European Geosciences Union - General Assembly 2011
3-8 April 2011 - Vienna

Session: SM5.2/NH4.7 Time-dependent earthquake occurrence and seismic hazard: physics and statistics

Statistical tests for the retrospective detection of space-time clusters of seismic events

R. Rotondi E. Varini

CNR - Istituto di Matematica Applicata e Tecnologie Informatiche, Milano, Italy

In the literature:
"The major feature of earthquake occurrence is time-space clustering, both short and long term." (Kagan, Jackson (2000) *GJI*, 143, p. 443)

"A class of time-dependent (forecasting) models is based on the long-term space-time clustering of earthquakes observed in historical catalogues" (International Commission on Earthquake Forecasting for Civil Protection, *Operational Earthquake Forecasting: State of Knowledge and Guidelines for Utilization*)

" ... worldwide M_s 7.0+ earthquakes tend to cluster in time and space, with features similar to smaller events." (Lombardi, Marzocchi (2007) *JGR*, 112, B02303)

Aim: quantify these sentences in Italian tectonic situation

Outline of the exercise:

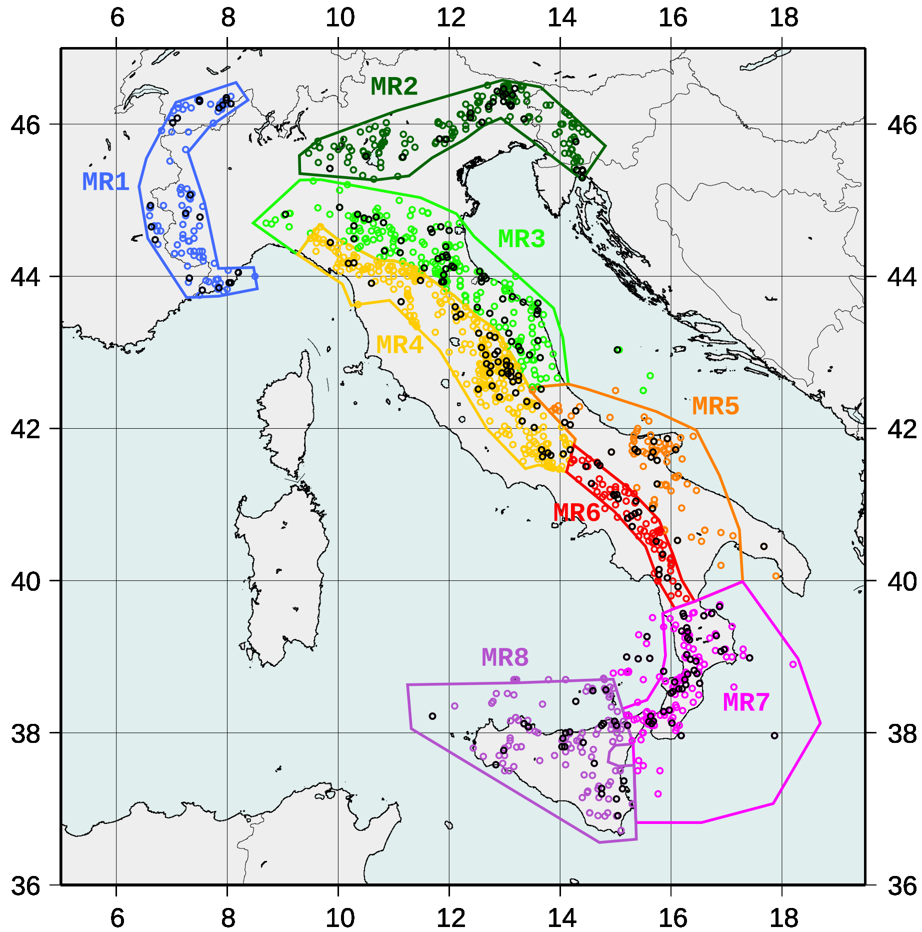
Data sets: 8 macro-regions, CPTI04, since 1600, $M_w \geq 4.5,\ 5.3,\ 6.0$

Statistical tests

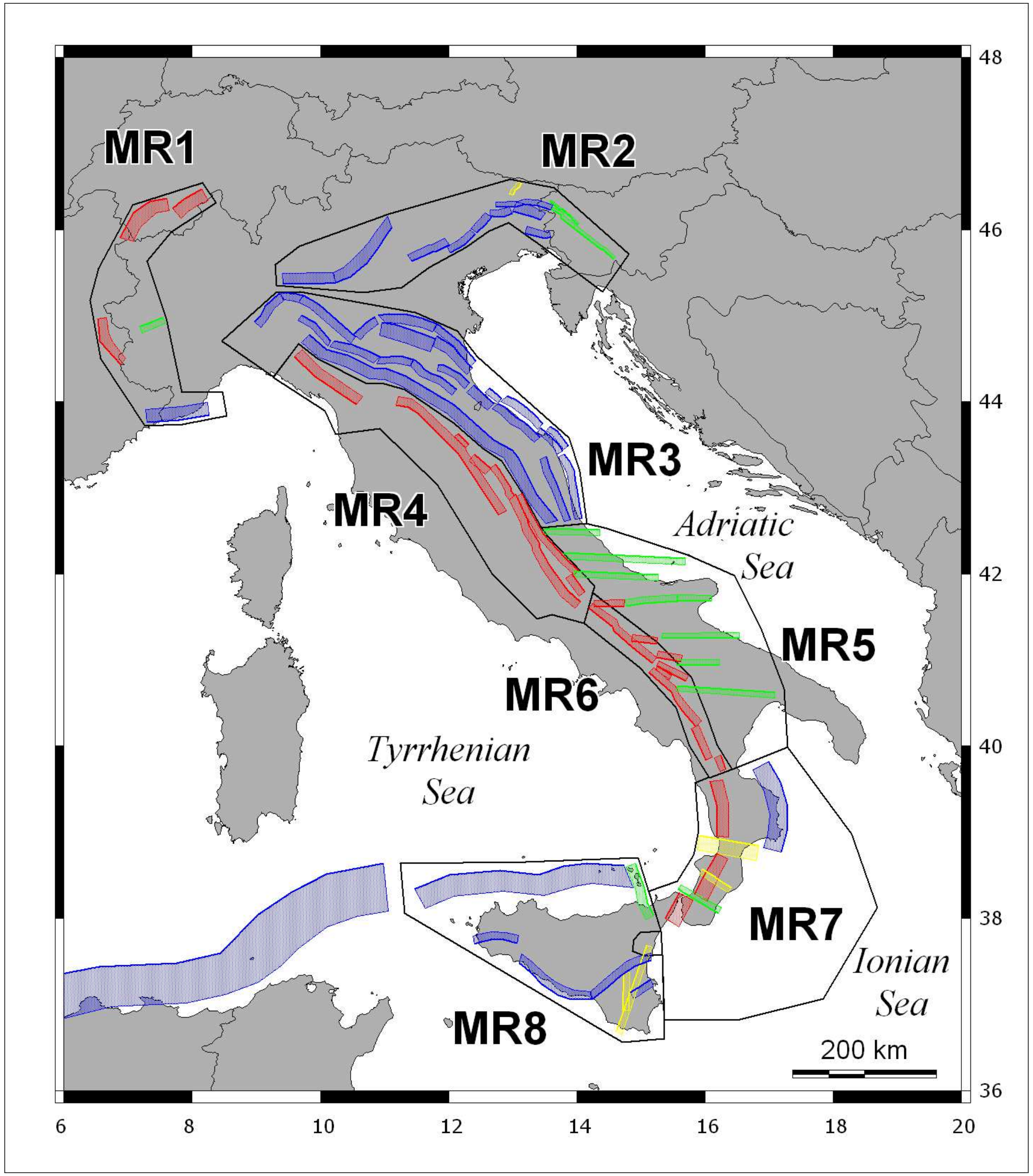
Software used: *R* functions available at

