# Early warning systems as a statistical decision problem

Jonathan Rougier

Dept. Maths, Univ. Bristol

CABOT/BRISK Summer School, July 2013

◆□▶ ◆□▶ ◆臣▶ ◆臣▶ 臣 の�?

# Merapi, Indonesia



Source: http://www.ghumakkar.com/2008/04/07/jogja-volcanoes-heritage-sites/

For natural hazards with precursors, where evacuation could be used to save lives.

▲□▶ ▲圖▶ ★ 国▶ ★ 国▶ - 国 - のへで



For natural hazards with precursors, where evacuation could be used to save lives.



#### No immediate risk.

▲□▶ ▲圖▶ ★ 国▶ ★ 国▶ - 国 - のへで

For natural hazards with precursors, where evacuation could be used to save lives.



**Prepare to evacuate.** Evacuate vulnerable members of the population.

◆□▶ ◆□▶ ◆三▶ ◆三▶ 三三 のへぐ

No immediate risk.

For natural hazards with precursors, where evacuation could be used to save lives.



#### Evacuate immediately.

**Prepare to evacuate.** Evacuate vulnerable members of the population.

< □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > <

No immediate risk.

#### Features of the hazard



◆□▶ ◆□▶ ◆臣▶ ◆臣▶ 三臣 - のへで



・ロト ・聞ト ・ヨト ・ヨト

₹ 9Q@

Likely precursor values for different types of eruption



Likely precursor values for different types of eruption



Likely precursor values for different types of eruption



Likely precursor values for different types of eruption



Likely precursor values for different types of eruption

# Conditional probabilities



Conditional probability plots

These are the conditional probabilities of the four volcano states as a function of the precursors: the four panels always add up to one.

The loss function encodes the risk manager's assessment of the loss incurred for each combination of action and outcome (large is worse).

◆□▶ ◆□▶ ◆三▶ ◆三▶ 三三 のへで

The loss function encodes the risk manager's assessment of the loss incurred for each combination of action and outcome (large is worse).

#### Cost of action Qui. Small Med. Large Green 0 0 0 0 Yellow 1 1 1 1 Red 10 10 10 10

The loss function encodes the risk manager's assessment of the loss incurred for each combination of action and outcome (large is worse).

Cost of action					FalseAlarm			
	Qui.	Small	Med.	Large	Qui.	Small	Med.	Large
Green	0	0	0	0	0	0	0	0
Yellow	1	1	1	1	5	1	0	0
Red	10	10	10	10	10	10	5	0

The loss function encodes the risk manager's assessment of the loss incurred for each combination of action and outcome (large is worse).

Cost of action					FalseAlarm			
	Qui.	Small	Med.	Large	Qui.	Small	Med.	Large
Green	0	0	0	0	0	0	0	0
Yellow	1	1	1	1	5	1	0	0
Red	10	10	10	10	10	10	5	0
		Harm						
	Qui.	Small	Med.	Large				
Green	0	1	10	100				
Yellow	0	0	1	50				
Red	0	0	0	20				

◆□▶ ◆□▶ ◆三▶ ◆三▶ 三三 のへで

The loss function encodes the risk manager's assessment of the loss incurred for each combination of action and outcome (large is worse).

Cost of action					FalseAlarm			
	Qui.	Small	Med.	Large	Qui.	Small	Med.	Large
Green	0	0	0	0	0	0	0	0
Yellow	1	1	1	1	5	1	0	0
Red	10	10	10	10	10	10	5	0
Harm					Total loss			
	Qui.	Small	Med.	Large	Qui.	Small	Med.	Large
Green	0	1	10	100	0	1	10	100
Yellow	0	0	1	50	6	2	2	51
Red	0	0	0	20	20	20	15	30

All of these losses and probabilities are those of the *risk manager* who acts as the agent for her stakeholders.

➤ X is the state of nature, A the action set, and L(a, x) the consequence (loss: large is worse) of choosing action a if the state turns out to be x.

< □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > <

All of these losses and probabilities are those of the *risk manager* who acts as the agent for her stakeholders.

- ➤ X is the state of nature, A the action set, and L(a, x) the consequence (loss: large is worse) of choosing action a if the state turns out to be x.
- ➤ Y is a measurement which is informative about X, with conditional distribution f<sub>Y|X</sub>. X itself has a distribution f<sub>X</sub>.

All of these losses and probabilities are those of the *risk manager* who acts as the agent for her stakeholders.

- ➤ X is the state of nature, A the action set, and L(a, x) the consequence (loss: large is worse) of choosing action a if the state turns out to be x.
- ➤ Y is a measurement which is informative about X, with conditional distribution f<sub>Y|X</sub>. X itself has a distribution f<sub>X</sub>.
- The objective is to find the optimal rule  $\delta^* : y \mapsto a$ , where " $\delta(y) = a$ " says "If the measurements are y, then choose a".

