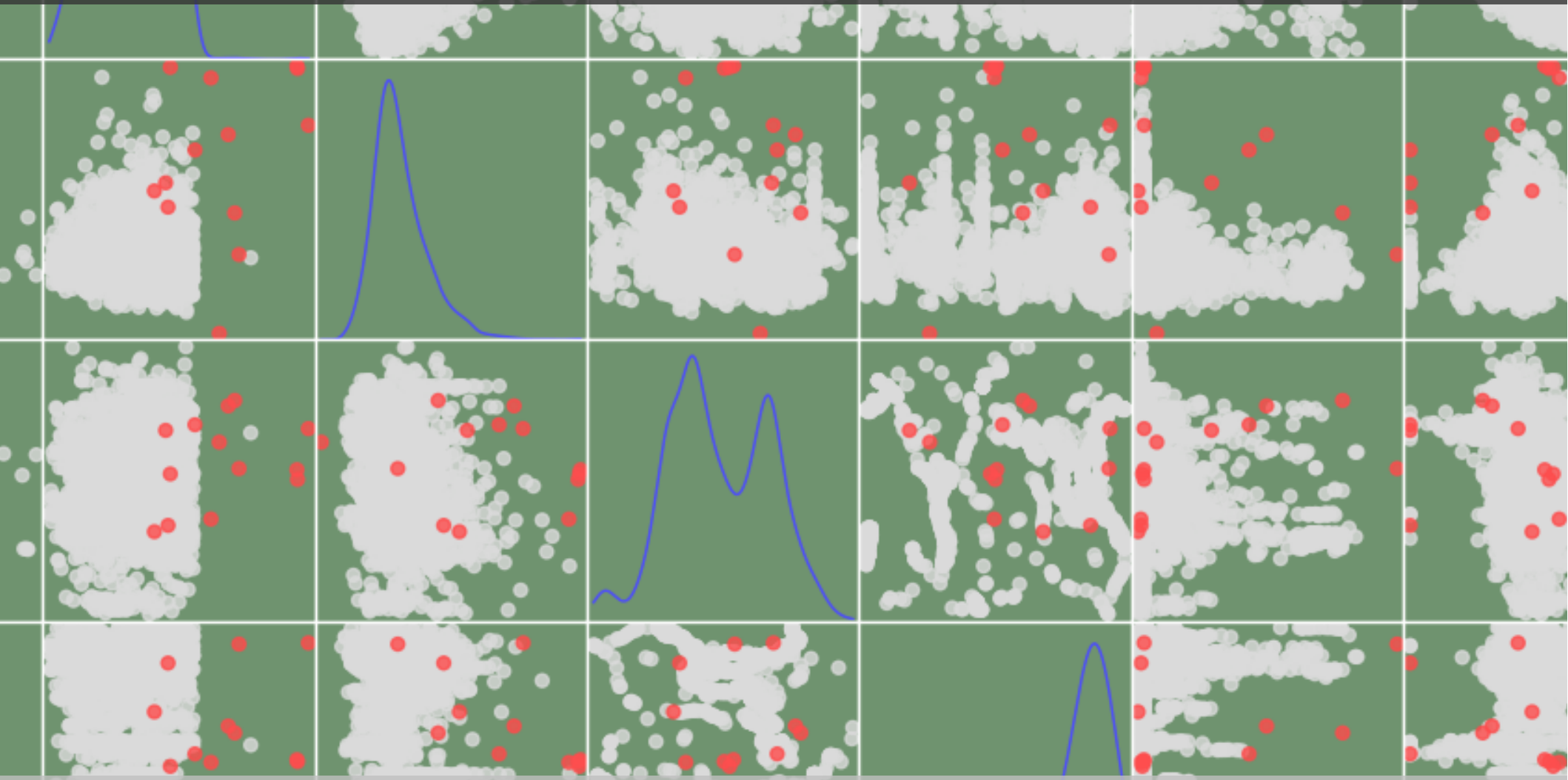
Intelligence and statistics for rapid and robust earthquake detection, association and location