http://www.niph.go.jp/soshiki/gijutsu/download/Rfunctions/index.html

Results



	$M_w \geq 4.5$	$M_w \geq 5.3$	$M_w \geq 6$
MR_1	108	21	2
MR_2	215	26	4
MR_3	257	50	2
MR_4	383	45	9
MR_5	89	17	6
MR_6	90	20	9
MR_7	190	43	14
MR_8	137	26	4



faulting mechanism

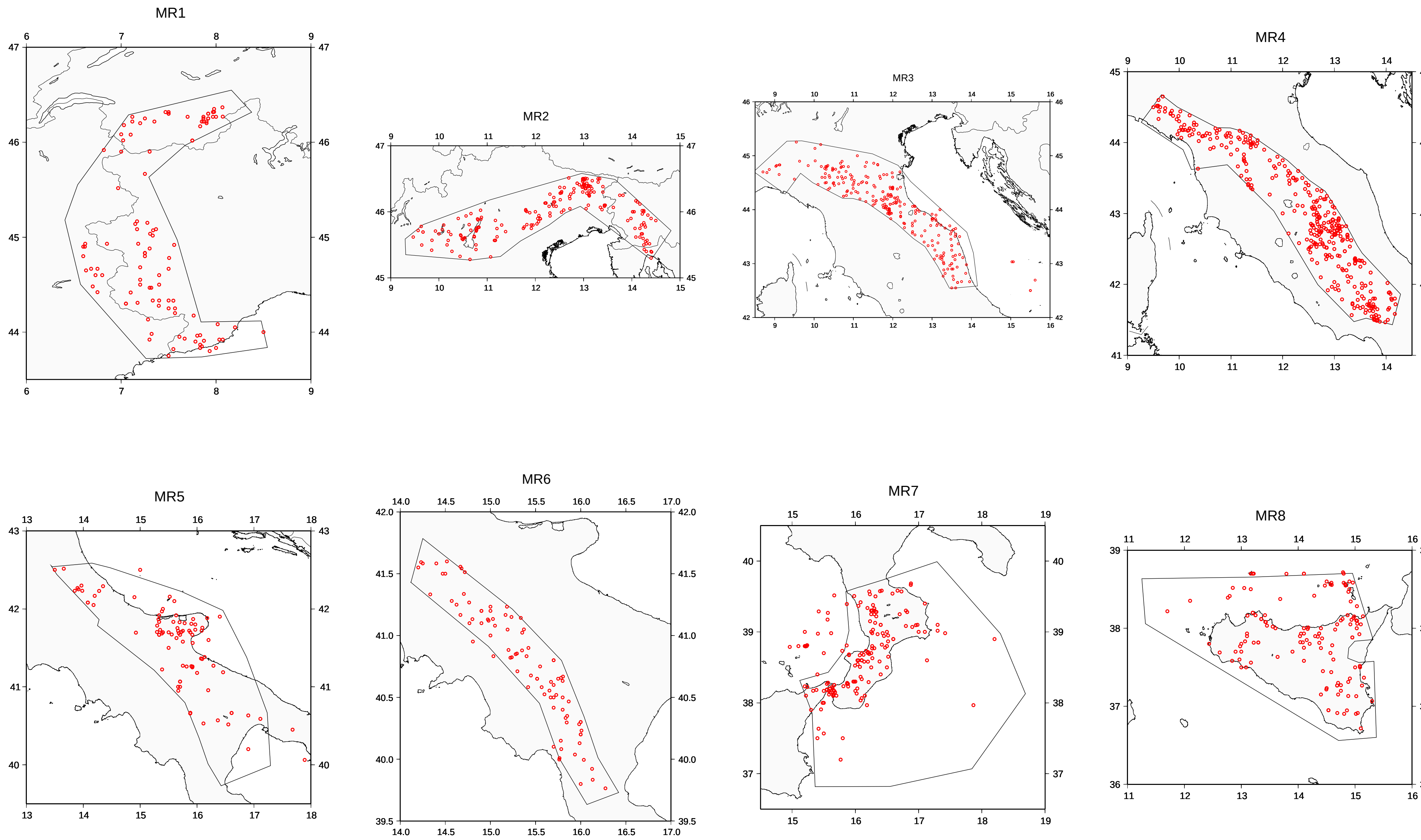
normal

reverse

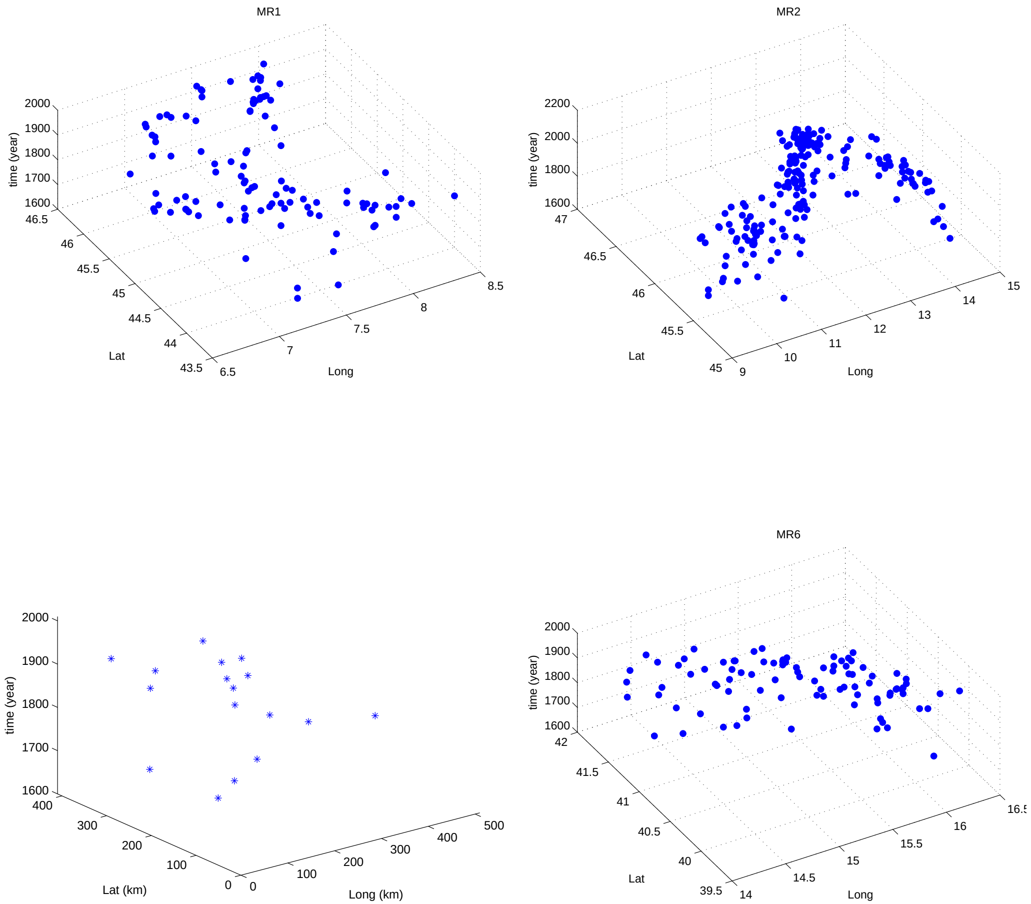
right-lateral strike-slip

left-lateral strike-slip

eight macro-regions, $M_w \geq 4.5$



eight macro-regions, $M_w \geq 4.5$



Null hypothesis vs Alternative hypothesis

H_0: the temporal distances between pairs of observations are independent of the spatial distances
H_1: not H_0

Knox's test: measures the excess of cases that are close both in space and time

$$T = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}^{S} a_{i,j}^{T}$$

where

$$a_{ij}^{S} = \begin{cases} 1, & i \neq j \text{ and } d_{i,j}^{S} < \delta_1 \ (km) \\ 0, & otherwise \end{cases} \qquad a_{ij}^{T} = \begin{cases} 1, & i \neq j \text{ and } d_{i,j}^{T} < \delta_2 \ (years) \\ 0, & otherwise \end{cases}$$

δ_1, δ_2 = critical space and time limits to be prespecified

$$E(T) = \frac{N_{1S} \ N_{1T}}{N}$$

