Verifying “Extreme Rainfall Alerts” for surface water flooding potential

Marion Mittermaier, Nigel Roberts and Clive Pierce

Pluvial (surface water) flooding

Flash floods

Why “Extreme Rainfall Alerts” ?

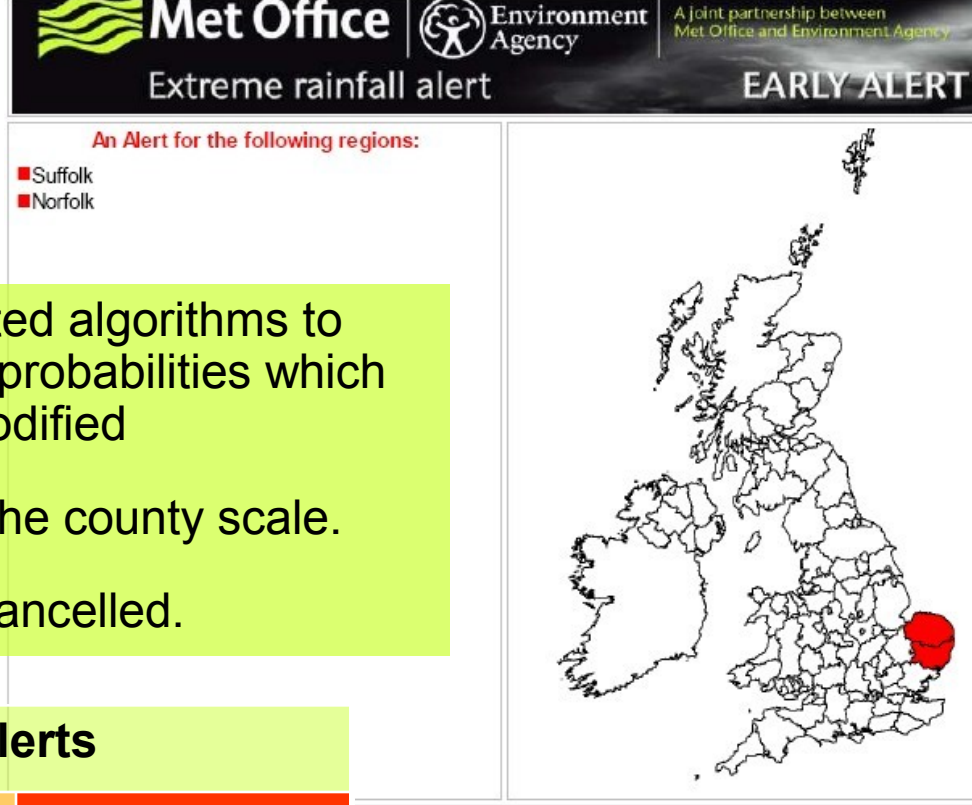
- **Much of the damage in the July 2007 UK floods was caused by surface water flooding.** The Environment Agency and the Met Office are working together to improve understanding and response to this flood risk.

- **Surface water flooding occurs as a direct result of extreme rainfall.** It differs from river flooding as it can happen before water enters a river or watercourse or where no river or watercourse exists.

- **Advance warning is difficult as it can happen very quickly** when the level of rainfall exceeds drainage capacity. Its impact is highly dependent on local landscapes and local conditions such as blocked culverts.

What are they?

- Based on sophisticated algorithms to generate first-guess probabilities which can be forecaster-modified
- Alerts are issued at the county scale.
- Can be updated or cancelled.



	Extreme Rainfall Alerts		
	Advisory	**Early**	**Imminent**
30 mm/h or 40 mm/3h or 50 mm/6h	Very low but prob >= 10%	Low with prob 20-40%	Moderate with prob >= 40%
Issued	14LT valid for the 24h starting from the next midnight	Lead time of 8 – 11h	Lead time of 1-3h

Difference between the National Severe Weather Warnings and Extreme Rainfall Alerts?

