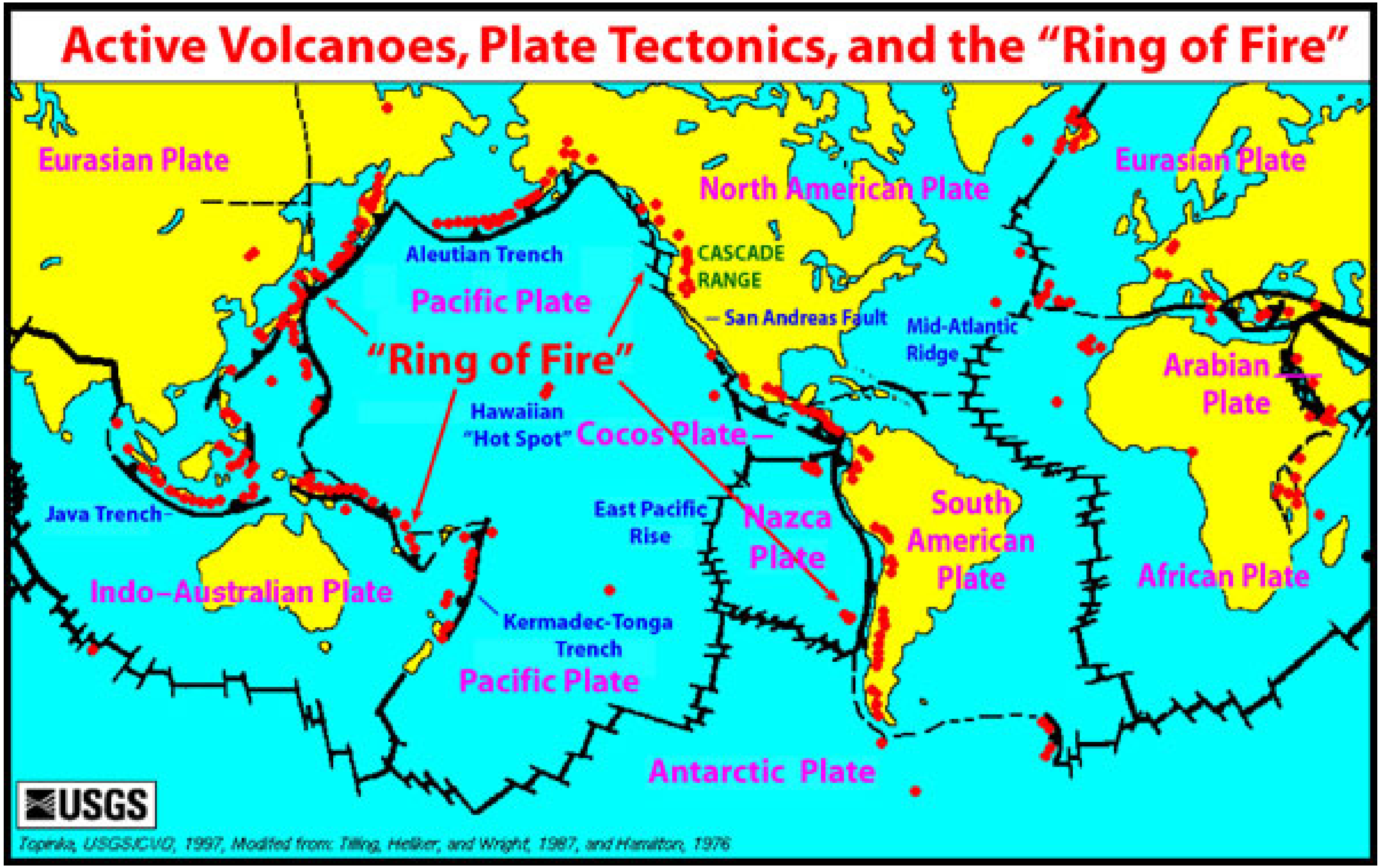

Number of earthquakes before and after a major one, magnitude of the main event, small events more
Number of earthquakes (magnitude >4) per 15 sec., average before=100
Main event, magnitude > 6.5
Main event, magnitude > 6.8
Main event, magnitude > 7.1
Main event, magnitude > 7.4
Time before and after a major eathquake (magnitude >6.5) in days
80
100
120
140
160
180
200
220
−15
−10
−5
0
5
10
15


Same techtonic plate as major one
Different techtonic plate as major one
Number of earthquakes (magnitude >2) per 15 sec., average before=100
0
200
400
600
800
1000
−15
−10
−5
0
5
10
15
Time before and after a major eathquake (magnitude >6.5) in days



Shapefiles from

http://www.colorado.edu/geography/foote/maps/assign/hotspots/hotspots.html

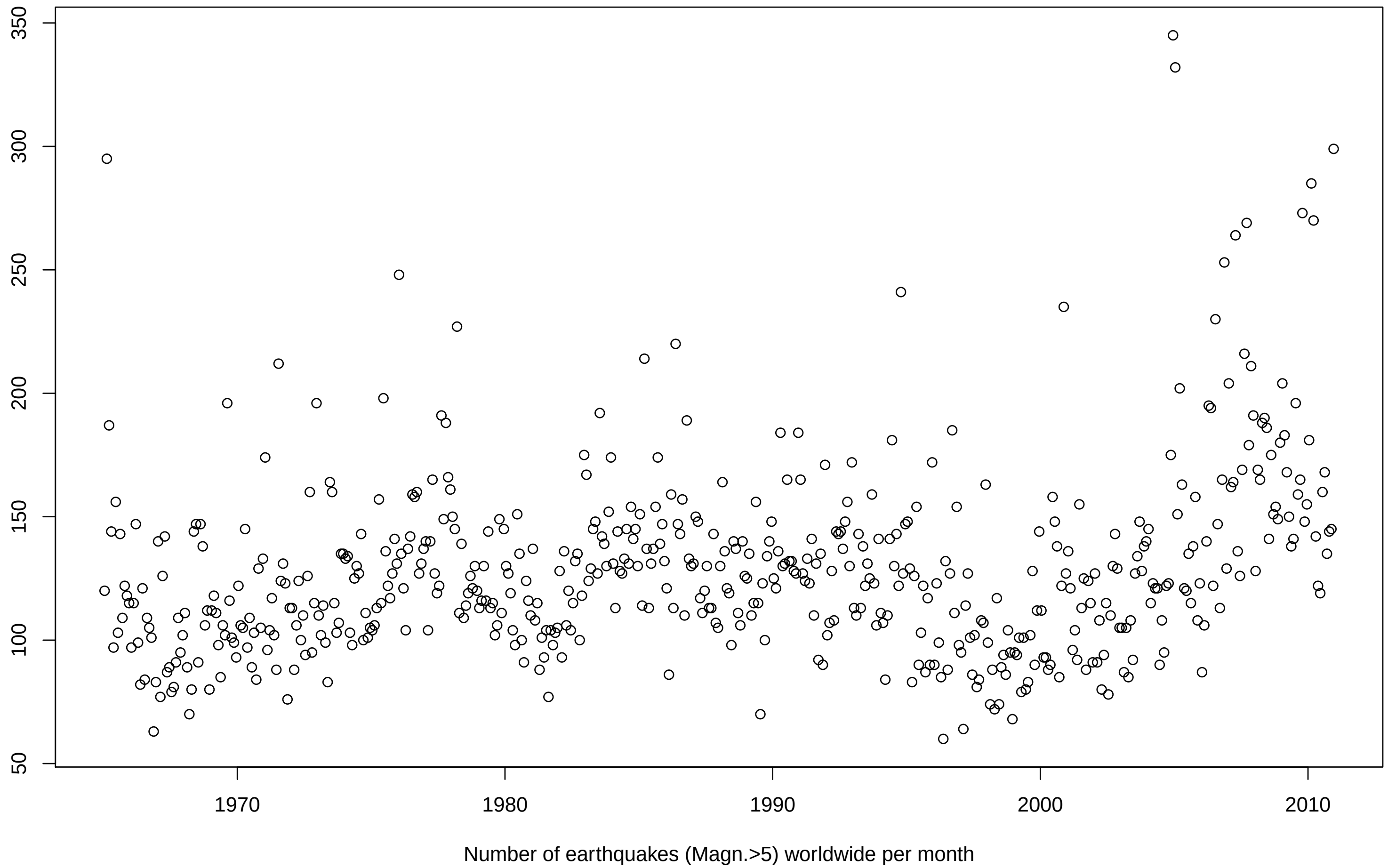

Active Volcanoes, Plate Tectonics, and the "Ring of Fire"
Eurasian Plate
North American Plate
Eurasian Plate
Aleutian Trench
CASCADE
RANGE
Pacific Plate
San Andreas Fault
Mid-Atlantic
Ridge
"Ring of Fire"
Arabian
Plate
Hawaiian
"Hot Spot"
Cocos Plate
Java Trench
East Pacific
Rise
Nazca
Plate
South
American
Plate
Indo-Australian Plate
African Plate
Kermadec-Tonga
Trench
Pacific Plate
Antarctic Plate
USGS
Topinka, USGS/CVO, 1997, Modified from: Tilling, Heliker, and Wright, 1987, and Hamilton, 1976

Agenda

- Motivation : earthquake risk and Parsons & Velasco (2011)
- Modeling dynamics
  - AR(1) : Gaussian autoregressive processes (as a starting point)
  - VAR(1) : multiple AR(1) processes, possible correlated
  - INAR(1) : autoregressive processes for counting variates
  - MINAR(1) : multiple counting processes
- Application to earthquakes frequency
  - counting earthquakes on tectonic plates
  - causality between different tectonic plates
  - counting earthquakes with different magnitudes