$$Var(T) = \frac{N_{1S} \ N_{1T}}{N} + \frac{4 \ N_{2S} \ N_{2T}}{n(n-1)(n-2)} - \left(\frac{N_{1S} \ N_{1T}}{N}\right)^2 + \frac{4\{N_{1S}(N_{1S}-1)-N_{2S}\}\times\{N_{1T}(N_{1T}-1)-N_{2T}\}}{n(n-1)(n-2)(n-3)}$$

$$N_{1S(T)} = \tfrac{1}{2} \textstyle\sum_{i=1}^{n} \ \sum_{j=1}^{n} a_{ij}^{S(T)} \qquad N_{2S(T)} = \tfrac{1}{2} \textstyle\sum_{i=1}^{n} \ \sum_{j=1}^{n} \ \sum_{k\neq j}^{n} a_{ij}^{S(T)} \ a_{ik}^{S(T)}$$

Null distribution

given $T = t$, p-value $= Prob\{T \geq t\}$ can be approximated by either one of the following:

- Poisson distribution when N_{1S} and N_{1T} are small compared with $N = \frac{n(n-1)}{2}$

$$\text{mid}-p-\text{value} = 1 \; - \; \sum_{k=0}^{t} \frac{E(T)^k}{k!} \; exp\{-E(T)\} \; + \; \frac{E(T)^t}{t!} \; exp\{-E(T)\}$$

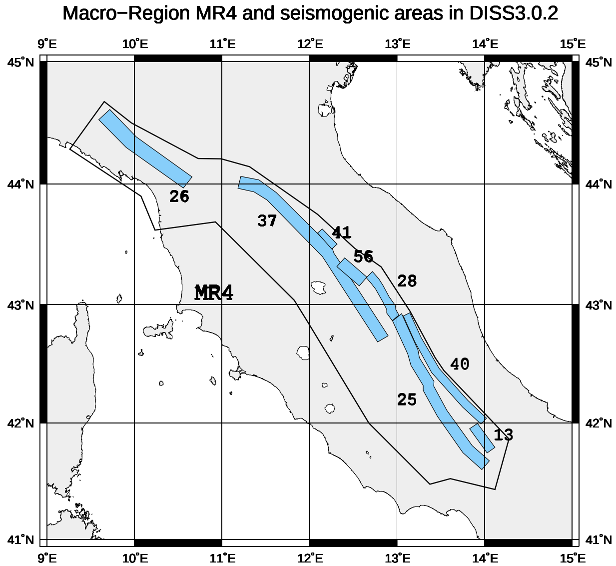
- Normal distribution

$$p - \text{value} = 1 \; - \; \Phi\left(\frac{t - E(T)}{\sqrt{var(T)}}\right)$$

- Monte Carlo hypothesis testing: we calculate the same statistic for a large number of data sets obtained by permuting the times among the fixed spatial locations (or viceversa)

$$\text{Simulated } p-\text{value} = \frac{1 \; + \; \sum_{\nu=1}^{Nrep} I(T_\nu \geq T_{obs})}{Nrep + 1}$$