- **NSWWS warnings are not designed to warn for surface water flooding.** ERA thresholds are specifically set to relate to the risk of surface water flooding (based on 30 mm rainfall per hour in urban areas);
- Because **rainfall thresholds leading to surface water flooding are higher** and there is **more uncertainty in forecasting rainfall quantities of high intensities**, the probabilities assigned to the ERA thresholds will be necessarily low compared to NSWWS;
- ERAs will be updated on regular basis, the NSWWS Heavy Rainfall Warning is only issued once. (Note: talks by Michael Sharpe, Clive Wilson and David Stephenson on verification of conventional warnings).

So how to verify these alerts?! Comparing two options…

Two approaches have been considered …

- taking the **"event"** view, and

  (did an event occur anywhere in the alert area during the time that the alert was in force)

- taking the **"time series"** (continuum) view

  (comparing the county accumulation totals hour-by-hour during the time that the alert was in force to establish if the threshold was exceeded)

Caveat: both of these approaches are inherently deterministic

Taking the “event” view

Defining an event

Events are defined as an occurrence of one of the ERA thresholds being exceeded, and whether or not there was an alert (i.e. it is not conditioned on just alerts or events)

Events have been split into two types:

- **An event** - it is extremely likely that an ERA threshold was exceeded.
- **A 'possible' event** - it is distinctly possible that an ERA threshold was exceeded but there is insufficient evidence to class as an event (or no event).

The distinction between **events** and **possible events** was necessary due to the **uncertainties in radar estimates** (especially on a coarse 5km grid).

Sometimes there were **several events** (for different geographical areas) or perhaps **several warnings** given (for different geographical areas) on the same day.

Events contingency table

Based on the pilot data for July-August 2008

Contingency table for: Advisories / Early / Imminent		Observed				
		Yes event occurred (possible events excluded)	Yes event occurred (possible events included)	No event occurred (possible events included as no)	No event occurred (possible events not included as no)	Total
Forecast	Advisory	a = 4	a' = 7	b = 12	b' = 9	16
	Early issued	a = 6	a' = 10	b = 13	b' = 9	19
	Imminent	a = 5	a' = 6	b = 4	b' = 3	9
	No advisory	c = 16	c' = 38	Can't be determined with this framework	Can't be determined with this framework	Can't be determined with this framework
	No early issued	c = 14	c' = 35			
	No imminent	c = 15	c' = 39			
Total		20	45'	Can't be determined with this framework	Can't be determined with this framework	Can't be determined with this framework

The "missing d's"

Events categorical statistics

Green=encouraging; red=cause for concern

Type	Hit rate (H)	Hit rate (H')	False Alarm Ratio (FAR)	False Alarm Ratio (FAR')	1-FAR	1-FAR'	CSI	CSI'
Advisory	0.2	**0.16**	0.75	0.56	**0.25**	**0.44**	0.13	0.13
Early	**0.3**	**0.22**	0.68	0.47	0.32	0.53	0.18	0.19
Imminent	**0.25**	**0.13**	0.44	0.33	**0.56**	**0.67**	**0.21**	**0.13**

H = # events for which alert was issued
H' = # events and possible events for which an alert was issued
FAR = # alerts for which there was no event
FAR' = # alerts for which there was no event or possible event
(1-FAR) = proportion of alerts (excluding possible events) that were correct
(1-FAR') = proportion of alerts (including possible events) that were correct
CSI = # occasions when either an alert was issued or an event occurred and forecast was correct
CSI' = # occasions when either an alert was issued or an event or possible event occurred and forecast was correct

A CSI of 0.13 for the advisories seems low but …

… based on the fact that advisories are issued on the basis of a **10% chance** then we can consider the following ideal situation:

b = a * 9 because 9 / 10 times nothing should happen, i.e. b should be 9 x greater than a;

c ~ 0 because in an ideal situation we should not miss any events, i.e. all events are warned for;

then **CSI ~ a / (a + 9 * a) ~ 0.1**

Therefore the value of **0.13 is in keeping with expectation** (given the 10% probability) and not as a bad a score as the deterministic framework would suggest (although the balance of a, b and c may not be as it should be).