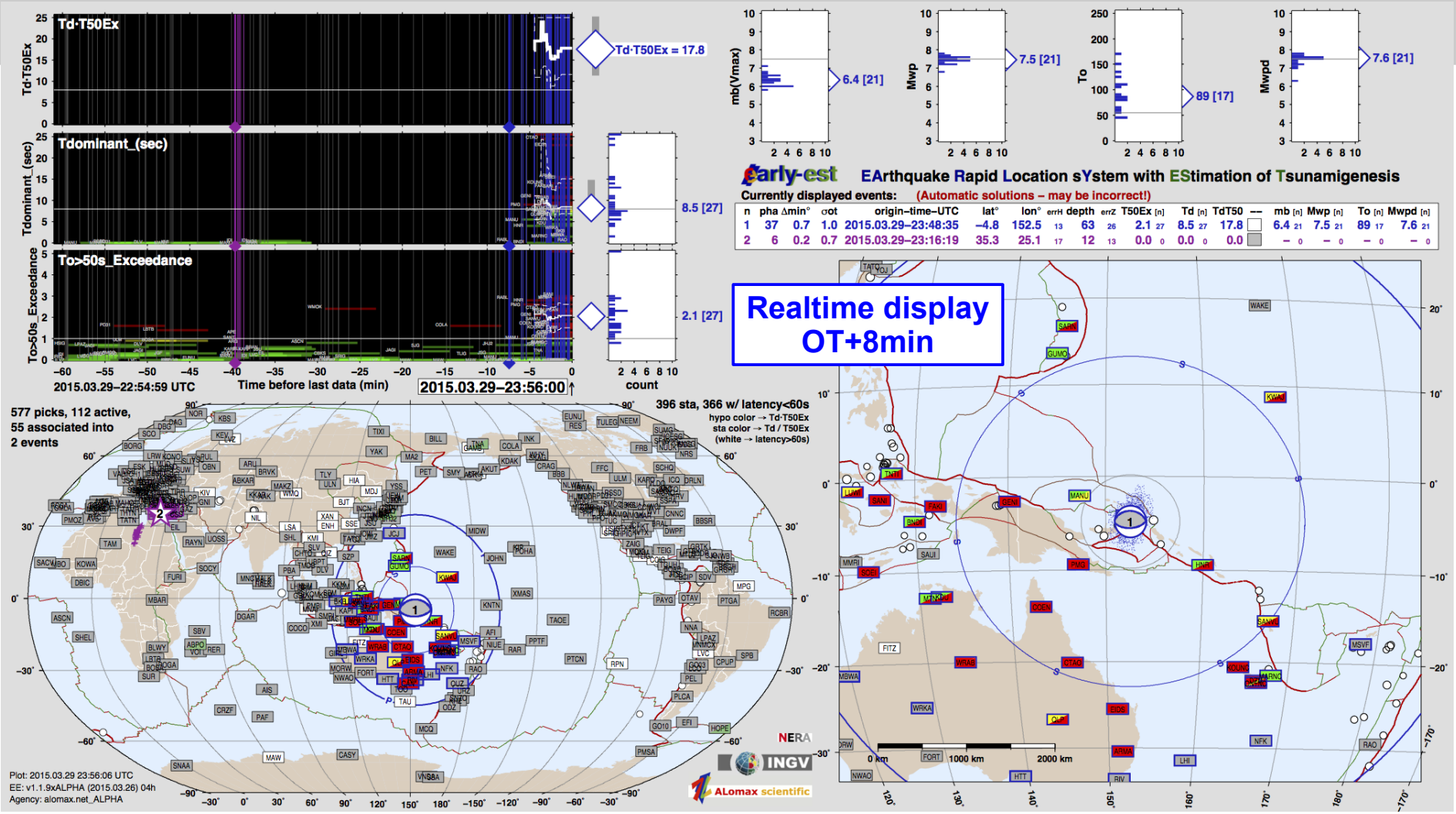
Anthony Lomax
ALomax Scientific, Mouans-Sartoux, France
anthony@alomax.net www.alomax.net @ALomaxNet

Alberto Michelini, Fabrizio Bernardi, and Valentino Lauciani
Istituto Nazionale di Geofisica e Vulcanologia, Roma, Italy

Early-est: rapid, fully automatic determination of the location, depth, magnitude, mechanism and tsunami potential of an earthquake