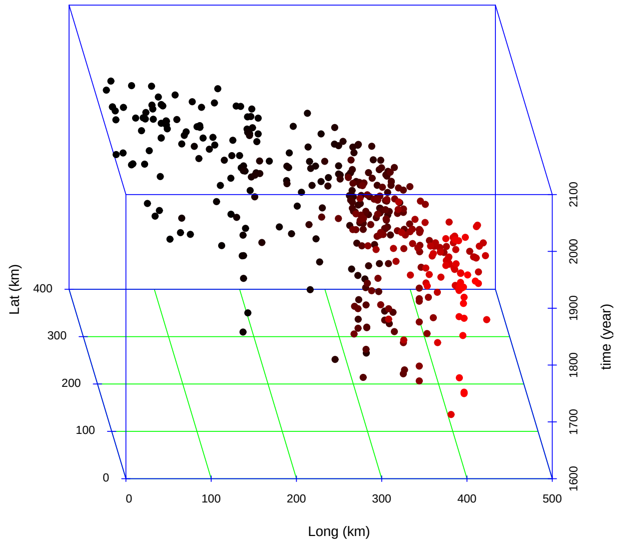
Macro-region MR_4

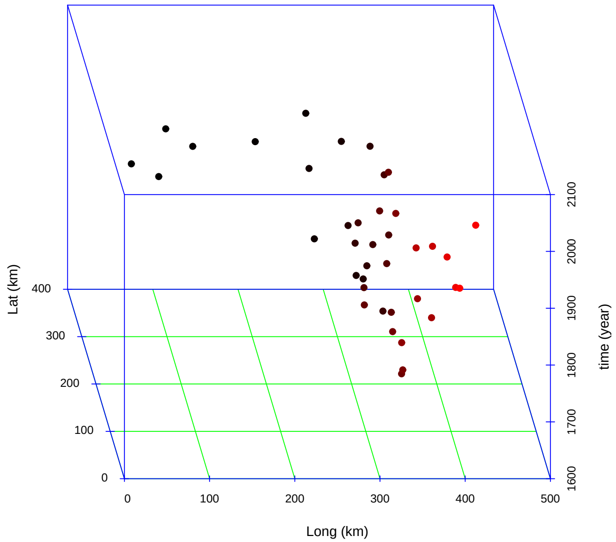


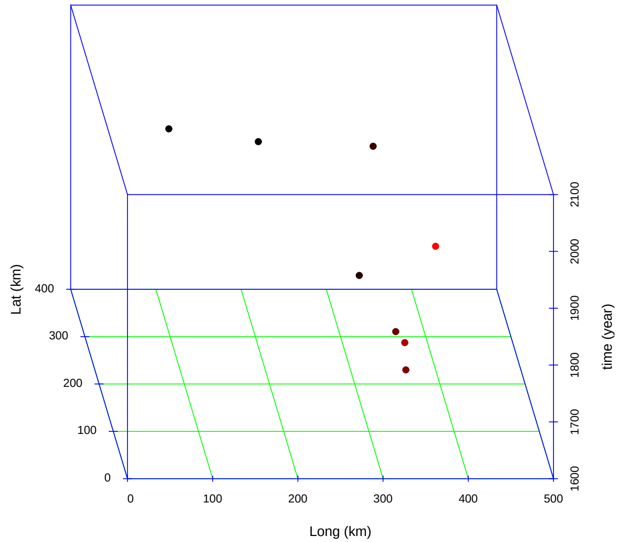
$M_w \geq 4.5 \quad n = 383$

$M_w \geq 5.3 \quad n = 39$

$M_w \geq 6.0 \quad n = 8$







Knox's test: $M_w \geq 4.5$

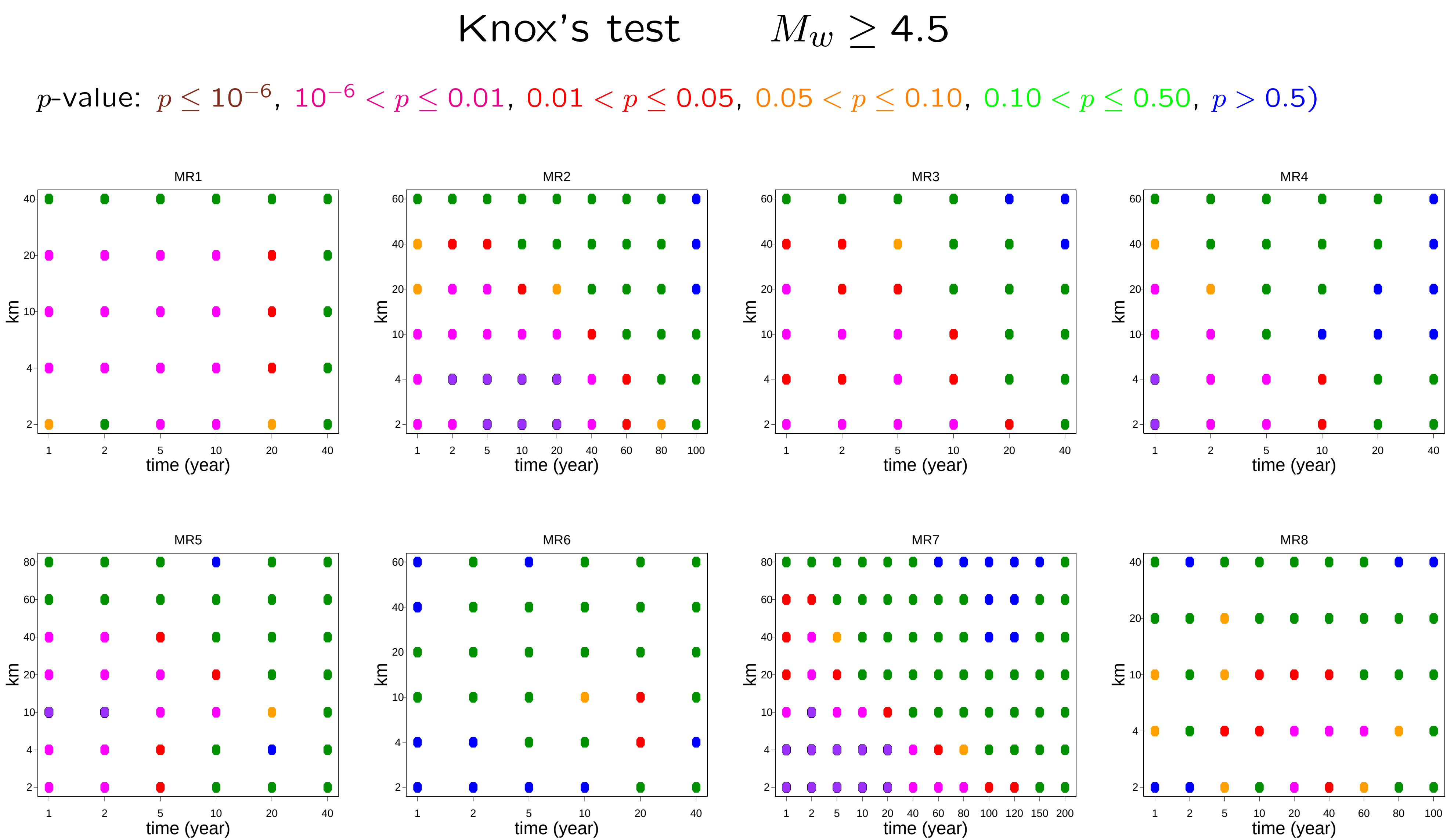
Critical Space limit δ_1 (km)		Critical time limit δ_2 (years)				
		≤ 1	≤ 2	≤ 5	≤ 10	≤ 20
≤ 2	Pairs	13	13	19	27	39^*
	Expected	1.789	3.577	8.644	16.827	31.944
	p-value	$3.334 \quad 10^{-8}$	$5.87 \quad 10^{-5}$	0.00111	0.0106	0.200
≤ 4	Pairs	17	20	33	48^*	72^*
	Expected	3.800	7.595	18.354	35.729	67.829
	p-value	$3.44 \quad 10^{-7}$	$8.82 \quad 10^{-5}$	$9.96 \; 10^{-4}$	0.0658	0.375
≤ 10	Pairs	29	46	84^*	136^*	244^*
	Expected	14.456	29.055	70.217	136.689	259.493
	p-value	0.000394	0.00178	0.131	0.513	0.664
≤ 20	Pairs	61	102^*	209^*	388^*	709^*
	Expected	40.657	81.265	196.395	382.315	725.792