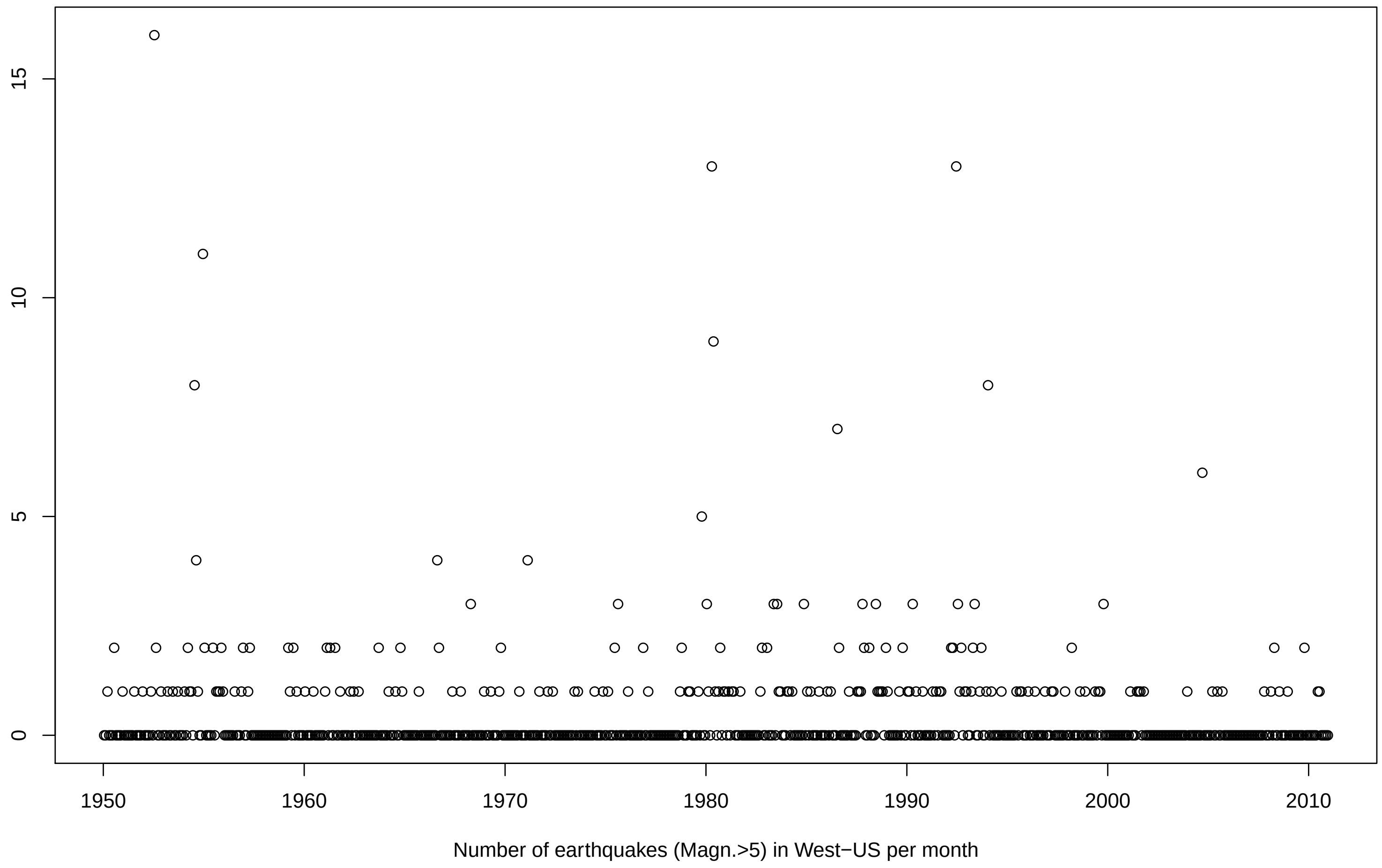
(ANSS) http://www.ncedc.org/cnss/catalog-search.html

Number of earthquakes (Magnitude $\geq$ 5) per month, worldwide



(ANSS) http://www.ncedc.org/cnss/catalog-search.html

Number of earthquakes (Magnitude ≥ 5) per month, in western U.S.



(Gaussian) Auto Regressive processes $AR(1)$

Definition A time series $(X_t)_{t\in\mathbb{N}}$ with values in $\mathbb{R}$ is called an AR(1) process if

$$X_t = \phi_0 + \phi_1 X_{t-1} + \varepsilon_t \quad (1)$$

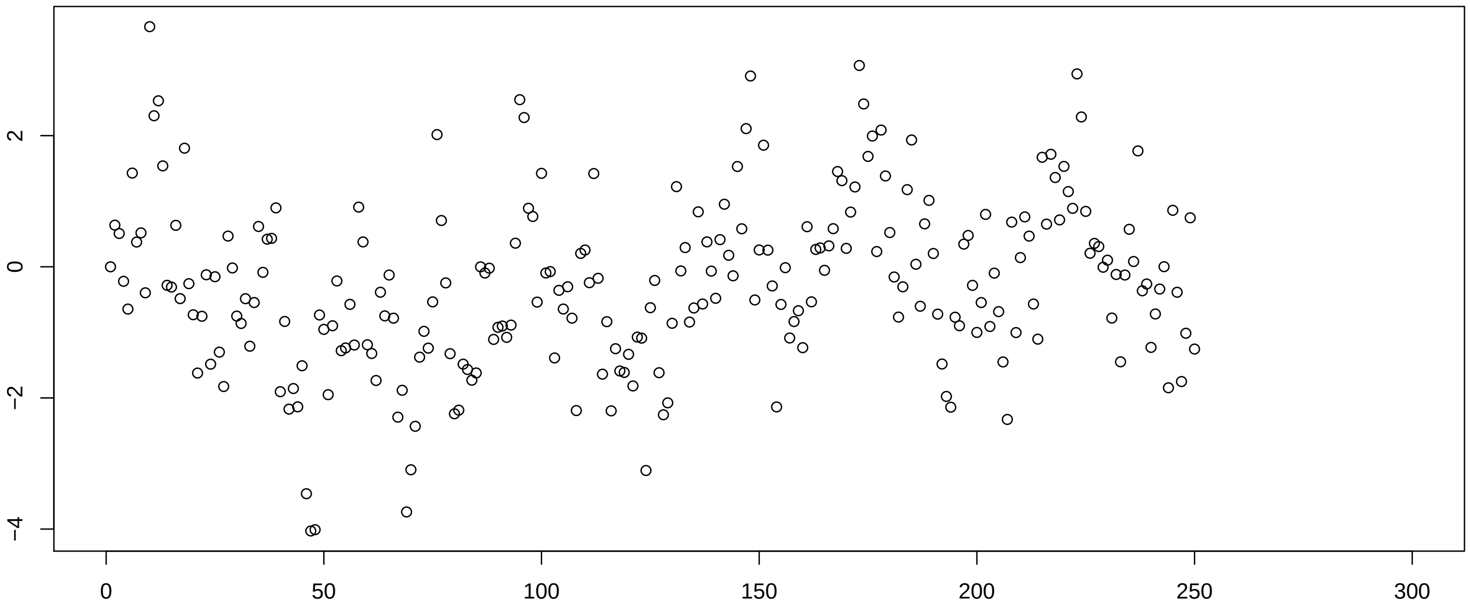
for all t, for real-valued parameters ϕ_0 and ϕ_1, and some i.i.d. random variables ε_t with values in $\mathbb{R}$.

It is common to assume that ε_t are independent variables, with a Gaussian distribution $\mathcal{N}(0,\sigma^2)$, with density

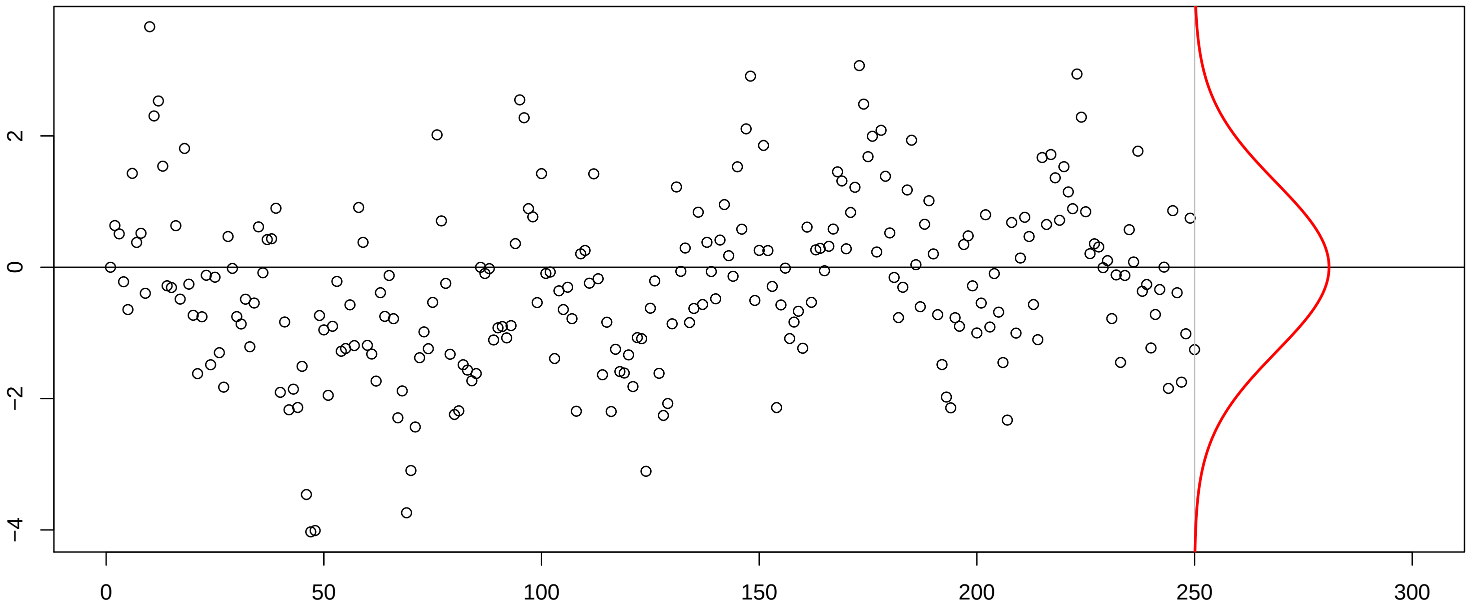
$$\varphi(\varepsilon) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{\varepsilon^2}{2\sigma^2}\right), \quad \varepsilon \in \mathbb{R}.$$

Note that we assume also that ε_t is independent of $\underline{X}_{t-1}$, i.e. past observations $X_0, X_1, \cdots, X_{t-1}$. Thus, $(\varepsilon_t)_{t\in\mathbb{N}}$ is called the innovation process.