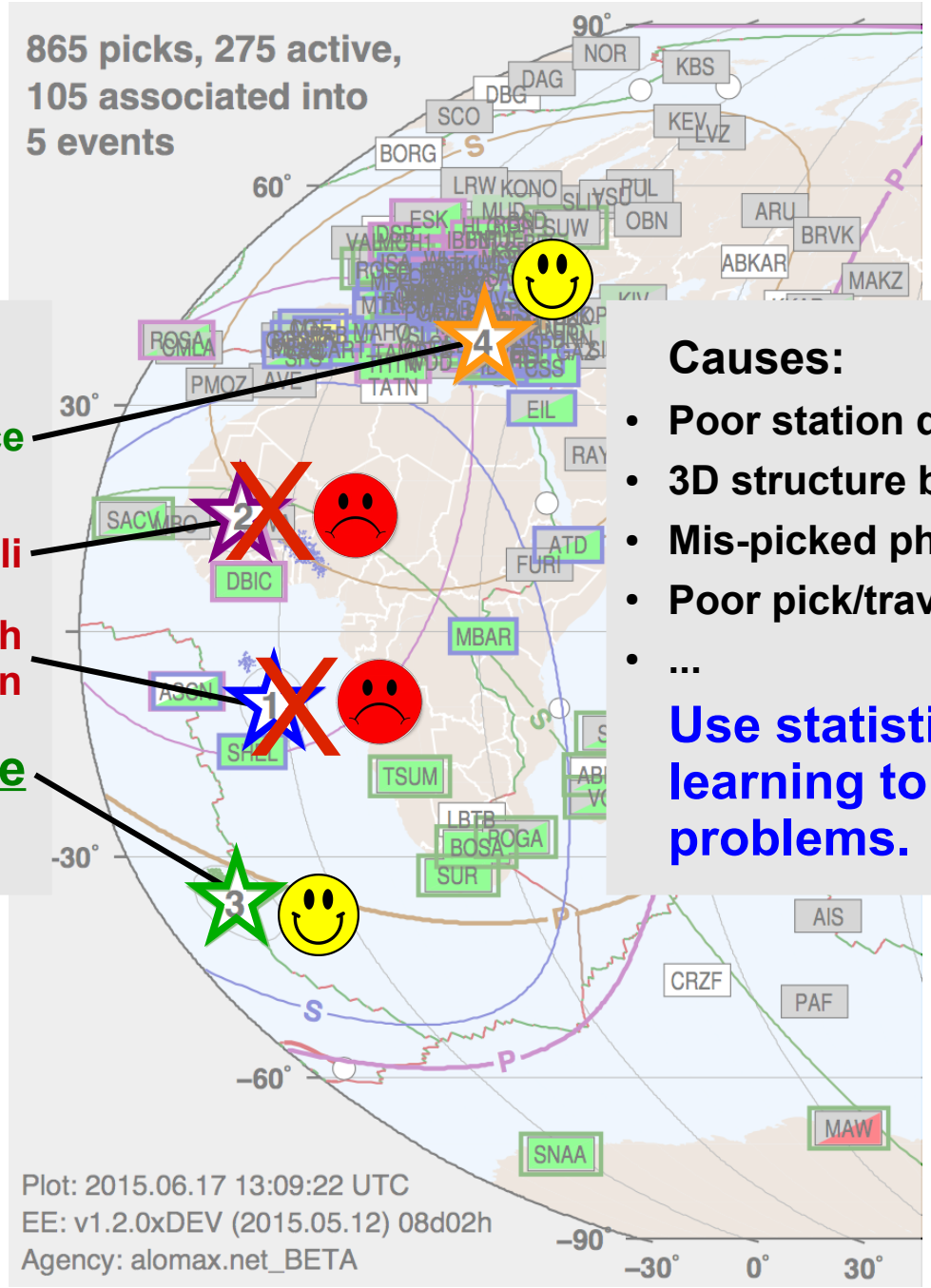
For effective earthquake and tsunami early-warning it is crucial that key earthquake parameters are determined as rapidly and reliably as possible.

EarlyEst: Rapid earthquake analysis module at INGV CAT tsunami alert center:



Rapid, early results use minimal data: prone to bias & errors, poor magnitudes, false events, ...

Example:
False events



M3.5 Greece

FALSE: M6 Mali

FALSE: M6 South Atlantic Ocean

M7 Mid-Atlantic Ridge

Causes:

- Poor station distribution
- 3D structure but 1D velocity model
- Mis-picked phases
- Poor pick/travel-time error model
- ...

Use statistics and machine learning to identify problems.

Identification of false events: apply statistics and machine learning to past true and (few) false events

"Data Frame" (2D array) of training data:

possible important attributes to discriminate true or false events

true/false	ph_assoc	ph_used	dmin(deg)	gap2(deg)	err_otime(sec)	lat(deg)	lon(deg)	depth(km)	mb	n	otime(UTC)	region
1	6	6	0.68	151.572	2.491	-13.31	132.195	10.16	-9	1	2015.05.01	Northern_Territory,_Australia
1	21	17	11.024	135.796	2.805	-16.84	66.467	0.78	5.154	6	2015.04.19	Mid_-_Indian_Ridge
0	11	7	1.771	166.38	2.639	-6.237	130.352	138.28	4.588	5	2015.06.17	Banda_Sea
0	7	7	0.636	247.126	1.508	38.496	21.876	10.16	-9	0	2015.06.17	Greece
0	45	39	30.86	75.262	5.01	-36.042	-16.471	7.81	5.573	26	2015.06.17	Southern_Mid_-_Atlantic_Ridge
0	44	38	3.688	98.833	1.073	2.598	124.161	332.42	4.907	31	2015.06.15	Celebes_Sea
0	65	56	1.702	90.896	1.419	27.306	140.285	383.2	4.892	48	2015.06.14	Bonin_Islands,_Japan_Region
0	49	42	3.731	72.411	1.928	-11.398	163.176	19.53	5.153	37	2015.06.14	Solomon_Islands
0	45	39	1.962	82.679	1.299	-0.232	124.559	66.41	5.041	32	2015.06.13	Southern_Molucca_Sea
0	36	32	1.678	125.296	1.572	45.355	142.046	283.2	4.623	22	2015.06.13	Hokkaido,_Japan_Region
0	48	38	3.7	99.016	1.389	-3.27	143.81	10.16	5.166	31	2015.06.13	Near_N_Coast_of_New_Guinea,_Png.
0	37	26	14.836	128.997	1.308	13.704	120.71	250.78	4.843	21	2015.06.13	Mindoro,_Philippines
0	78	65	5.116	121.641	3.102	-24.504	-175.774	19.53	5.133	37	2015.06.13	South_of_Tonga_Islands
0	34	34	0.699	113.06	1.23	34.446	137.382	322.66	4.604	18	2015.06.13	Near_S._Coast_of_Honshu,_Japan
0	104	98	2.493	134.463	0.862	56.689	-156.725	34.77	5.21	84	2015.06.13	Alaska_Peninsula
0	53	50	1.731	125.087	1.592	38.851	142.607	34.77	5.027	36	2015.06.12	Near_East_Coast_of_Honshu,_Japan
0	47	32	9.439	120.563	2.042	6.173	116.677	92.19	5.064	29	2015.06.12	Borneo
0	44	33	1.378	129.993	1.811	-5.562	152.304	19.53	5.138	31	2015.06.12	New_Britain_Region,_P.N.G.
0	170	158	2.059	39.853	1.415	-15.56	-173.049	10.16	5.455	97	2015.06.12	Tonga_Islands
0	50	41	4.058	93.874	2.858	-20.805	-173.832	10.16	5.058	23	2015.06.12	Tonga_Islands
0	53	38	1.663	144.499	1.695	8.009	126.963	96.09	5.026	34	2015.06.11	Mindanao,_Philippines
0	36	26	3.828	118.986	1.688	12.902	146.272	19.53	5.179	23	2015.06.11	South_of_Mariana_Islands
0	41	33	1.943	161.243	0.814	29.73	131.359	9.96	5.032	27	2015.06.11	Southeast_of_Ryukyu_Islands
0	83	74	1.981	133.715	1.902	39.677	143.213	19.53	5.284	67	2015.06.11	Off_East_Coast_of_Honshu,_Japan
0	178	165	2.192	81.346	1.391	39.503	143.387	19.73	5.328	136	2015.06.11	Off_East_Coast_of_Honshu,_Japan