	p-value	0.00142	0.0598	0.326	0.456	0.572
≤ 40	Pairs	141^*	258^*	590^*	1109^*	2106^*
	Expected	116.776	233.414	564.096	1098.105	2084.654
	p-value	0.0796	0.219	0.360	0.468	0.466
≤ 60	Pairs	225^*	438^*	987^*	1862^*	3484^*
	Expected	193.174	386.121	933.144	1816.516	3448.495
	p-value	0.108	0.145	0.320	0.418	0.465

Knox's test: $M_w \geq 5.3$

Critical Space limit δ_1 (km)		Critical time limit δ_2 (years)					
		≤ 2	≤ 5	≤ 10	≤ 20	≤ 40	≤ 60
≤ 10	Pairs	1	2	2	3	5	6
	Expected	0.222	0.822	1.333	2.467	4.111	6.311
	MC p-value	0.214	0.190	0.387	0.458	0.406	0.650
≤ 20	Pairs	1	2	2	5	12	23
	Expected	0.859	3.177	5.152	9.530	15.884	24.384
	MC p-value	0.576	0.857	0.975	0.979	0.905	0.673
≤ 40	Pairs	2	3	6	19	37	62
	Expected	2.182	8.073	13.091	24.218	40.364	61.964
	MC p-value	0.694	0.994	0.995	0.927	0.765	0.495
≤ 60	Pairs	2	6	16	40	65	96
	Expected	3.444	12.744	20.667	38.233	63.722	97.822
	MC p-value	0.921	0.999	0.906	0.362	0.469	0.568