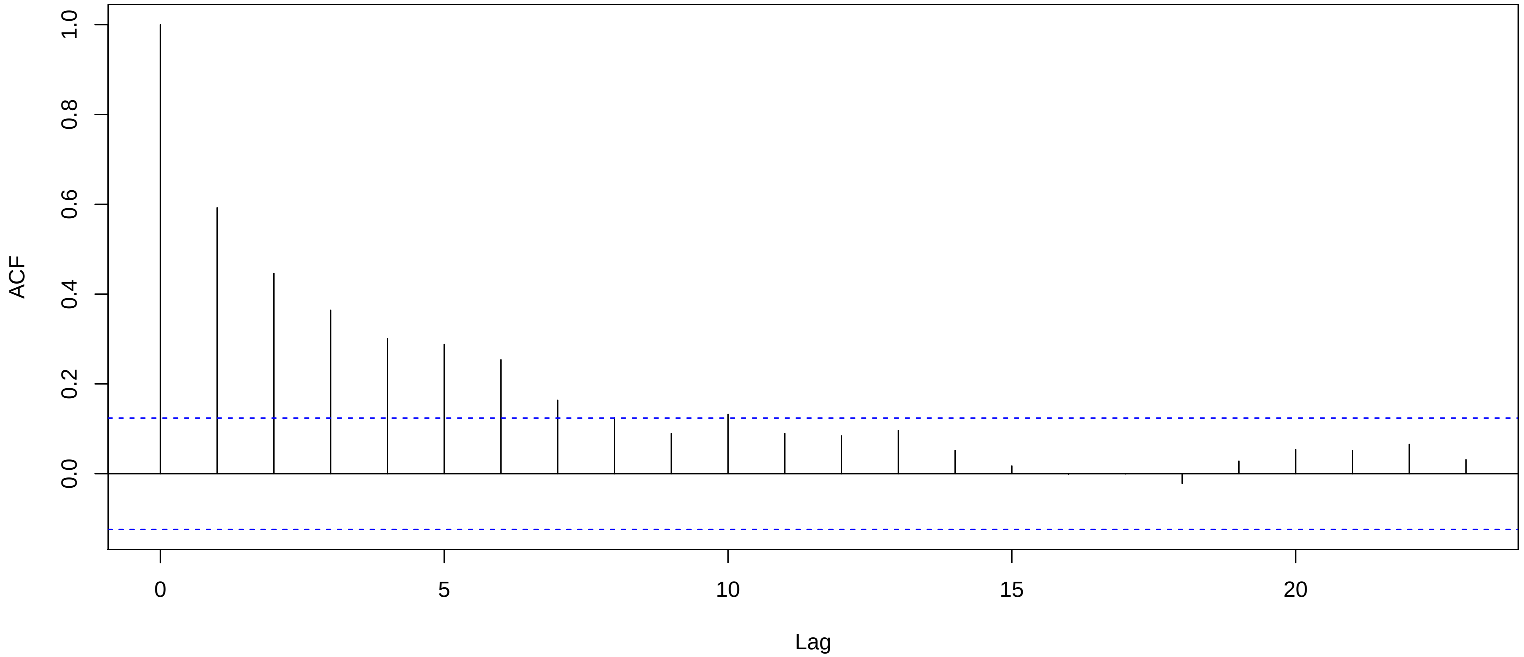
Example : $X_t = \phi_1 X_{t-1} + \varepsilon_t$ with $\varepsilon_t \sim \mathcal{N}(0,1)$, i.i.d., and $\phi = 0.6$



Example : $X_t = \phi_1 X_{t-1} + \varepsilon_t$ with $\varepsilon_t \sim \mathcal{N}(0,1)$, i.i.d., and $\phi = 0.6$



Example : $X_t = \phi_1 X_{t-1} + \varepsilon_t$: autocorrelation $\rho(h) = \text{corr}(X_t, X_{t-h}) = \phi_1^h$



Definition A time series $(X_t)_{t\in\mathbb{N}}$ is said to be (weakly) stationary if

- $\mathbb{E}(X_t)$ is independent of t ($=: \mu$)
- $\text{cov}(X_t, X_{t-h})$ is independent of t ($=: \gamma(h)$), called autocovariance function

Remark As a consequence, $\text{var}(X_t) = \mathbb{E}([X_t - \mathbb{E}(X_t)]^2)$ is independent of t ($=: \gamma(0)$). Define the autocorrelation function $\rho(\cdot)$ as

$$\rho(h) := \text{corr}(X_t, X_{t-h}) = \frac{\text{cov}(X_t, X_{t-h})}{\sqrt{\text{var}(X_t)\text{var}(X_{t-h})}} = \frac{\gamma(h)}{\gamma(0)}, \quad \forall h \in \mathbb{N}.$$

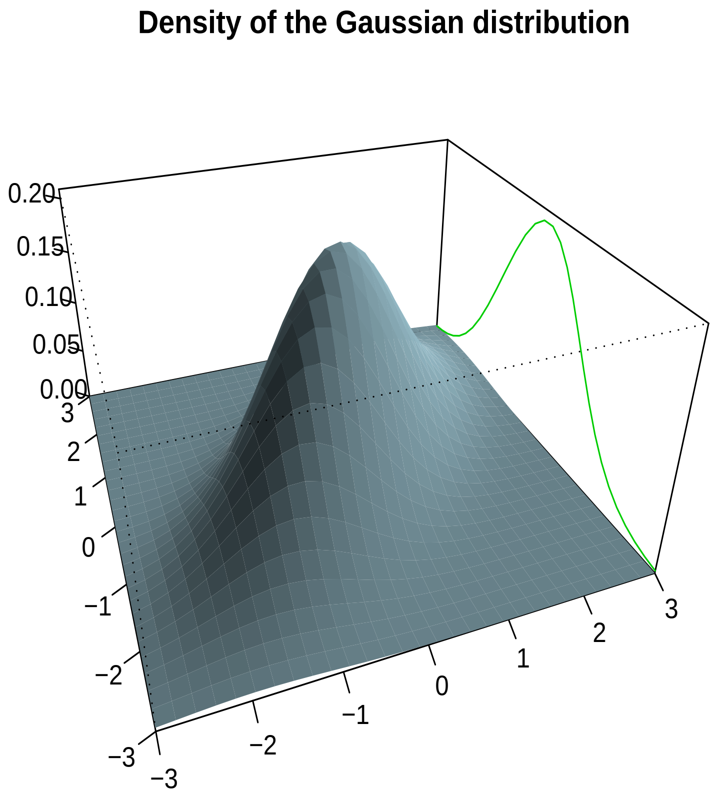
Proposition $(X_t)_{t\in\mathbb{N}}$ is a stationary AR(1) time series if and only if $\phi_1 \in (-1, 1)$.

Remark If $\phi_1 = 1$, $(X_t)_{t\in\mathbb{N}}$ is called a random walk.

Proposition If $(X_t)_{t\in\mathbb{N}}$ is a stationary AR(1) time series,

$$\rho(h) = \phi_1^h, \quad \forall h \in \mathbb{N}.$$

From univariate to multivariate models



Univariate gaussian distribution $\mathcal{N}(0,\sigma^2)$

$$\varphi(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{x^2}{2\sigma^2}\right),\text{ for all } x\in\mathbb{R}$$

Multivariate gaussian distribution $\mathcal{N}(0,\boldsymbol{\Sigma})$

$$\varphi(\boldsymbol{x})=\frac{1}{\sqrt{(2\pi)^d|\det\Sigma|}}\exp\left(-\frac{\boldsymbol{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{x}}{2}\right),$$

for all $\boldsymbol{x}\in\mathbb{R}^d$.

$\boldsymbol{X}=\boldsymbol{A}\boldsymbol{Z}$ where $\boldsymbol{A}\boldsymbol{A}'=\boldsymbol{\Sigma}$ and $\boldsymbol{Z}\sim\mathcal{N}(0,\mathbb{I})$

(geometric interpretation)

Vector (Gaussian) AutoRegressive processes $VAR(1)$

Definition A time series $(\boldsymbol{X}_t = (X_{1,t}, \cdots, X_{d,t}))_{t\in\mathbb{N}}$ with values in $\mathbb{R}^d$ is called a VAR(1) process if

$$\begin{cases} X_{1,t} = \phi_{1,1}X_{1,t-1} + \phi_{1,2}X_{2,t-1} + \cdots + \phi_{1,d}X_{d,t-1} + \varepsilon_{1,t} \\ X_{2,t} = \phi_{2,1}X_{1,t-1} + \phi_{2,2}X_{2,t-1} + \cdots + \phi_{2,d}X_{d,t-1} + \varepsilon_{2,t} \\ \cdots \\ X_{d,t} = \phi_{d,1}X_{1,t-1} + \phi_{d,2}X_{2,t-1} + \cdots + \phi_{d,d}X_{d,t-1} + \varepsilon_{d,t} \end{cases} \tag{2}$$

or equivalently

$$\underbrace{\begin{pmatrix} X_{1,t} \\ X_{2,t} \\ \vdots \\ X_{d,t} \end{pmatrix}}_{\boldsymbol{X}_t} = \underbrace{\begin{pmatrix} \phi_{1,1} & \phi_{1,2} & \cdots & \phi_{1,d} \\ \phi_{2,1} & \phi_{2,2} & \cdots & \phi_{2,d} \\ \vdots & \vdots & & \vdots \\ \phi_{d,1} & \phi_{d,2} & \cdots & \phi_{d,d} \end{pmatrix}}_{\boldsymbol{\Phi}} \underbrace{\begin{pmatrix} X_{1,t-1} \\ X_{2,t-1} \\ \vdots \\ X_{d,t-1} \end{pmatrix}}_{\boldsymbol{X}_{t-1}} + \underbrace{\begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \\ \vdots \\ \varepsilon_{d,t} \end{pmatrix}}_{\boldsymbol{\varepsilon}_t}$$

for all t, for some real-valued $d \times d$ matrix Φ, and some i.i.d. random vectors $\boldsymbol{\varepsilon}_t$ with values in $\mathbb{R}^d$.

It is common to assume that $\boldsymbol{\varepsilon}_t$ are independent variables, with a Gaussian distribution $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, with density

$$\varphi(\boldsymbol{\varepsilon}) = \frac{1}{\sqrt{(2\pi)^d |\det \Sigma|}} \exp\left(-\frac{\boldsymbol{\varepsilon}' \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon}}{2}\right), \quad \forall \boldsymbol{\varepsilon} \in \mathbb{R}^d.$$

Thus, independent means *time independent*, but can be dependent componentwise.

Note that we assume also that $\boldsymbol{\varepsilon}_t$ is independent of $\underline{\boldsymbol{X}}_{t-1}$, i.e. past observations $\boldsymbol{X}_0, \boldsymbol{X}_1, \cdots, \boldsymbol{X}_{t-1}$. Thus, $(\boldsymbol{\varepsilon}_t)_{t\in\mathbb{N}}$ is called the innovation process.

Definition A time series $(\boldsymbol{X}_t)_{t\in\mathbb{N}}$ is said to be (weakly) stationary if

- $\mathbb{E}(\boldsymbol{X}_t)$ is independent of t $(=: \boldsymbol{\mu})$
- $\text{cov}(\boldsymbol{X}_t, \boldsymbol{X}_{t-h})$ is independent of t $(=: \boldsymbol{\gamma}(h))$, called autocovariance matrix

Remark As a consequence, $\text{var}(\boldsymbol{X}_t) = \mathbb{E}([\boldsymbol{X}_t - \mathbb{E}(\boldsymbol{X}_t)]'[\boldsymbol{X}_t - \mathbb{E}(\boldsymbol{X}_t)])$ is independent of t (=: $\boldsymbol{\gamma}(0)$). Define finally the autocorrelation matrix,

$$\boldsymbol{\rho}(h) = \boldsymbol{\Delta}^{-1}\boldsymbol{\gamma}(h)\boldsymbol{\Delta}^{-1}, \text{ where } \boldsymbol{\Delta} = \text{diag}\left(\sqrt{\gamma_{i,i}(0)}\right).$$

Proposition $(\boldsymbol{X}_t)_{t\in\mathbb{N}}$ is a stationary AR(1) time series if and only if the d eignvalues of $\boldsymbol{\Phi}$ should have a norm lower than 1.