All of these losses and probabilities are those of the *risk manager* who acts as the agent for her stakeholders.

- ➤ X is the state of nature, A the action set, and L(a, x) the consequence (loss: large is worse) of choosing action a if the state turns out to be x.
- ➤ Y is a measurement which is informative about X, with conditional distribution f<sub>Y|X</sub>. X itself has a distribution f<sub>X</sub>.
- The objective is to find the optimal rule  $\delta^* : y \mapsto a$ , where " $\delta(y) = a$ " says "If the measurements are y, then choose a".
- An elegant piece of theory shows that

$$\delta^*(y) = \operatorname*{argmin}_{a \in \mathcal{A}} \mathrm{E}\{L(a, X) \mid Y = y\}$$

where  $E\{\cdot | \cdot\}$  indicates conditional expectation; Bayes's Theorem is used to find this.

So in our case:

•  $X \in \{$ Quiescent, Small, Medium, Large $\}$ .

◆□▶ ◆□▶ ◆臣▶ ◆臣▶ 臣 のへぐ

So in our case:

•  $X \in \{$ Quiescent, Small, Medium, Large $\}$ .

◆□▶ ◆□▶ ◆三▶ ◆三▶ 三三 のへぐ

•  $\mathcal{A} = \{ \mathsf{Green}, \mathsf{Yellow}, \mathsf{Red} \}.$ 

So in our case:

- $X \in \{$ Quiescent, Small, Medium, Large $\}$ .
- $\blacktriangleright \ \mathcal{A} = \{\mathsf{Green}, \mathsf{Yellow}, \mathsf{Red}\}.$

$$L(a,x) = \begin{pmatrix} 0 & 1 & 10 & 100 \\ 6 & 2 & 2 & 51 \\ 20 & 20 & 15 & 30 \end{pmatrix}$$

◆□▶ ◆□▶ ◆三▶ ◆三▶ 三三 のへぐ

So in our case:

- $X \in \{$ Quiescent, Small, Medium, Large $\}$ .
- $\mathcal{A} = \{ \mathsf{Green}, \mathsf{Yellow}, \mathsf{Red} \}.$

$$L(a,x) = \begin{pmatrix} 0 & 1 & 10 & 100 \\ 6 & 2 & 2 & 51 \\ 20 & 20 & 15 & 30 \end{pmatrix}$$

Y = (Seismicity, Expansion). f<sub>Y|X</sub> was shown in the precursors picture;

$$f_X = (0.70, 0.20, 0.07, 0.03).$$

#### The optimal rule for our risk manager



**Optimal traffic light setting** 

Seismicity

◆□▶ ◆□▶ ◆三▶ ◆三▶ 三三 のへぐ

#### Outcomes under this decision rule

These can be found by simulation (1000 'weeks'):

	Quiescent	Small	Medium	Large	sum
Green	654	160	28	2	844
Yellow	30	46	39	16	131
Red	0	1	5	19	25
sum	684	207	72	37	1000

# Solving the wrong problem

 Undoubtedly the hardest part is specifying the loss function. (The probability distributions f<sub>Y|X</sub> and f<sub>X</sub> can often be approximated using historical relative frequencies.)

# Solving the wrong problem

- Undoubtedly the hardest part is specifying the loss function. (The probability distributions f<sub>Y|X</sub> and f<sub>X</sub> can often be approximated using historical relative frequencies.)
- In this case it is tempting to propose a 0-1 loss function, which is simply a classifier,

		Loss		
	Qui.	Small	Med.	Large
Green	0	0	1	1
Yellow	1	1	0	1
Red	1	1	1	0

i.e. which describes the state of the volcano without any thought to the losses which follow.

# Solving the wrong problem

- Undoubtedly the hardest part is specifying the loss function. (The probability distributions f<sub>Y|X</sub> and f<sub>X</sub> can often be approximated using historical relative frequencies.)
- In this case it is tempting to propose a 0-1 loss function, which is simply a classifier,

		Loss		
	Qui.	Small	Med.	Large
Green	0	0	1	1
Yellow	1	1	0	1
Red	1	1	1	0

i.e. which describes the state of the volcano without any thought to the losses which follow.

The effect is to choose the action/state which has the highest conditional probability.

# Solving the wrong problem (cont)

Optimal traffic light setting (0-1 loss)



Seismicity

◆□ > ◆□ > ◆臣 > ◆臣 > ─ 臣 ─ のへで

# Solving the wrong problem (cont)



・ロト ・聞ト ・ヨト ・ヨト

æ

#### Reflections

- 1. The statistician does not 'solve' this decision problem. Instead, he guides the risk manager, allowing her to explore her judgements about the hazard and about the loss.
- 2. Hazard process probabilities can often be assessed from the historical record, guided by experts.
- 3. Losses are much less well-specified. Undoubtedly there will be some negotiation between the risk manager and her stakeholders. But it is much better to be negotiating about the loss table than about larger and less well-specified aspects of the early-warning system.