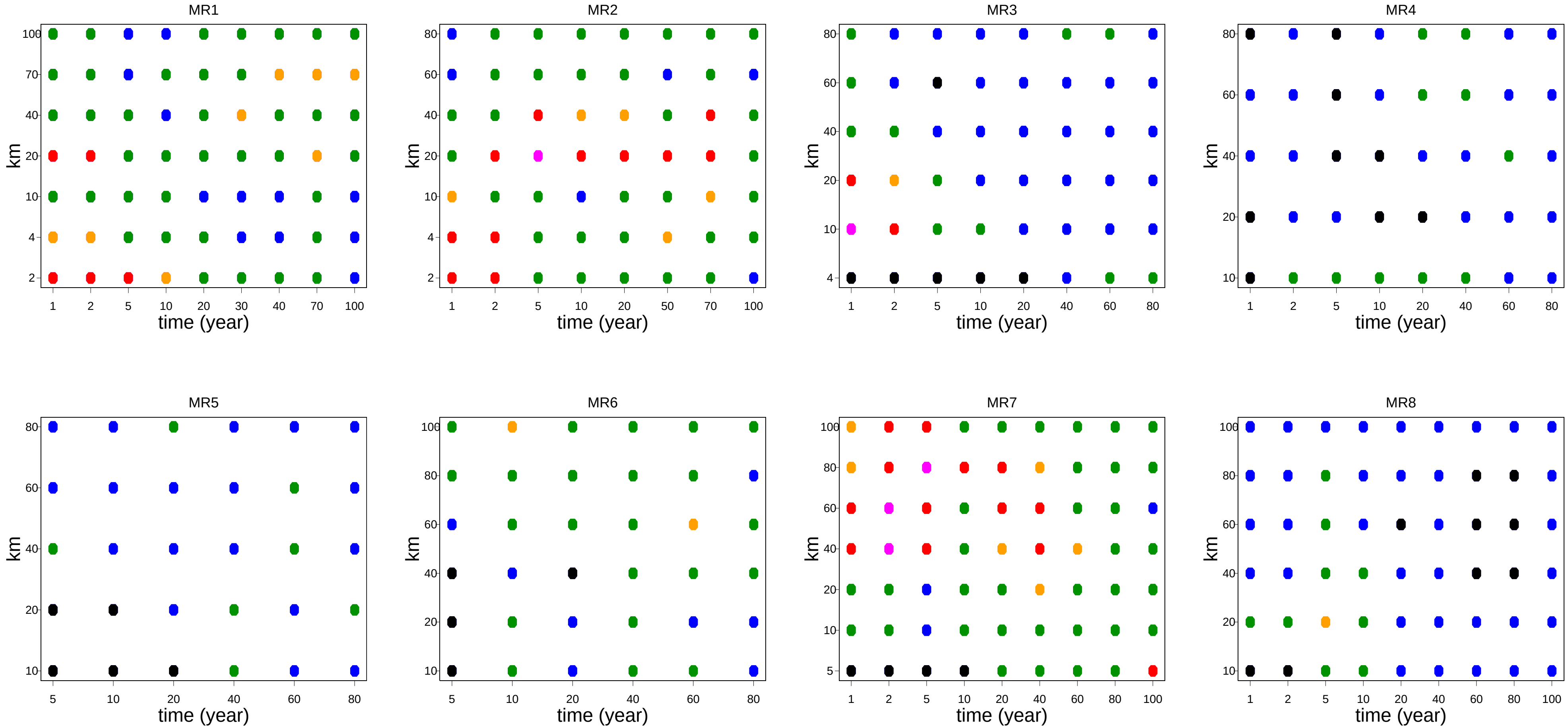
Knox's test: $M_w \geq 6.0$

Critical Space limit δ_1 (km)		Critical time limit δ_2 (years)				
		≤ 2	≤ 10	≤ 20	≤ 50	≤ 100
≤ 30	Pairs	1	1	1	1	3
	Expected	0.222	0.444	0.556	1.000	1.667
	MC p-value	0.212	0.388	0.463	0.734	0.152
≤ 50	Pairs	1	1	1	1	3
	Expected	0.333	0.667	0.833	1.500	2.500
	MC p-value	0.321	0.520	0.658	0.864	0.467
≤ 100	Pairs	1	1	1	3	5
	Expected	0.778	1.556	1.944	3.500	5.833
	MC p-value	0.664	0.889	0.974	0.802	0.914


Knox's test $M_w \geq 4.5$
p-value: $p \leq 10^{-6}$, $10^{-6} < p \leq 0.01$, $0.01 < p \leq 0.05$, $0.05 < p \leq 0.10$, $0.10 < p \leq 0.50$, $p > 0.5$)
MR1
MR2
MR3
MR4
MR5
MR6
MR7
MR8
km
time (year)

Knox's test $\quad M_w \geq 5.3$

p-value: $p \leq 10^{-6}$, $10^{-6} < p \leq 0.01$, $0.01 < p \leq 0.05$, $0.05 < p \leq 0.10$, $0.10 < p \leq 0.50$, $0.50 < p \leq 0.95$, $p > 0.95$



Jacquez's k-NN test: proposes to substitute the distance by the Kth nearest neighbours (k-NN)

$$T_k = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}^S a_{i,j}^T$$

where

$$a_{ij}^S = \begin{cases} 1, & \text{if case } j \text{ is a } k \text{ NN of case } i \ (\neq j) \text{ in space} \\ 0, & otherwise \end{cases}$$

$$a_{ij}^T = \begin{cases} 1, & \text{if case } j \text{ is a } k \text{ NN of case } i \ (\neq j) \text{ in time} \\ 0, & otherwise \end{cases}$$

Null distribution: simulated applying Monte Carlo hypothesis testing

Jacquez's test: K-NN for MR_4

K nearest neighbour

$M_w \geq 4.5$	1	2	3	4	5	10	20	30	40
T	0.5	3	10.5	20	28.5	69.5	214	467	847.5
Expected	0.658	2.319	4.983	8.649	13.32	51.70	203.7	455.9	808.4
MC p-value	0.629	0.297	0.002	0.001	0.001	0.007	0.214	0.239	0.057
$\widehat{s}_k$	4.633	6.550	7.914	9.293	10.54	15.84	24.70	31.63	38.24
$\widehat{t}_k$	0.564	1.059	1.536	2.053	2.572	5.538	10.63	15.24	19.56

	50	60	70	80	90	100	150
T	1315	1871	2521	3255.5	4108.5	5052.5	11273.5
Expected	1261.1	1814.1	2467.4	3220.9	4074.7	5028.8	11303.
MC p-value	0.033	0.035	0.081	0.185	0.183	0.245	0.551
$\widehat{s}_k$	45.36	52.14	60.90	67.76	73.27	80.09	122.1
$\widehat{t}_k$	23.94	28.18	33.18	38.20	42.99	47.69	60.99