Proposition If $(\boldsymbol{X}_t)_{t\in\mathbb{N}}$ is a stationary AR(1) time series,

$$\boldsymbol{\rho}(h) = \boldsymbol{\Phi}^h, h \in \mathbb{N}.$$

Statistical inference for AR(1) time series

Consider a series of observations $X_1, \cdots, X_n$. The likelihood is the joint distribution of the vectors $\boldsymbol{\mathcal{X}} = (X_1, \cdots, X_n)$, which is not the product of marginal distribution, since consecutive observations are not independent ($\text{cov}(X_t, X_{t-h}) = \phi^h$). Nevertheless

$$\mathcal{L}(\phi, \sigma; (X_0, \boldsymbol{\mathcal{X}})) = \prod_{t=1}^{n} \pi_{\phi,\sigma}(X_t | X_{t-1})$$

where $\pi_{\phi,\sigma}(\cdot | X_{t-1})$ is a Gaussian density.

Maximum likelihood estimators are

$$(\widehat{\phi}, \widehat{\sigma}) \in \text{argmax} \log \mathcal{L}(\phi, \sigma; (X_0, \boldsymbol{\mathcal{X}}))$$

Poisson distribution - and process - for counts

N as a Poisson distribution is $\mathbb{P}(N=k)=e^{-\lambda}\dfrac{\lambda^k}{k!}$ where $k\in\mathbb{N}$.

If $N\sim\mathcal{P}(\lambda)$, then $\mathbb{E}(N)=\lambda$.

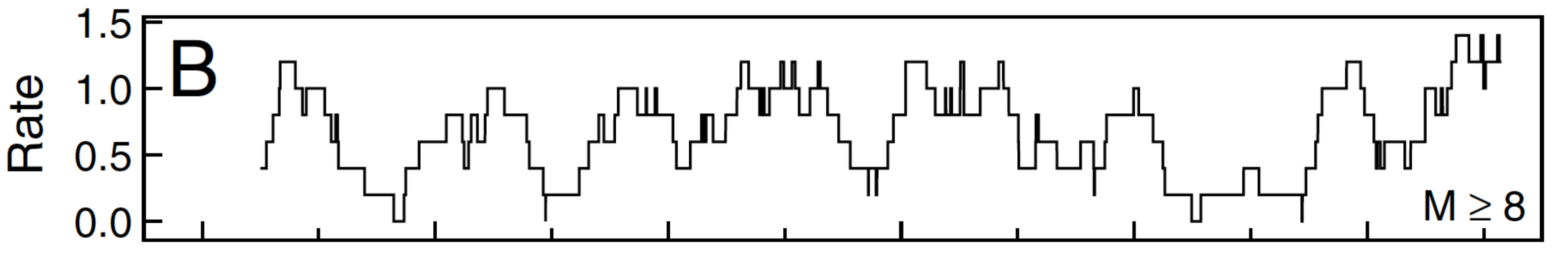
$(N_t)_{t\geq 0}$ is an homogeneous Poisson process, with parameter $\lambda\in\mathbb{R}^+$ if

- on time frame $[t,t+h]$, $(N_{t+h}-N_t)\sim\mathcal{P}(\lambda\cdot h)$
- on $[t_1,t_2]$ and $[t_3,t_4]$ counts are independent, if $0\leq t_1<t_2<t_3<t_4$, $(N_{t_2}-N_{t_1})\perp\!\!\!\perp(N_{t_4}-N_{t_3})$

Poisson processes and counting models

Earthquake count models are mostly based upon the Poisson process (see Utsu (1969), Gardner & Knopoff (1974), Lomnitz (1974), Kagan & Jackson (1991)), Cox process (self-exciting, cluster or branching processes, stress-release models (see Rathbun (2004) for a review), or Hidden Markov Models (HMM) (see Zucchini & MacDonald (2009) and Orfanogiannaki et al. (2010)).

See also Vere-Jones (2010) for a summary of statistical and stochastic models in seismology. Recently, Shearer & Starkb (2012) and Beroza (2012) rejected homogeneous Poisson model,



Thinning operator $\circ$

Steutel & van Harn (1979) defined a thinning operator as follows

Definition Define operator $\circ$ as

$$p \circ N = Y_1 + \cdots + Y_N \text{ if } N \neq 0, \text{ and } 0 \text{ otherwise},$$

where N is a random variable with values in $\mathbb{N}$, $p \in [0,1]$, and $Y_1, Y_2, \cdots$ are i.i.d. Bernoulli variables, independent of N, with $\mathbb{P}(Y_i = 1) = p = 1 - \mathbb{P}(Y_i = 0)$. Thus $p \circ N$ is a compound sum of i.i.d. Bernoulli variables.

Hence, given N, $p \circ N$ has a binomial distribution $\mathcal{B}(N,p)$.

Note that $p \circ (q \circ \boldsymbol{N}) \stackrel{\mathcal{L}}{=} [pq] \circ \boldsymbol{N}$ for all $p, q \in [0,1]$.

Further

$$\mathbb{E}(p \circ N) = p\mathbb{E}(N) \text{ and } \text{var}(p \circ N) = p^2 \text{var}(N) + p(1-p)\mathbb{E}(N).$$

(Poisson) Integer AutoRegressive processes $INAR(1)$

Based on that thinning operator, Al-Osh & Alzaid (1987) and McKenzie (1985) defined the integer autoregressive process of order 1 :

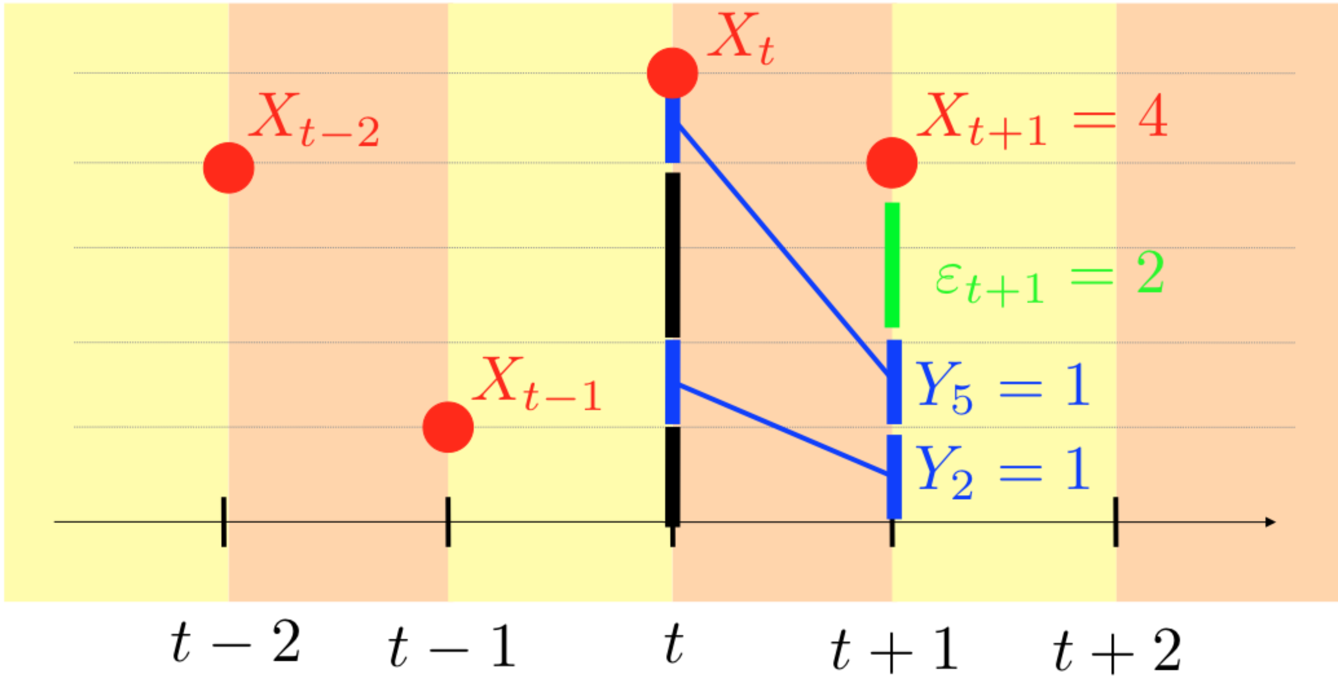
Definition A time series $(X_t)_{t\in\mathbb{N}}$ with values in $\mathbb{R}$ is called an INAR(1) process if

$$X_t = p \circ X_{t-1} + \varepsilon_t, \tag{3}$$

where (ε_t) is a sequence of i.i.d. integer valued random variables, i.e.

$$X_t = \sum_{i=1}^{X_{t-1}} Y_i + \varepsilon_t, \text{ where } Y_i's \text{ are } i.i.d.\ \mathcal{B}(p).$$

Such a process can be related to Galton-Watson processes with immigration, or physical branching model.



$$X_{t+1} = \sum_{i=1}^{X_t} Y_i + \varepsilon_{t+1},\ \text{where } Y_i' s \text{ are } i.i.d.\ \mathcal{B}(p)$$

Proposition $\mathbb{E}(X_t) = \dfrac{\mathbb{E}(\varepsilon_t)}{1-p}$, $\text{var}(X_t) = \gamma(0) = \dfrac{p\mathbb{E}(\varepsilon_t) + \text{var}(\varepsilon_t)}{1-p^2}$ and

$$\gamma(h) = \text{cov}(X_t, X_{t-h}) = p^h.$$

It is common to assume that ε_t are independent variables, with a Poisson distribution $\mathcal{P}(\lambda)$, with probability function

$$\mathbb{P}(\varepsilon_t = k) = e^{-\lambda}\frac{\lambda^k}{k!}, k \in \mathbb{N}.$$

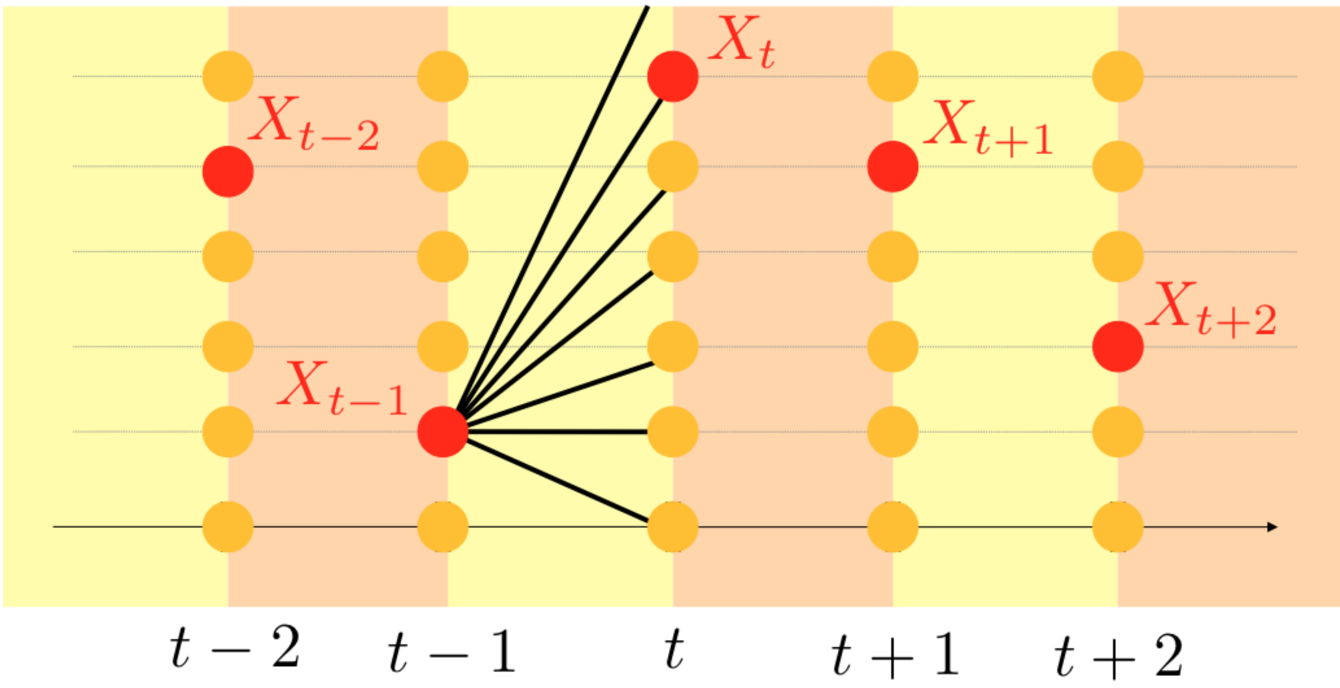
Proposition If (ε_t) are Poisson random variables, then (N_t) will also be a sequence of Poisson random variables.