Taking the time series view

Using a “unit” of time

- Max accumulations per county were extracted from the 2-km radar running **hourly, three hourly and six-hourly** accumulations
- Results are calculated at the **county scale**.
- A **precise matching** in both space and time was applied.
- A **3 x 2 contingency table** is compiled using two thresholds, **T1 (lesser) and T2**. The T1 threshold is used to differentiate between *hits* and *near hits*, *close misses* and *misses*.

The criteria can be summarised as follows:

Hit = rainfall accumulation >= T2 and alert issued

Near hit = rainfall accumulation >= T1 but < T2 and alert issued

Close miss = rainfall accumulation >= T1 but < T2 with no alert issued

Miss = rainfall accumulation >= T2 but no alert issued

False alarm = rainfall accumulation < T1 but alert issued

Correct non-event = rainfall accumulation < T1 and no alert issued

Observations “uncertainty”

The two-tier thresholds enable the introduction of “uncertainty” to the analysis.

	Hourly	*Three hourly*	*Six hourly*
T1	20 mm/h	30 mm/3h	40 mm/6h
T2 (for which the alert is issued)	**30 mm/h**	**40 mm/3h**	**50 mm/6h**

The T2 threshold could have been exceeded …

An event could have occurred with a lesser accumulation …

Time series categorical statistics

Summed over all counties

Met Office

Observations uncertainty

30 mm/h

	Hits	Near hits	Close miss	Miss	False alarms	Correct non-events
Advisories*	28	76	482	181	4823	226339
Early alerts	6	25	533	203	1427	229735
Imminent alerts	1	12	546	208	131	231031

d's not missing but ***HUGE***!