K nearest neighbour

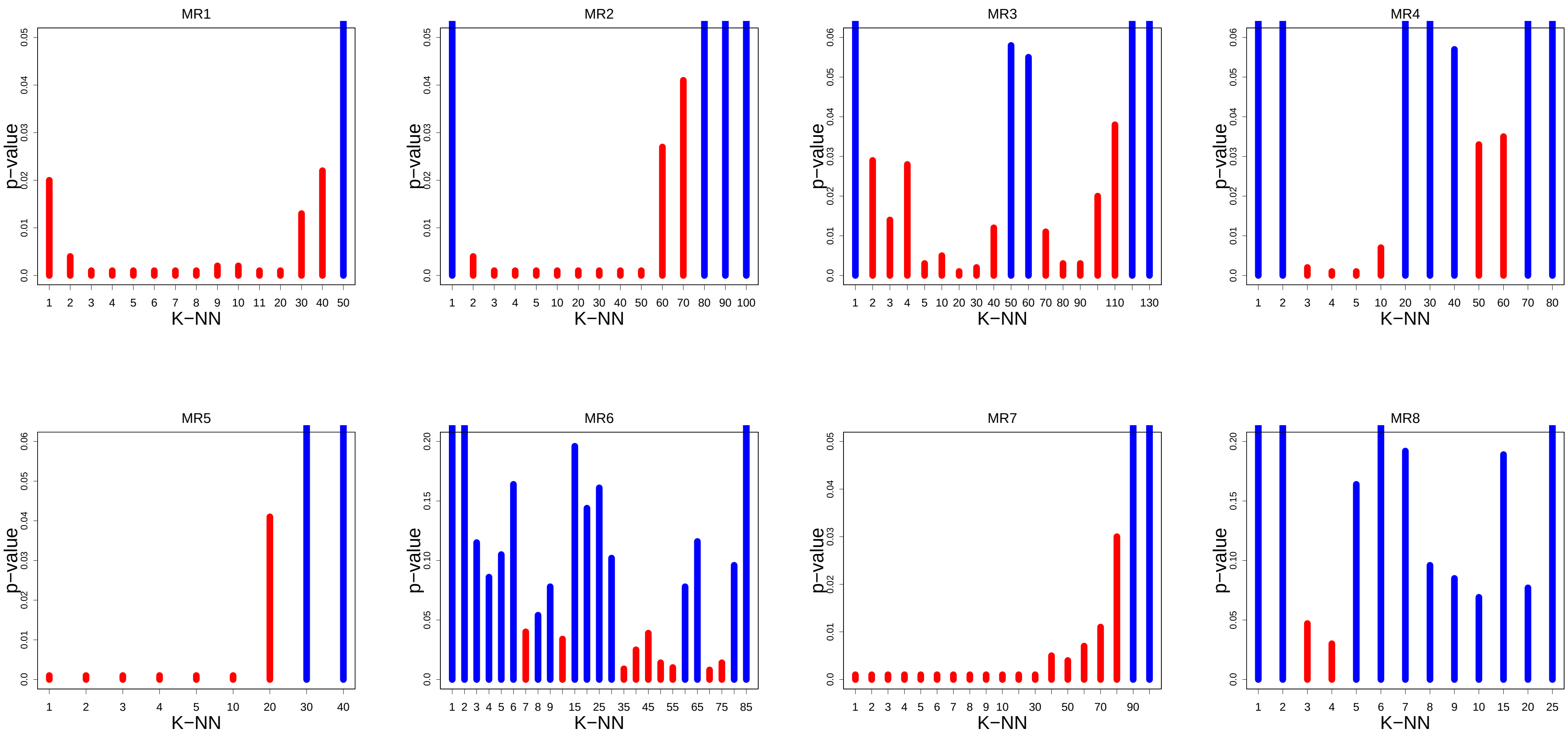
$\mathbf{M}_w \geq 5.3$	1	2	3	4	5	6	7	8	9
T	2.5	4	5.5	8	13	22	30	36.5	43
Expected	0.511	2.045	4.602	8.182	12.78	18.41	25.06	32.73	41.42
MC p-value	0.012	0.084	0.305	0.540	0.458	0.158	0.094	0.209	351
$\widehat{s}_k$	13.18	19.82	26.00	34.56	38.15	47.40	52.68	56.84	63.38
$\widehat{t}_k$	3.730	8.554	14.50	19.29	23.87	30.34	33.58	38.78	42.28

	10	15	20	30
T	53.5	120.5	206	458
Expected	51.14	115.1	204.5	460.2
MC p-value	0.283	0.196	0.415	0.547
$\widehat{s}_k$	68.87	82.70	93.92	139.2
$\widehat{t}_k$	46.76	69.74	96.49	150.4