Note that we assume also that ε_t is independent of $\underline{X}_{t-1}$, i.e. past observations $X_0, X_1, \cdots, X_{t-1}$. Thus, $(\varepsilon_t)_{t\in\mathbb{N}}$ is called the innovation process.

Proposition $(X_t)_{t\in\mathbb{N}}$ is a stationary INAR(1) time series if and only if $p \in [0,1)$.

Proposition If $(X_t)_{t\in\mathbb{N}}$ is a stationary INAR(1) time series, $(X_t)_{t\in\mathbb{N}}$ is an homogeneous Markov chain.

$$\pi(x_t, x_{t-1}) \quad = \quad \mathbb{P}(X_t = x_t | X_{t-1} = x_{t-1}) = \sum_{k=0}^{x_t} \underbrace{\mathbb{P}\left(\sum_{i=1}^{x_{t-1}} Y_i = x_t - k\right)}_{\text{Binomial}} \cdot \underbrace{\mathbb{P}(\varepsilon = k)}_{\text{Poisson}}.$$



Inference of Integer AutoRegressive processes $INAR(1)$

Consider a Poisson INAR(1) process, then the likelihood is

$$\mathcal{L}(p,\lambda;X_0,\boldsymbol{\mathcal{X}})=\left[\prod_{t=1}^{n}f_t(X_t)\right]\cdot\frac{\lambda^{X_0}}{(1-p)^{X_0}X_0!}\exp\left(-\frac{\lambda}{1-p}\right)$$

where

$$f_t(y)=\exp(-\lambda)\sum_{i=0}^{\min\{X_t,X_{t-1}\}}\frac{\lambda^{y-i}}{(y-i)!}\binom{Y_{t-1}}{i}p^i(1-p)^{Y_{t-1}-y},\ \text{for } t=1,\cdots,n.$$

Maximum likelihood estimators are

$$(\widehat{p},\widehat{\lambda})\in\text{argmax}\log\mathcal{L}(p,\lambda;(X_0,\boldsymbol{\mathcal{X}}))$$

Multivariate Integer Autoregressive processes $MINAR(1)$

Let $\boldsymbol{N_t} := (N_{1,t}, \cdots, N_{d,t})$, denote a multivariate vector of counts.

Definition Let $\boldsymbol{P} := [p_{i,j}]$ be a $d \times d$ matrix with entries in $[0,1]$. If $\boldsymbol{N} = (N_1, \cdots, N_d)$ is a random vector with values in $\mathbb{N}^d$, then $\boldsymbol{P} \circ \boldsymbol{N}$ is a d-dimensional random vector, with i-th component

$$[\boldsymbol{P} \circ \boldsymbol{N}]_i = \sum_{j=1}^{d} p_{i,j} \circ N_j,$$

for all $i = 1, \cdots, d$, where all counting variates Y in $p_{i,j} \circ N_j$'s are assumed to be independent.

Note that $\boldsymbol{P} \circ (\boldsymbol{Q} \circ \boldsymbol{N}) \overset{\mathcal{L}}{=} [\boldsymbol{P}\boldsymbol{Q}] \circ \boldsymbol{N}$.

Further, $\mathbb{E}\left(\boldsymbol{P} \circ \boldsymbol{N}\right) = \boldsymbol{P}\mathbb{E}(\boldsymbol{N})$, and

$$\mathbb{E}\left((\boldsymbol{P} \circ \boldsymbol{N})(\boldsymbol{P} \circ \boldsymbol{N})'\right) = \boldsymbol{P}\mathbb{E}(\boldsymbol{N}\boldsymbol{N}')\boldsymbol{P}' + \Delta,$$

with $\Delta := \text{diag}(\boldsymbol{V}\mathbb{E}(\boldsymbol{N}))$ where $\boldsymbol{V}$ is the $d \times d$ matrix with entries $p_{i,j}(1 - p_{i,j})$.

Definition A time series $(\boldsymbol{X}_t)$ with values in $\mathbb{N}^d$ is called a d-variate MINAR(1) process if

$$\boldsymbol{X}_t = \boldsymbol{P} \circ \boldsymbol{X}_{t-1} + \boldsymbol{\varepsilon}_t \tag{4}$$

for all t, for some $d \times d$ matrix $\boldsymbol{P}$ with entries in $[0,1]$, and some i.i.d. random vectors $\boldsymbol{\varepsilon}_t$ with values in $\mathbb{N}^d$.

$(\boldsymbol{X}_t)$ is a Markov chain with states in $\mathbb{N}^d$ with transition probabilities

$$\pi(\boldsymbol{x}_t, \boldsymbol{x}_{t-1}) = \mathbb{P}(\boldsymbol{X}_t = \boldsymbol{x}_t | \boldsymbol{X}_{t-1} = \boldsymbol{x}_{t-1}) \tag{5}$$

satisfying

$$\pi(\boldsymbol{x}_t, \boldsymbol{x}_{t-1}) = \sum_{\boldsymbol{k}=0}^{\boldsymbol{x}_t} \mathbb{P}(\boldsymbol{P} \circ \boldsymbol{x}_{t-1} = \boldsymbol{x}_t - \boldsymbol{k}) \cdot \mathbb{P}(\boldsymbol{\varepsilon} = \boldsymbol{k}).$$

Parameter inference for $MINAR(1)$

Proposition Let $(\boldsymbol{X}_t)$ be a d-variate MINAR(1) process satisfying stationary conditions, as well as technical assumptions (called C1-C6 in Franke & Subba Rao (1993)), then the conditional maximum likelihood estimate $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta} = (\boldsymbol{P}, \boldsymbol{\Lambda})$ is asymptotically normal,

$$\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, \Sigma^{-1}(\boldsymbol{\theta})), \text{ as } n \to \infty.$$

Further,

$$2[\log \mathcal{L}(\underline{\boldsymbol{N}}, \widehat{\boldsymbol{\theta}}|\boldsymbol{N}_0) - \log \mathcal{L}(\underline{\boldsymbol{N}}, \boldsymbol{\theta}|\boldsymbol{N}_0)] \xrightarrow{\mathcal{L}} \chi^2(d^2 + \dim(\boldsymbol{\lambda})), \text{ as } n \to \infty.$$

Granger causality with $BINAR(1)$

$$\underbrace{\begin{pmatrix} X_{1,t} \\ X_{2,t} \end{pmatrix}}_{\boldsymbol{X}_t} = \underbrace{\begin{pmatrix} p_{1,1} & p_{1,2} \\ p_{2,1} & p_{2,2} \end{pmatrix}}_{\boldsymbol{P}} \circ \underbrace{\begin{pmatrix} X_{1,t-1} \\ X_{2,t-1} \end{pmatrix}}_{\boldsymbol{X}_{t-1}} + \underbrace{\begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix}}_{\boldsymbol{\varepsilon}_t}, \text{ with var} \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} = \begin{pmatrix} \lambda_1 & \varphi \\ \varphi & \lambda_2 \end{pmatrix}$$

Granger causality with $BINAR(1)$

1. (X_1) and (X_2) are instantaneously related if $\boldsymbol{\varepsilon}$ is a noncorrelated noise,

$$\underbrace{\begin{pmatrix} X_{1,t} \\ X_{2,t} \end{pmatrix}}_{\boldsymbol{X}_t} = \underbrace{\begin{pmatrix} p_{1,1} & p_{1,2} \\ p_{2,1} & p_{2,2} \end{pmatrix}}_{\boldsymbol{P}} \circ \underbrace{\begin{pmatrix} X_{1,t-1} \\ X_{2,t-1} \end{pmatrix}}_{\boldsymbol{X}_{t-1}} + \underbrace{\begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix}}_{\boldsymbol{\varepsilon}_t}, \text{ with var} \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} = \begin{pmatrix} \lambda_1 & \star \\ \star & \lambda_2 \end{pmatrix}$$

Granger causality with $BINAR(1)$

2. (X_1) and (X_2) are independent, $(X_1)\perp(X_2)$ if $\boldsymbol{P}$ is diagonal, i.e. $p_{1,2} = p_{2,1} = 0$, and ε_1 and ε_2 are independent,

$$\underbrace{\begin{pmatrix} X_{1,t} \\ X_{2,t} \end{pmatrix}}_{\boldsymbol{X}_t} = \underbrace{\begin{pmatrix} p_{1,1} & 0 \\ 0 & p_{2,2} \end{pmatrix}}_{\boldsymbol{P}} \circ \underbrace{\begin{pmatrix} X_{1,t-1} \\ X_{2,t-1} \end{pmatrix}}_{\boldsymbol{X}_{t-1}} + \underbrace{\begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix}}_{\boldsymbol{\varepsilon}_t}, \text{ with var} \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$$

Granger causality with $BINAR(1)$

3. (N_1) causes (N_2) but (N_2) does not cause (X_1), $(X_1)\rightarrow(X_2)$, if $\boldsymbol{P}$ is a lower triangle matrix, i.e. $p_{2,1}\neq 0$ while $p_{1,2}=0$,

$$\underbrace{\begin{pmatrix} X_{1,t} \\ X_{2,t} \end{pmatrix}}_{\boldsymbol{X}_t} = \underbrace{\begin{pmatrix} p_{1,1} & 0 \\ \star & p_{2,2} \end{pmatrix}}_{\boldsymbol{P}} \circ \underbrace{\begin{pmatrix} X_{1,t-1} \\ X_{2,t-1} \end{pmatrix}}_{\boldsymbol{X}_{t-1}} + \underbrace{\begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix}}_{\boldsymbol{\varepsilon}_t}, \text{ with var} \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} = \begin{pmatrix} \lambda_1 & \varphi \\ \varphi & \lambda_2 \end{pmatrix}$$