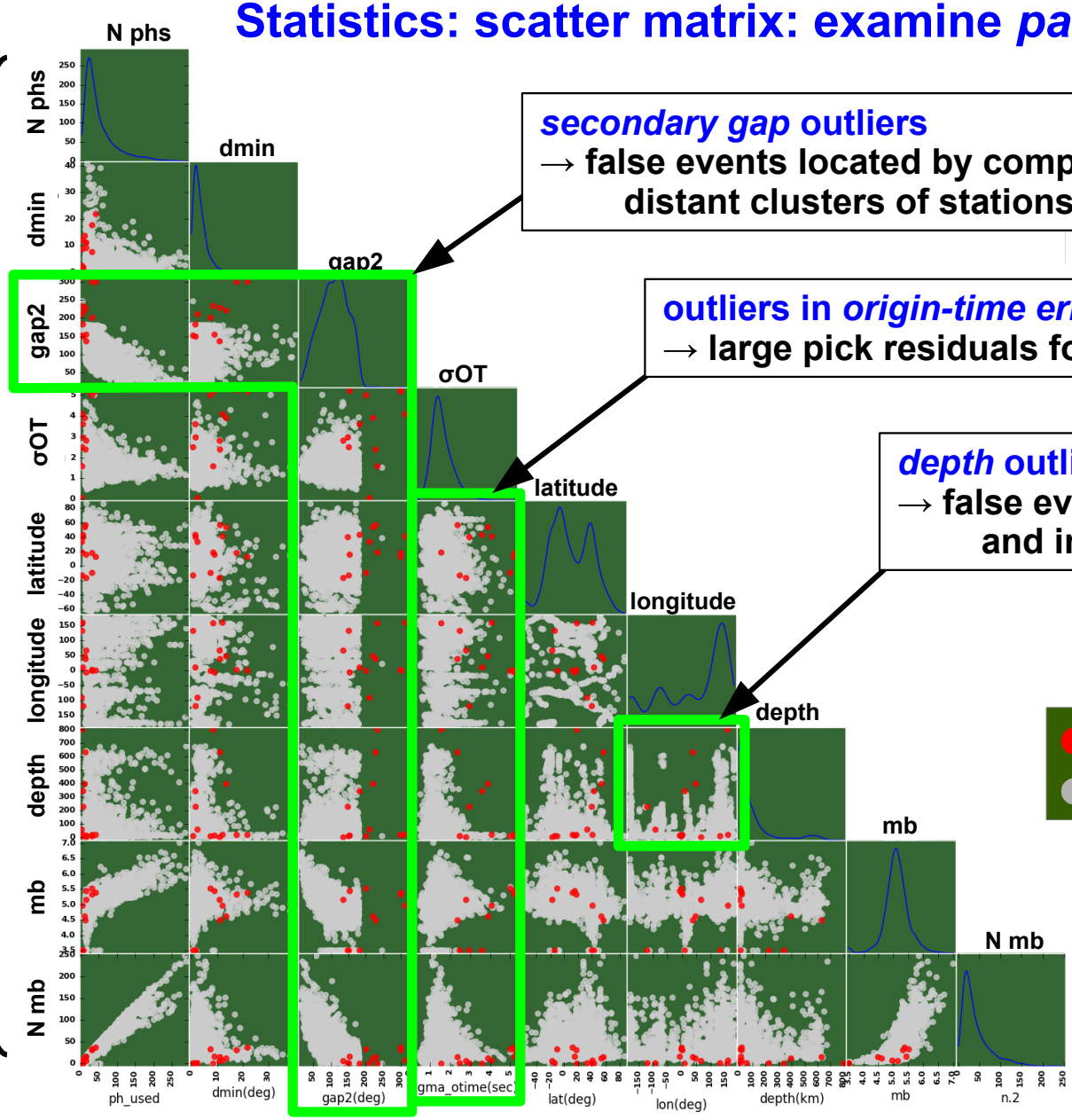
"labelled" - identified as true or false event