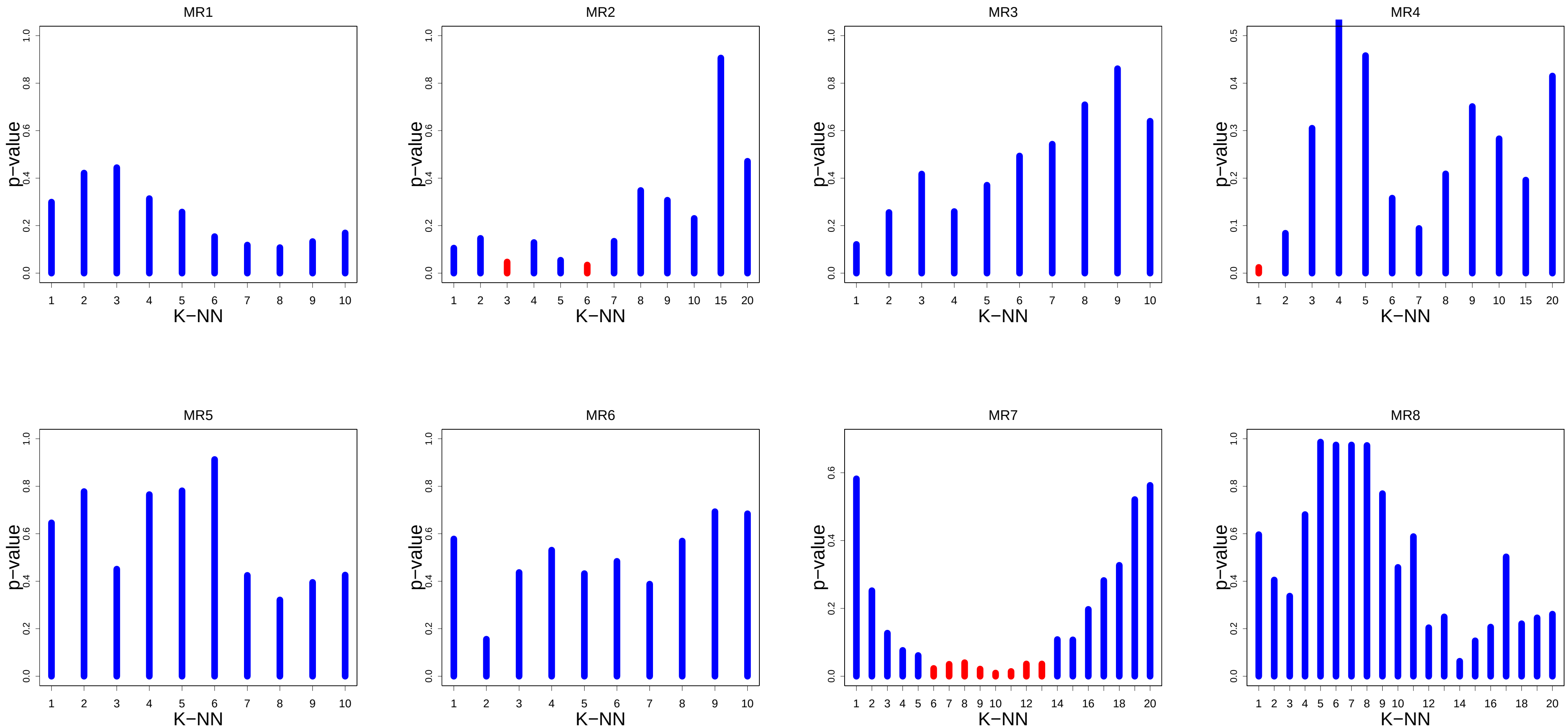
K nearest neighbour

$\mathbf{M}_w \geq 6.0$	1	2	3	4	5
T	1.	3.	5.5	10.	15.
Expected	0.563	2.250	5.063	9.	14.063
MC p-value	0.361	0.275	0.414	0.238	0.317

Jacquez's K-NN test $\quad M_w \geq 4.5$



Jacquez's K-NN test $M_w \geq 5.3$



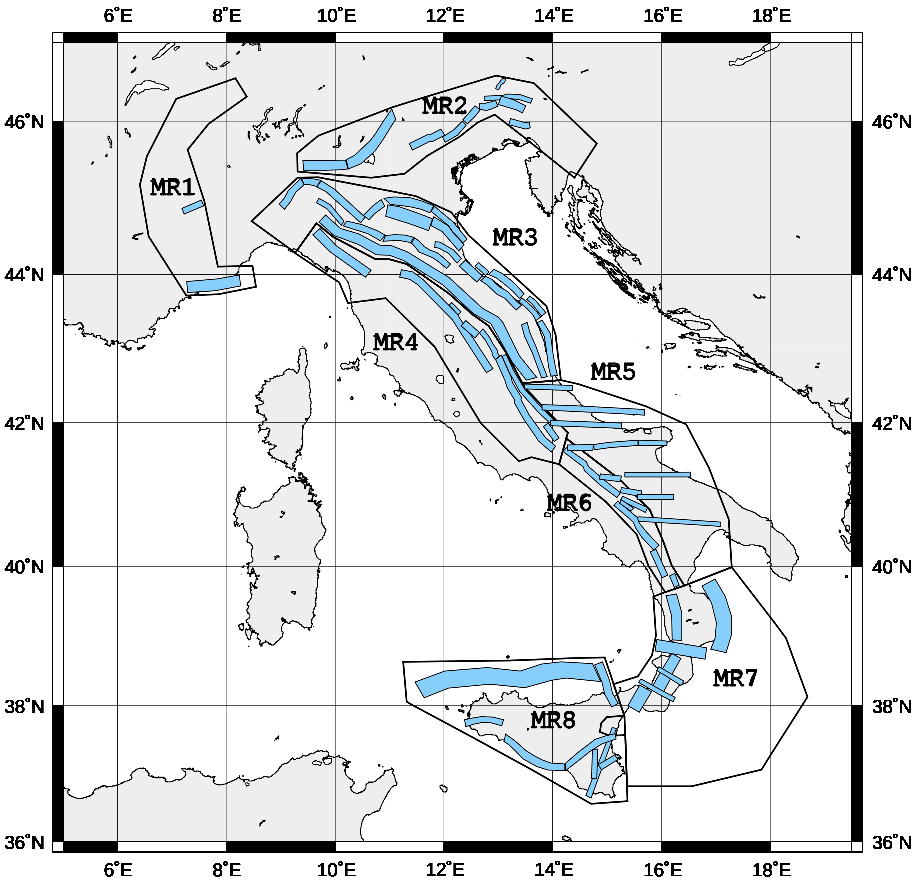
Summarizing

we have performed global tests: they provide a **single statistic** assessing the degree to which a map **pattern deviates from** the null hypothesis of **space-time randomness**, do not provide information on the size and location of clusters

(1) essential agreement between Knox's and K-NN tests
(2) clustering effect is not scale-invariant

$$\begin{array}{ll} \text{high clustering } \wedge & \left\{ \begin{array}{l} \text{Eastern Alps (MR}_2) \\ \text{Central Northern Apennines Alps (MR}_3) \\ \text{Central Northern Apennines West (MR}_4) \\ \text{Calabrian Arc (MR}_7) \end{array} \right. \\ \text{space} - \text{time negative correlation} & \\ & \\ \text{low clustering} & \left\{ \text{Southern Apennines (MR}_6) \right. \\ & \\ \text{clustering } \wedge & \left\{ \begin{array}{l} \text{Western Alps (MR}_1) \\ \text{Southern Apennines Apulia (MR}_5) \end{array} \right. \\ \text{uncorrelated space} - \text{time} & \\ & \\ \text{mixed} & \left\{ \text{Sicily (MR}_8) \right. \end{array}$$

in MR_2 and MR_3 the average half-sum of the length of consecutive seismogenic areas is comparable with the average distance of the "rejected" K-NN $\Longrightarrow$ clusters are due to fault interaction (?)



when a global test finds

- no significant deviation from randomness, local tests may be useful to uncover isolated hot spots of increased activity

- significant degree of clustering, local statistics may establish whether (1) the study area is homogeneous, or (2) there are local outliers which contribute to the significant global statistic

Future work:

- tests for the detection of clusters used when there is no a priori idea of **where and how large** the clusters may be; many local tests are carried out simultaneously

- focused (local) tests used to evaluate whether clustering occurs around particular foci in prospective way