Granger causality with $BINAR(1)$

4. (N_2) causes (N_1) but $(N_{1,t})$ does not cause (N_2), $(N_1)\leftarrow(N_{2,t})$, if $\boldsymbol{P}$ is a upper triangle matrix, i.e. $p_{1,2}\neq 0$ while $p_{2,1}=0$,

$$\underbrace{\begin{pmatrix}X_{1,t}\\X_{2,t}\end{pmatrix}}_{\boldsymbol{X}_t}=\underbrace{\begin{pmatrix}p_{1,1}&\star\\0&p_{2,2}\end{pmatrix}}_{\boldsymbol{P}}\circ\underbrace{\begin{pmatrix}X_{1,t-1}\\X_{2,t-1}\end{pmatrix}}_{\boldsymbol{X}_{t-1}}+\underbrace{\begin{pmatrix}\varepsilon_{1,t}\\\varepsilon_{2,t}\end{pmatrix}}_{\boldsymbol{\varepsilon}_t},\text{ with var}\begin{pmatrix}\varepsilon_{1,t}\\\varepsilon_{2,t}\end{pmatrix}=\begin{pmatrix}\lambda_1&\varphi\\\varphi&\lambda_2\end{pmatrix}$$

Granger causality with $BINAR(1)$

5. (N_1) causes (N_2) and conversely, i.e. a feedback effect $(N_1)\leftrightarrow(N_2)$, if $\boldsymbol{P}$ is a full matrix, i.e. $p_{1,2}, p_{2,1} \neq 0$

$$\underbrace{\begin{pmatrix} X_{1,t} \\ X_{2,t} \end{pmatrix}}_{\boldsymbol{X}_t} = \underbrace{\begin{pmatrix} p_{1,1} & \star \\ \star & p_{2,2} \end{pmatrix}}_{\boldsymbol{P}} \circ \underbrace{\begin{pmatrix} X_{1,t-1} \\ X_{2,t-1} \end{pmatrix}}_{\boldsymbol{X}_{t-1}} + \underbrace{\begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix}}_{\boldsymbol{\varepsilon}_t}, \text{ with var} \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} = \begin{pmatrix} \lambda_1 & \varphi \\ \varphi & \lambda_2 \end{pmatrix}$$

Bivariate Poisson $BINAR(1)$

A classical distribution for $\boldsymbol{\varepsilon}_t$ is the bivariate Poisson distribution, with one common shock, i.e.

$$\begin{cases} \varepsilon_{1,t} = M_{1,t} + M_{0,t} \\ \varepsilon_{2,t} = M_{2,t} + M_{0,t} \end{cases}$$

where $M_{1,t}$, $M_{2,t}$ and $M_{0,t}$ are independent Poisson variates, with parameters $\lambda_1 - \varphi$, $\lambda_2 - \varphi$ and φ, respectively. In that case, $\boldsymbol{\varepsilon}_t = (\varepsilon_{1,t}, \varepsilon_{2,t})$ has joint probability function

$$e^{-[\lambda_1+\lambda_2-\varphi]} \frac{(\lambda_1-\varphi)^{k_1}}{k_1!} \frac{(\lambda_2-\varphi)^{k_2}}{k_2!} \sum_{i=0}^{\min\{k_1,k_2\}} \binom{k_1}{i} \binom{k_2}{i} i! \left(\frac{\varphi}{[\lambda_1-\varphi][\lambda_2-\varphi]} \right)$$

with $\lambda_1, \lambda_2 > 0$, $\varphi \in [0, \min\{\lambda_1, \lambda_2\}]$.

$$\boldsymbol{\lambda} = \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} \text{ and } \boldsymbol{\Lambda} = \begin{pmatrix} \lambda_1 & \varphi \\ \varphi & \lambda_2 \end{pmatrix}$$

Bivariate Poisson $BINAR(1)$ and Granger causality

For instantaneous causality, we test

$$H_0 : \varphi = 0 \text{ against } H_1 : \varphi \neq 0$$

Proposition Let $\widehat{\boldsymbol{\lambda}}$ denote the conditional maximum likelihood estimate of $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \varphi)$ in the non-constrained MINAR(1) model, and $\boldsymbol{\lambda}^{\perp}$ denote the conditional maximum likelihood estimate of $\boldsymbol{\lambda}^{\perp} = (\lambda_1, \lambda_2, 0)$ in the constrained model (when innovation has independent margins), then under suitable conditions,

$$2[\log \mathcal{L}(\underline{\boldsymbol{N}}, \widehat{\boldsymbol{\lambda}}|\boldsymbol{N}_0) - \log \mathcal{L}(\underline{\boldsymbol{N}}, \widehat{\boldsymbol{\lambda}}^{\perp}|\boldsymbol{N}_0)] \xrightarrow{\mathcal{L}} \chi^2(1), \text{ as } n \to \infty, \text{ under } H_0.$$

Bivariate Poisson $BINAR(1)$ and Granger causality

For lagged causality, we test

$$H_0 : \boldsymbol{P} \in \mathcal{P} \text{ against } H_1 : \boldsymbol{P} \notin \mathcal{P},$$

where $\mathcal{P}$ is a set of constrained shaped matrix, e.g. $\mathcal{P}$ is the set of $d \times d$ diagonal matrices for lagged independence, or a set of block triangular matrices for lagged causality.

Proposition Let $\widehat{\boldsymbol{P}}$ denote the conditional maximum likelihood estimate of $\boldsymbol{P}$ in the non-constrained MINAR(1) model, and $\widehat{\boldsymbol{P}}^c$ denote the conditional maximum likelihood estimate of $\boldsymbol{P}$ in the constrained model, then under suitable conditions,

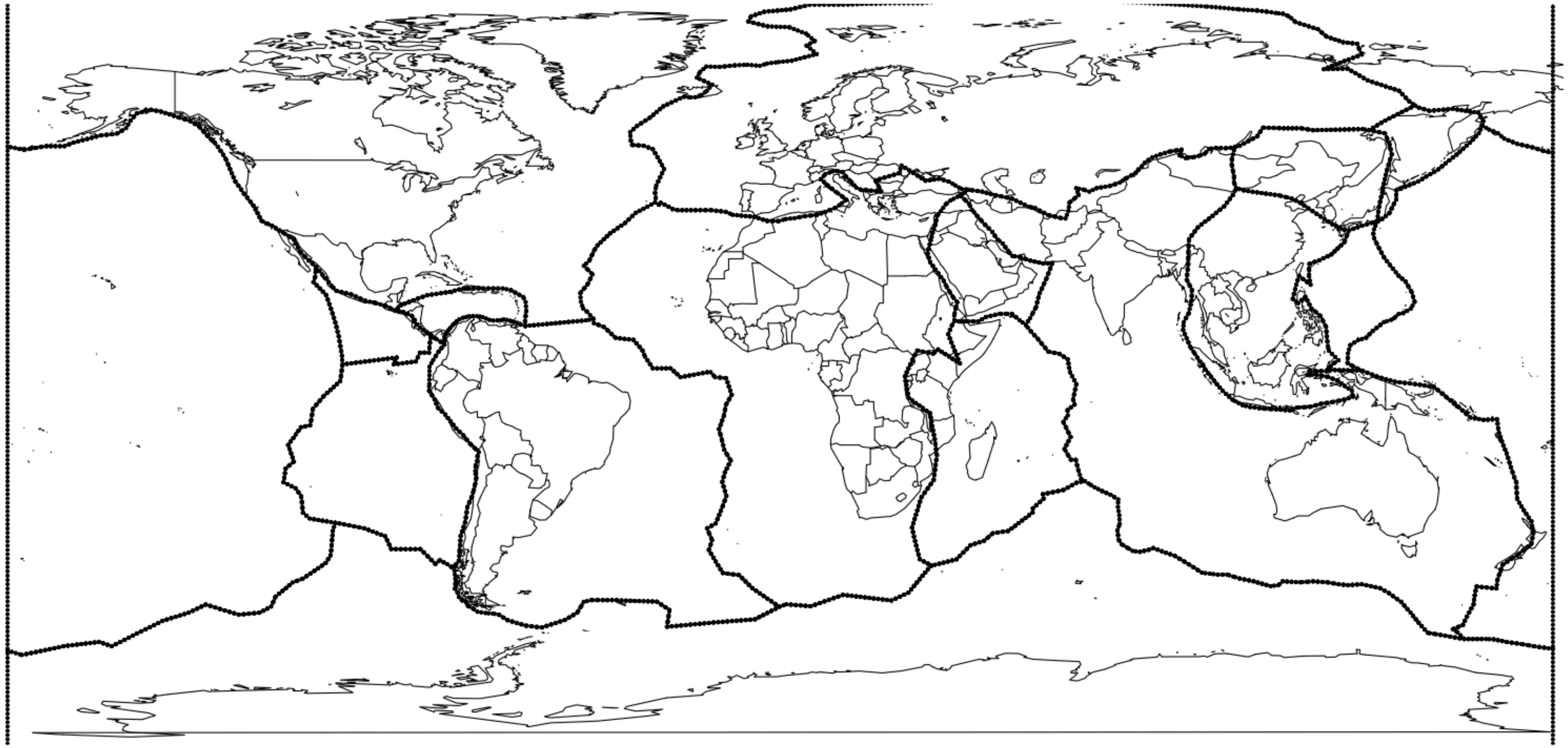
$$2[\log \mathcal{L}(\underline{\boldsymbol{N}}, \widehat{\boldsymbol{P}}|\boldsymbol{N}_0) - \log \mathcal{L}(\underline{\boldsymbol{N}}, \widehat{\boldsymbol{P}}^c|\boldsymbol{N}_0)] \xrightarrow{\mathcal{L}} \chi^2(d^2 - \dim(\mathcal{P})), \text{ as } n \to \infty, \text{ under } H_0.$$

Example Testing $(N_{1,t}) \leftarrow (N_{2,t})$ is testing whether $p_{1,2} = 0$, or not.

Autocorrelation of $MINAR(1)$ processes

Proposition Consider a MINAR(1) process with representation $\boldsymbol{X}_t = \boldsymbol{P} \circ \boldsymbol{X}_{t-1} + \boldsymbol{\varepsilon}_t$, where $(\boldsymbol{\varepsilon}_t)$ is the innovation process, with $\boldsymbol{\lambda} := \mathbb{E}(\boldsymbol{\varepsilon}_t)$ and $\boldsymbol{\Lambda} := \text{var}(\boldsymbol{\varepsilon}_t)$. Let $\boldsymbol{\mu} := \mathbb{E}(\boldsymbol{X}_t)$ and $\boldsymbol{\gamma}(h) := \text{cov}(\boldsymbol{X}_t, \boldsymbol{X}_{t-h})$. Then $\boldsymbol{\mu} = [\mathbb{I} - \boldsymbol{P}]^{-1}\boldsymbol{\lambda}$ and for all $h \in \mathbb{Z}$, $\boldsymbol{\gamma}(h) = \boldsymbol{P}^h\boldsymbol{\gamma}(0)$ with $\boldsymbol{\gamma}(0)$ solution of $\boldsymbol{\gamma}(0) = \boldsymbol{P}\boldsymbol{\gamma}(0)\boldsymbol{P}' + (\boldsymbol{\Delta} + \boldsymbol{\Lambda})$.