Problem: many true and few false events!

First step: basic statistical & expert analysis of past true and (few) false events

Statistics: scatter matrix: examine *pairs of attributes*

possible important attributes to discriminate true or false events



***secondary gap* outliers**
→ false events located by compact, distant clusters of stations

outliers in *origin-time error*
→ large pick residuals for false events

depth* outliers in *longitude
→ false events sometimes deep and in a-seismic regions

false events
true events

pandas.tools.plotting.scatter_matrix

Second step: Machine learning (classification, regression, …) for identifying outliers, making decisions, finding patterns...

Examine semi-automatically in high-dimension *many attributes*

What is machine learning?

Given training data, construct an algorithm to make predications on new data.

1. Learn (select and tune algorithms) using training data.
2. Test algorithm on testing data.
3. Apply algorithm to new data.

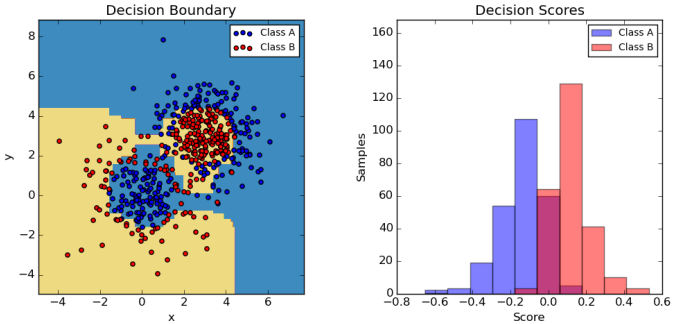
- Supervised learning: predict attributes of data:
  - Classification: learn from labeled, *xy* training data how to predict the (*discrete*) class *y* of new, unlabeled data *x*.
  - Regression: learn from *xy* training data how to predict the (*continuous*) values of *y* variables in new data *x*.

- Unsupervised learning:
  No target attributes, try to discover clustering or distribution of the data, or reduce the dimensionality of the data.

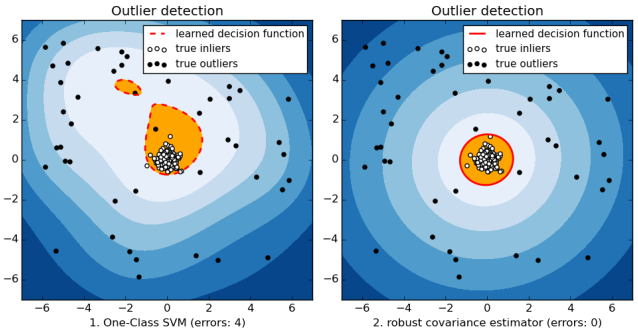
http://scikit-learn.org

Applications:

Decision / classification
(e.g. False event?
Tsunamigenic earthquake?)