Multivariate models ?



The dataset, and stationarity issues

We work with 16 (17) tectonic plates,

– Japan is at the limit of 4 tectonic plates (Pacific, Okhotsk, Philippine and Amur),
– California is at the limit of the Pacific, North American and Juan de Fuca plates.

Data were extracted from the Advanced National Seismic System database (ANSS) `http://www.ncedc.org/cnss/catalog-search.html`

– 1965-2011 for magnitude $M > 5$ earthquakes (70,000 events) ;
– 1992-2011 for $M > 6$ earthquakes (3,000 events) ;
– To count the number of earthquakes, used time ranges of 3, 6, 12, 24, 36 and 48 hours ;
– Approximately 8,500 to 135,000 periods of observation ;

Multivariate models : comparing dynamics

$$\begin{pmatrix} X_{1,t} \\ X_{2,t} \end{pmatrix} = \begin{pmatrix} p_{1,1} & p_{1,2} \\ p_{2,1} & p_{2,2} \end{pmatrix} \circ \begin{pmatrix} X_{1,t-1} \\ X_{2,t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} \text{ with var} \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} = \begin{pmatrix} \lambda_1 & \varphi \\ \varphi & \lambda_2 \end{pmatrix}$$

Complete model, with full dependence

Multivariate models : comparing dynamics

$$\begin{pmatrix} X_{1,t} \\ X_{2,t} \end{pmatrix} = \begin{pmatrix} p_{1,1} & 0 \\ 0 & p_{2,2} \end{pmatrix} \circ \begin{pmatrix} X_{1,t-1} \\ X_{2,t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} \text{ with var} \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} = \begin{pmatrix} \lambda_1 & \varphi \\ \varphi & \lambda_2 \end{pmatrix}$$

Partial model, with diagonal thinning matrix, no-crossed lag correlation

Multivariate models : comparing dynamics

$$\begin{pmatrix} X_{1,t} \\ X_{2,t} \end{pmatrix} = \begin{pmatrix} p_{1,1} & 0 \\ 0 & p_{2,2} \end{pmatrix} \circ \begin{pmatrix} X_{1,t-1} \\ X_{2,t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} \text{ with var} \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$$

Two independent INAR processes

Multivariate models : comparing dynamics

$$\begin{pmatrix} X_{1,t} \\ X_{2,t} \end{pmatrix} = \begin{pmatrix} p_{1,1} & 0 \\ 0 & p_{2,2} \end{pmatrix} \circ \begin{pmatrix} X_{1,t-1} \\ X_{2,t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} \text{ with var} \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$$

Two independent INAR processes

Multivariate models : comparing dynamics

$$\begin{pmatrix} X_{1,t} \\ X_{2,t} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \circ \begin{pmatrix} X_{1,t-1} \\ X_{2,t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} \text{ with var} \begin{pmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{pmatrix} = \begin{pmatrix} \lambda_1 & \varphi \\ \varphi & \lambda_2 \end{pmatrix}$$

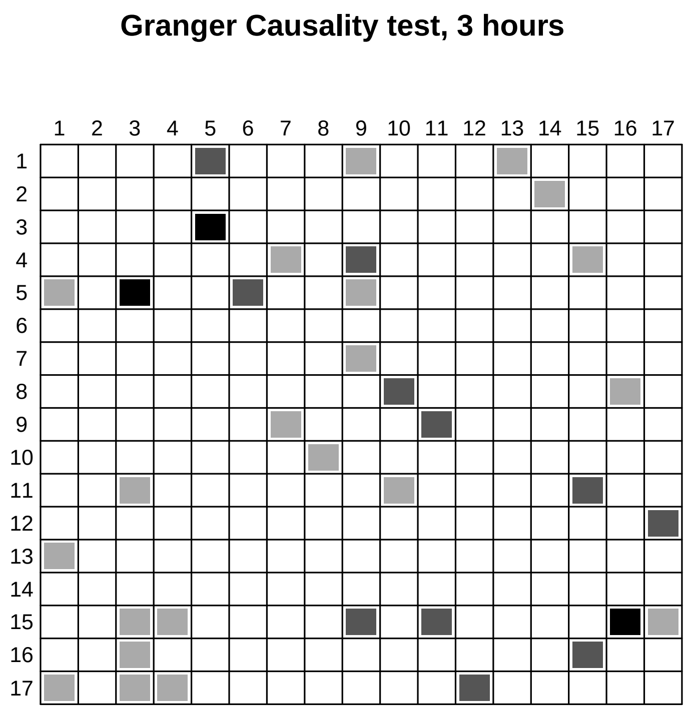
Two (possibly dependent) Poisson processes

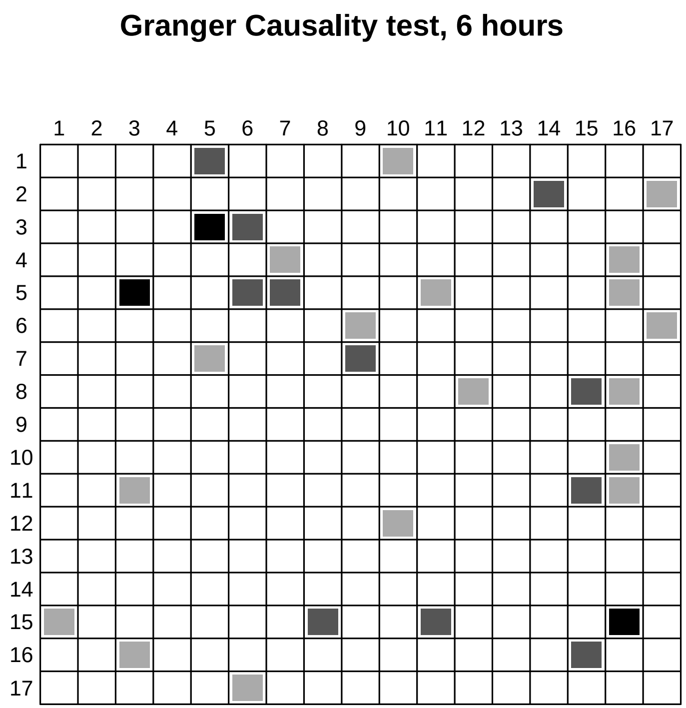
Multivariate models : tectonic plates interactions

- For all pairs of tectonic plates, at all frequencies, autoregression **in time** is important (very high statistical significance) ;
  - Long sequence of zeros, then mainshocks and aftershocks ;
  - Rate of aftershocks decreases exponentially over time (Omori's law) ;
- For 7-13% of pairs of tectonic plates, diagonal BINAR has significant better fit than independent INARs ;
  - Contribution of dependence in noise ;
  - Spatial contagion of order 0 (within h hours) ;
  - Contiguous tectonic plates ;
- For 7-9% of pairs of tectonic plates, proposed BINAR has significant better fit than diagonal BINAR ;
  - Contribution of spatial contagion of order 1 (in time interval $[h, 2h]$) ;
  - Contiguous tectonic plates ;
- for approximately 90%, there is no significant spatial contagion for $M > 5$ earthquakes

Granger causality $N_1 \rightarrow N_2$ or $N_1 \leftarrow N_2$

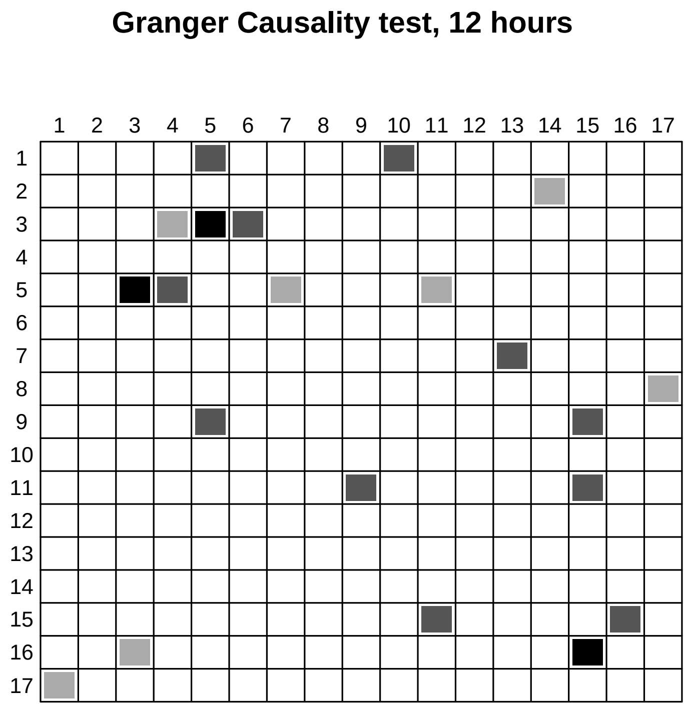
1. North American Plate, 2. Eurasian Plate,3. Okhotsk Plate, 4. Pacific Plate (East), 5. Pacific Plate (West), 6. Amur Plate, 7. Indo-Australian Plate, 8. African Plate, 9. Indo-Chinese Plate, 10. Arabian Plate, 11. Philippine Plate, 12. Coca Plate, 13. Caribbean Plate, 14. Somali Plate, 15. South American Plate, 16. Nasca Plate, 17. Antarctic Plate

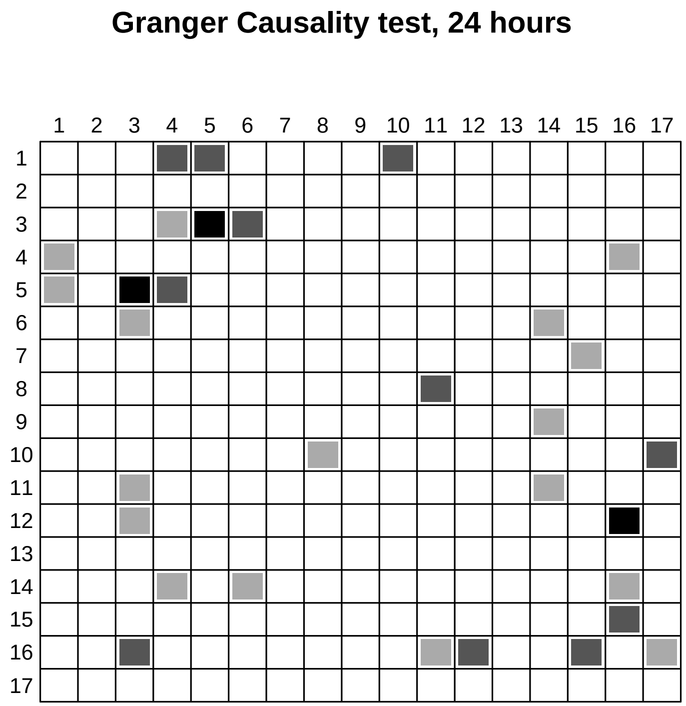




Granger causality $N_1 \rightarrow N_2$ or $N_1 \leftarrow N_2$

1. North American Plate, 2. Eurasian Plate,3. Okhotsk Plate, 4. Pacific Plate (East), 5. Pacific Plate (West), 6. Amur Plate, 7. Indo-Australian Plate, 8. African Plate, 9. Indo-Chinese Plate, 10. Arabian Plate, 11. Philippine Plate, 12. Coca Plate, 13. Caribbean Plate, 14. Somali Plate, 15. South American Plate, 16. Nasca Plate, 17. Antarctic Plate