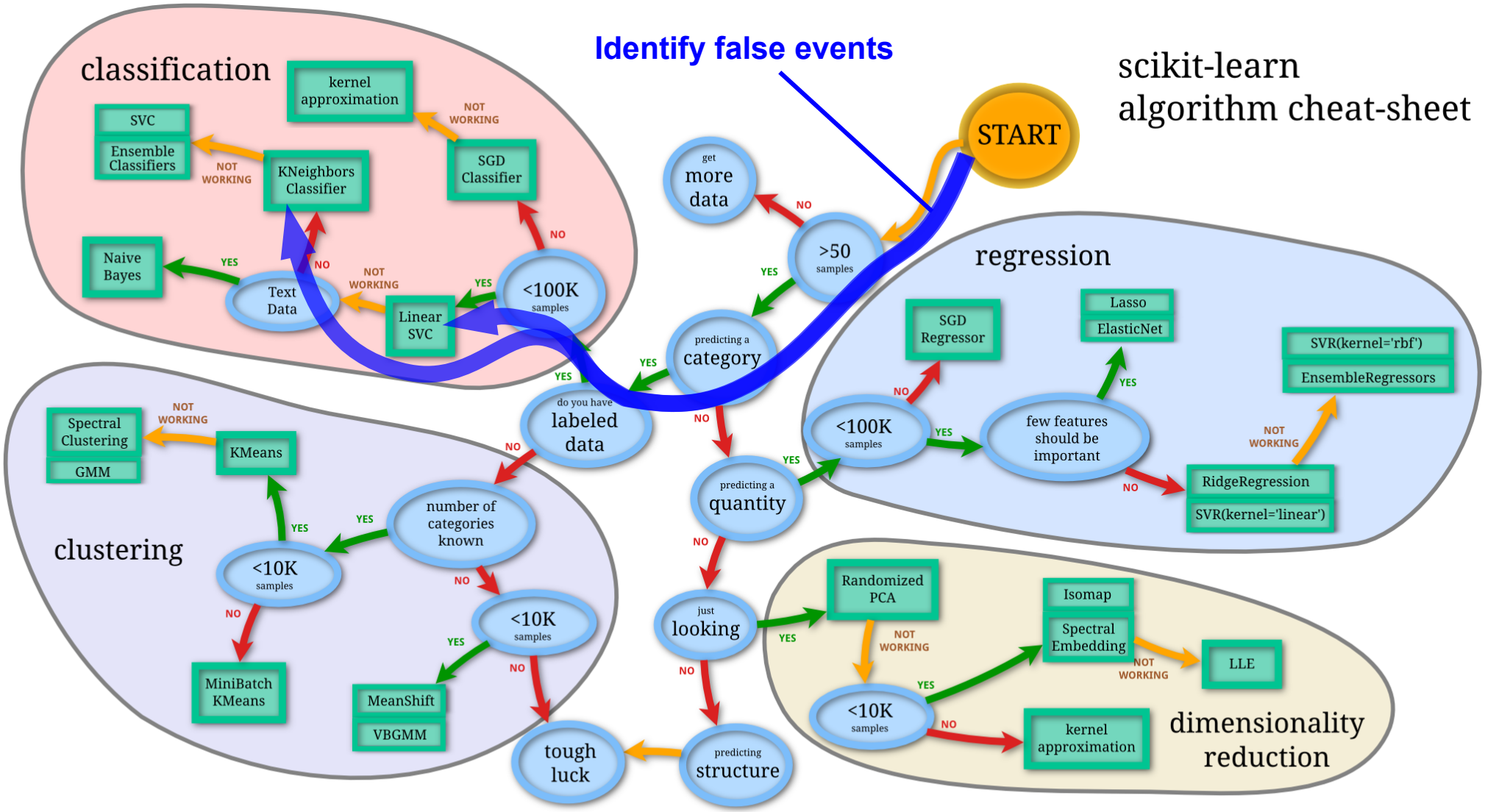


Outlier detection
(e.g. False event? Unusual event?)



and many more...

Machine learning: multitude of methods depending on goals and characteristics of data set



http://scikit-learn.org

Multiple machine learning algorithms: train and test with past true and (few) false events

Data Frame (2D array) of training data:

possible important attributes to discriminate true or false events

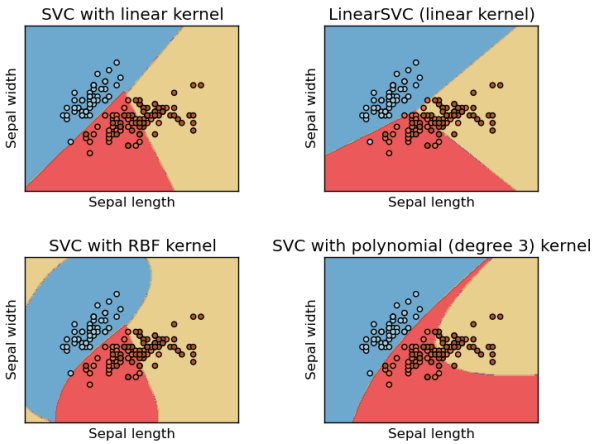
true/false	ph_assoc	ph_used	dmin(deg)	gap2(deg)	err_otime(sec)	lat(deg)	lon(deg)	depth(km)	mb	n	otime(UTC)	region
1	6	6	0.68	151.572	2.491	-13.31	132.195	10.16	-9	1	2015.05.01	Northern_Territory,_Australia
1	21	17	11.024	135.796	2.805	-16.84	66.467	0.78	5.154	6	2015.04.19	Mid_-_Indian_Ridge
0	11	7	1.771	166.38	2.639	-6.237	130.352	138.28	4.588	5	2015.06.17	Banda_Sea
0	7	7	0.636	247.126	1.508	38.496	21.876	10.16	-9	0	2015.06.17	Greece
0	45	39	30.86	75.262	5.01	-36.042	-16.471	7.81	5.573	26	2015.06.17	Southern_Mid_-_Atlantic_Ridge
0	44	38	3.688	98.833	1.073	2.598	124.161	332.42	4.907	31	2015.06.15	Celebes_Sea
0	65	56	1.702	90.896	1.419	27.306	140.285	383.2	4.892	48	2015.06.14	Bonin_Islands,_Japan_Region
0	49	42	3.731	72.411	1.928	-11.398	163.176	19.53	5.153	37	2015.06.14	Solomon_Islands
0	45	39	1.962	82.679	1.299	-0.232	124.559	66.41	5.041	32	2015.06.13	Southern_Molucca_Sea
0	36	32	1.678	125.296	1.572	45.355	142.046	283.2	4.623	22	2015.06.13	Hokkaido,_Japan_Region
0	48	38	3.7	99.016	1.389	-3.27	143.81	10.16	5.166	31	2015.06.13	Near_N_Coast_of_New_Guinea,_Png.
0	37	26	14.836	128.997	1.308	13.704	120.71	250.78	4.843	21	2015.06.13	Mindoro,_Philippines
0	78	65	5.116	121.641	3.102	-24.504	-175.774	19.53	5.133	37	2015.06.13	South_of_Tonga_Islands
0	34	34	0.699	113.06	1.23	34.446	137.382	322.66	4.604	18	2015.06.13	Near_S._Coast_of_Honshu,_Japan
0	104	98	2.493	134.463	0.862	56.689	-156.725	34.77	5.21	84	2015.06.13	Alaska_Peninsula
0	53	50	1.731	125.087	1.592	38.851	142.607	34.77	5.027	36	2015.06.12	Near_East_Coast_of_Honshu,_Japan
0	47	32	9.439	120.563	2.042	6.173	116.677	92.19	5.064	29	2015.06.12	Borneo
0	44	33	1.378	129.993	1.811	-5.562	152.304	19.53	5.138	31	2015.06.12	New_Britain_Region,_P.N.G.
0	170	158	2.059	39.853	1.415	-15.56	-173.049	10.16	5.455	97	2015.06.12	Tonga_Islands
0	50	41	4.058	93.874	2.858	-20.805	-173.832	10.16	5.058	23	2015.06.12	Tonga_Islands
0	53	38	1.663	144.499	1.695	8.009	126.963	96.09	5.026	34	2015.06.11	Mindanao,_Philippines
0	36	26	3.828	118.986	1.688	12.902	146.272	19.53	5.179	23	2015.06.11	South_of_Mariana_Islands
0	41	33	1.943	161.243	0.814	29.73	131.359	9.96	5.032	27	2015.06.11	Southeast_of_Ryukyu_Islands
0	83	74	1.981	133.715	1.902	39.677	143.213	19.53	5.284	67	2015.06.11	Off_East_Coast_of_Honshu,_Japan
0	178	165	2.192	81.346	1.391	39.503	143.387	19.73	5.328	136	2015.06.11	Off_East_Coast_of_Honshu,_Japan

"labelled" true or false event

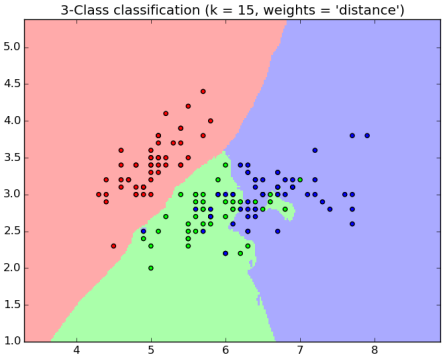
Problem: many true and few false events!

Classifier Algorithms:

Support vector machines (SVMs)



Nearest Neighbors Classification



and many more...

Multiple machine learning algorithms: train and test with past true and (few) false events

Test Target:		[1 1 1 0 0 0 0 0 0 0 0 0 0]		
Method	***Misfit***	***Result***	***Test Set***	
SVC_linear	1	[1 1 0 0 0 0 0 0 0 0 0 0 0]	A	promising
SVC_linear	1	[1 0 1 0 0 0 0 0 0 0 0 0 0]	B	
SVC_rbf	3	[0 0 0 0 0 0 0 0 0 0 0 0 0]	A	poor
SVC_rbf	2	[1 0 0 0 0 0 0 0 0 0 0 0 0]	B	
SVC_poly	3	[0 0 0 0 0 0 0 0 0 0 0 0 0]	A	poor
SVC_poly	3	[0 0 0 0 0 0 0 0 0 0 0 0 0]	B	
KNeighborsClassifier_uniform_5	3	[0 0 0 0 0 0 0 0 0 0 0 0 0]	A	poor
KNeighborsClassifier_uniform_5	2	[1 0 0 0 0 0 0 0 0 0 0 0 0]	B	
KNeighborsClassifier_distance_5	3	[0 0 0 0 0 0 0 0 0 0 0 0 0]	A	poor
KNeighborsClassifier_distance_5	2	[1 0 0 0 0 0 0 0 0 0 0 0 0]	B	
KNeighborsClassifier_uniform_10	3	[0 0 0 0 0 0 0 0 0 0 0 0 0]	A	poor
KNeighborsClassifier_uniform_10	3	[0 0 0 0 0 0 0 0 0 0 0 0 0]	B	
KNeighborsClassifier_distance_10	3	[0 0 0 0 0 0 0 0 0 0 0 0 0]	A	poor
KNeighborsClassifier_distance_10	3	[0 0 0 0 0 0 0 0 0 0 0 0 0]	B	
DecisionTreeClassifier	0	[1 1 1 0 0 0 0 0 0 0 0 0 0]	A	unstable
DecisionTreeClassifier	4	[1 1 1 1 0 0 0 1 1 1 0 0 0]	B	
AdaBoostClassifier_SAMME_50	0	[1 1 1 0 0 0 0 0 0 0 0 0 0]	A	unstable
AdaBoostClassifier_SAMME_50	2	[1 0 1 0 0 0 0 0 0 1 0 0 0]	B	
AdaBoostClassifier_SAMME_100	0	[1 1 1 0 0 0 0 0 0 0 0 0 0]	A	good
AdaBoostClassifier_SAMME_100	1	[1 0 1 0 0 0 0 0 0 0 0 0 0]	B	
AdaBoostClassifier_SAMME_200	1	[1 1 0 0 0 0 0 0 0 0 0 0 0]	A	unstable
AdaBoostClassifier_SAMME_200	3	[1 0 1 0 0 0 0 0 1 1 0 0 0]	B	
AdaBoostClassifier_SAMME.R_50	0	[1 1 1 0 0 0 0 0 0 0 0 0 0]	A	good
AdaBoostClassifier_SAMME.R_50	1	[1 1 1 0 0 0 0 0 0 1 0 0 0]	B	
RandomForestClassifier_gini_5	1	[1 1 0 0 0 0 0 0 0 0 0 0 0]	A	unstable
RandomForestClassifier_gini_5	3	[1 0 1 1 0 0 0 0 0 1 0 0 0]	B	
RandomForestClassifier_entropy_5	1	[1 1 0 0 0 0 0 0 0 0 0 0 0]	A	unstable
RandomForestClassifier_entropy_5	2	[1 0 1 0 0 0 0 0 0 0 0 1 0]	B	
RandomForestClassifier_gini_10	1	[1 1 0 0 0 0 0 0 0 0 0 0 0]	A	promising
RandomForestClassifier_gini_10	1	[1 0 1 0 0 0 0 0 0 0 0 0 0]	B	
RandomForestClassifier_entropy_10	1	[1 1 0 0 0 0 0 0 0 0 0 0 0]	A	unstable
RandomForestClassifier_entropy_10	3	[0 0 0 0 0 0 0 0 0 0 0 0 0]	B	
ExtraTreesClassifier_gini_5	0	[1 1 1 0 0 0 0 0 0 0 0 0 0]	A	unstable
ExtraTreesClassifier_gini_5	2	[1 0 1 0 0 0 0 1 0 0 0 0 0]	B	
ExtraTreesClassifier_entropy_5	2	[1 0 0 0 0 0 0 0 0 0 0 0 0]	A	poor
ExtraTreesClassifier_entropy_5	2	[1 0 1 0 0 0 0 0 0 0 0 1 0]	B	
ExtraTreesClassifier_gini_10	2	[0 1 0 0 0 0 0 0 0 0 0 0 0]	A	poor
ExtraTreesClassifier_gini_10	2	[1 0 1 0 0 0 0 0 0 1 0 0 0]	B	
ExtraTreesClassifier_entropy_10	1	[1 1 0 0 0 0 0 0 0 0 0 0 0]	A	unstable
ExtraTreesClassifier_entropy_10	2	[1 0 0 0 0 0 0 0 0 0 0 0 0]	B	

False & true events

Algorithms act in high-dimension using many attributes
→ may discover complex relationships between attributes,
→ may be difficult to understand in terms of expert knowledge & scientific theory.

Many algorithms to select and parameters to tune
→ great open software helps.

Intelligence and statistics for rapid and robust earthquake analysis, identification of false events: Conclusions

- **Statistical analysis** aids in acting on individual or few attributes,
  (e.g. stronger filtering on azimuth gaps)
  Direct use of expert knowledge & scientific theory
- **Machine learning** acts in high-dimension using many attributes:
  Powerful and shows much promise for improving early warning reliability,
  Many machine learning algorithms are very familiar in geophysics,
  and powerful new algorithms for big data, image recognition, …
  BUT, automated, not theory based
- False events: Include past event history? → Recursive Neural Networks?

- Easy to use with well documented, **open tools** in Python, R, Java, …
- What advantages and trade-offs for science?
  Machine learning, Automation ↔ Expert knowledge, Scientific Theory

Support: Centro Nazionale Terremoti, INGV
Data: ingv.it, geofon.gfz-potsdam.de, geosbud.ipgp.fr, resif.fr, ird.nc, iris.washington.edu, usgs.gov
Software: Python: scikit-learn.org, pandas.pydata.org, matplotlib.org; R statistics language

Anthony Lomax
ALomax Scientific, Mouans-Sartoux, France
anthony@alomax.net www.alomax.net @ALomaxNet

Alberto Michelini, Fabrizio Bernardi, and Valentino Lauciani
Istituto Nazionale di Geofisica e Vulcanologia, Roma, Italy