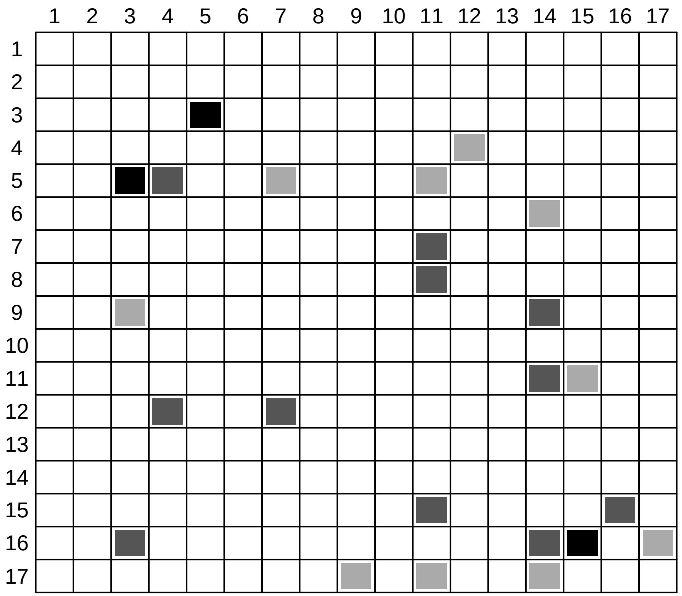




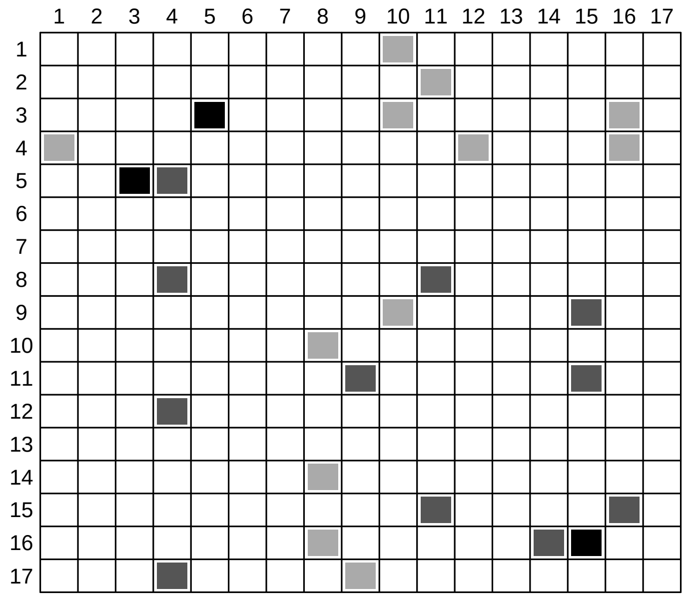
Granger causality $N_1 \rightarrow N_2$ or $N_1 \leftarrow N_2$

1. North American Plate, 2. Eurasian Plate,3. Okhotsk Plate, 4. Pacific Plate (East), 5. Pacific Plate (West), 6. Amur Plate, 7. Indo-Australian Plate, 8. African Plate, 9. Indo-Chinese Plate, 10. Arabian Plate, 11. Philippine Plate, 12. Coca Plate, 13. Caribbean Plate, 14. Somali Plate, 15. South American Plate, 16. Nasca Plate, 17. Antarctic Plate

Granger Causality test, 36 hours



Granger Causality test, 48 hours



Multivariate models : frequency versus magnitude

$$X_{1,t}=\sum_{i=1}\mathbf{1}(T_i\in[t,t+1),M_i\leq s)\text{ and }X_{2,t}=\sum_{i=1}\mathbf{1}(T_i\in[t,t+1),M_i>s)$$

Here we work on two sets of data : medium-size earthquakes ($M\in(5,6)$) and large-size earthquakes ($M>6$).

- Investigate direction of relationship (which one causes the other, or both) ;
- Pairs of tectonic plates :
  - Uni-directional causality : most common for contiguous plates (North American causes West Pacific, Okhotsk causes Amur) ;
  - Bi-directional causality : Okhotsk and West Pacific, South American and Nasca for example ;
- Foreshocks and aftershocks :
  - Aftershocks much more significant than foreshocks (as expected) ;
  - Foreshocks announce arrival of larger-size earthquakes ;
  - Foreshocks significant for Okhotsk, West Pacific, Indo-Australian, Indo-Chinese, Philippine, South American ;

Risk management issues

- Interested in computing $\mathbb{P}\left(\sum_{t=1}^{T}(N_{1,t}+N_{2,t})\geq n \middle| \mathcal{F}_0\right)$ for various values of T (time horizons) and n (tail risk measure) ;
  - Total number of earthquakes on a set of two tectonic plates ;
- 100 000 simulated paths of diagonal and proposed BINAR models ;
  - Use estimated parameters of both models ;
  - Pair : Okhotsk and West Pacific ;
- Scenario : on a 12-hour period, 23 earthquakes on Okhotsk and 46 earthquakes on West Pacific (second half of March 10th, 2011) ;

Diagonal model				
n / days	**1 day**	**3 days**	**7 days**	**14 days**
5	0.9680	0.9869	0.9978	0.9999
10	0.5650	0.7207	0.8972	0.9884
15	0.1027	0.2270	0.4978	0.8548
20	0.0067	0.0277	0.1308	0.4997
Proposed model				
n / days	**1 day**	**3 days**	**7 days**	**14 days**
5	0.9946	0.9977	0.9997	1.0000
10	0.8344	0.9064	0.9712	0.9970
15	0.3638	0.5288	0.7548	0.9479
20	0.0671	0.1573	0.3616	0.7256

Al-Osh, M.A. & A.A. Alzaid (1987) "First-order integer-valued autoregressive process", *Journal of Time Series Analysis* **8**, 261-275.

Beroza, G.C. (2012) How many great earthquakes should we expect ? *PNAS*, 109, 651-652.

Cabrera, B., B. (2006). Pricing catastrophic bonds for earthquakes in Mexico. *Master thesis*, Humboldt-Universität zu Berlin.

Dion, J.-P., G. Gauthier & A. Latour (1995), "Branching processes with immigration and integer-valued time series", *Serdica Mathematical Journal* **21**, 123-136.

Du, J.-G. & Y. Li (1991), "The integer-valued autoregressive (INAR(p)) model", *Journal of Time Series Analysis* **12**, 129-142.

Ferland, R.A., Latour, A. & Oraichi, D. (2006), Integer-valued GARCH process. *Journal of Time Series Analysis* **27**, 923-942.

Fokianos, K. (2011). Count time series models. *to appear in* Handbook of Time Series Analysis.

Franke, J. & Subba Rao, T. (1993). Multivariae first-order integer-valued autoregressions. *Forschung Universität Kaiserslautern*, **95**.

Gourieroux, C. & J. Jasiak (2004) "Heterogeneous INAR(1) model with application to car insurance", *Insurance : Mathematics & Economics* **34**, 177-192.

Gardner, J.K. & I. Knopoff (1974) "Is the sequence of earthquakes in Southern California, with aftershocks removed, Poissonean ?" *Bulletin of the Seismological Society of America* **64**, 1363-1367.

Joe, H. (1996) Time series models with univariate margins in the convolution-closed infinitely divisible class. *Journal of Time Series Analysis* **104**, 117-133.

Johnson, N.L, Kotz, S. & Balakrishnan, N. (1997). Discrete Multivariate Distributions. Wiley Interscience.

Kagan, Y.Y. & D.D. Jackson (1991) Long-term earthquake clustering, *Geophysical Journal International* **104**, 117-133.

Kocherlakota, S. & Kocherlakota K. (1992). Bivariate Discrete Distributions. CRC Press.

Latour, A. (1997) "The multivariate GINAR(p) process", *Advances in Applied Probability* **29**, 228-248.

Latour, A. (1998) Existence and stochastic structure of a non-negative integer-valued autoregressive process, *Journal of Time Series Analysis* **19**, 439-455.

Lomnitz, C. (1974) "Global Tectonic and Earthquake Risk", Elsevier.

Lütkepohl, H. (2005). New Introduction to Multiple Time Series Analysis. Springer Verlag.

Mahamunulu, D. M. (1967). A note on regression in the multivariate Poisson distribution. *Journal of the American Statistical Association*, **62**, 25 1-258.

McKenzie, E. (1985) Some simple models for discrete variate time series, *Water Resources Bulletin* **21**, 645-650.

Orfanogiannaki, K., D. Karlis & G.A. Papadopoulos (2010) "Identifying Seismicity Levels via Poisson Hidden Markov Models", *Pure and Applied Geophysics* **167**, 919-931.

Parsons, T., & A.A. Velasco (2011) "Absence of remotely triggered large earthquakes beyond the mainshock region", Nature Geoscience, March 27th, 2011.

Pedeli, X. & Karlis, D. (2011) "A bivariate Poisson INAR(1) model with application", *Statistical Modelling* **11**, 325-349.

Pedeli, X. & Karlis, D. (2011). "On estimation for the bivariate Poisson INAR process", To appear in *Communications in Statistics, Simulation and Computation.*

Rathbun, S.L. (2004), "Seismological modeling", *Encyclopedia of Environmetrics.*

Rosenblatt, M. (1971). Markov processes, Structure and Asymptotic Behavior. Springer-Verlag.

Shearer, P.M. & Stark, P.B (2012). Global risk of big earthquakes has not recently increased. *PNAS*, 109, 717-721.

Steutel, F. & K. van Harn (1979) "Discrete analogues of self-decomposability and stability", *Annals of Probability* **7**, 893-899.

Utsu, T. (1969) "Aftershocks and earthquake statistics (I) - Some parameters which characterize an aftershock sequence and their interrelations", Journal of the Faculty of Science of Hokkaido University, 121-195.

Vere-Jones, D. (2010), "Foundations of Statistical Seismology", *Pure and Applied Geophysics* **167**, 645-653.

Weiβ, C.H. (2008) "Thinning operations for modeling time series of counts : a survey", *Advances in Statistical Analysis* **92**, 319-341.

Zucchini, W. & I.L. MacDonald (2009) "Hidden Markov Models for Time Series : An Introduction Using R", 2nd edition, Chapman & Hall

The article can be downloaded from http ://arxiv.org/abs/1112.0929