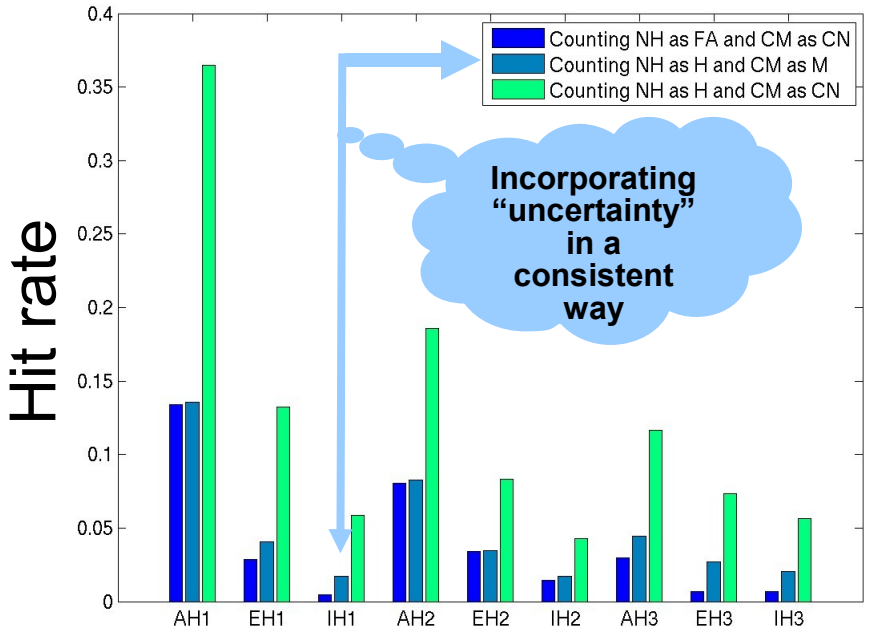
40 mm/3h

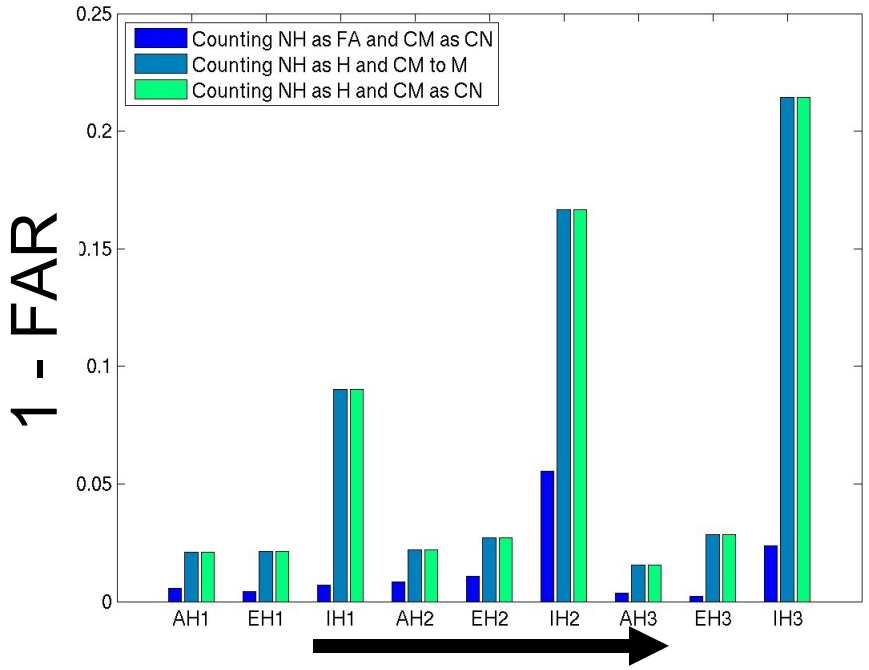
	Hits	Near hits	Close miss	Miss	False alarms	Correct non-events
Advisories*	33	53	575	377	3816	215795
Early alerts	14	22	606	396	1284	218327
Imminent alerts	6	12	616	404	90	219521

50 mm/6h

	Hits	Near hits	Close miss	Miss	False alarms	Correct non-events
Advisories*	9	30	546	295	2447	294056
Early alerts	2	22	554	302	823	195680
Imminent alerts	2	16	560	302	66	196437

All days analysed for all threshold options



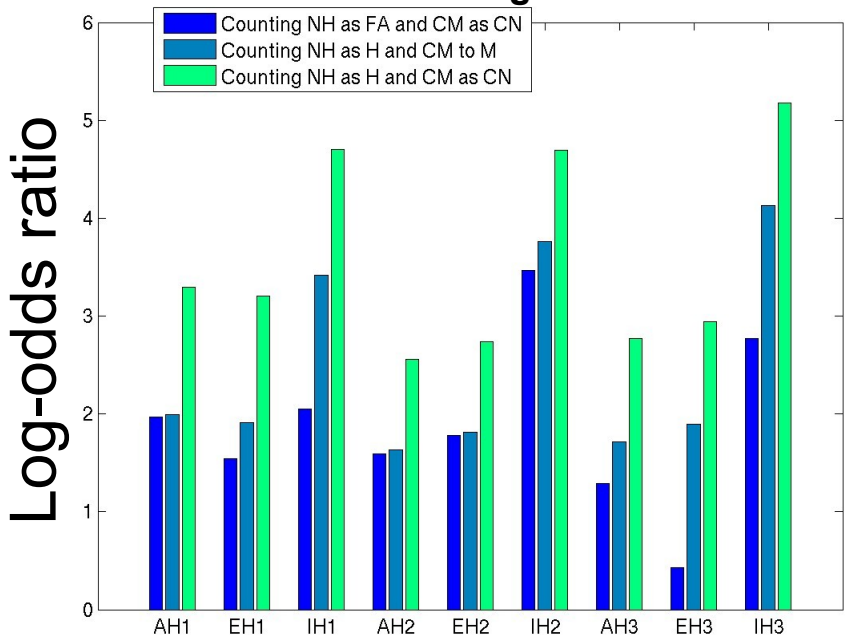


- For hit rate:
advisory > early > imminent

- For proportion correct and log-odds ratio: **imminent > early > advisory**

- Larger accumulations seem to score marginally better

- Scores are lower than for the event-based analysis but trends the same



General conclusions

- **The two approaches provide broadly similar results:**
  - e.g. in agreement that **too few advisories and alerts** were being issued; **forecasters are too conservative** (worrying about false alarms)
- There are also **some notable differences**:
  - the more detailed analysis suggests many **more events missed**; but also higher false alarms;
  - the **scores** for the time series analysis are **much lower**
- **Neither of the methodologies is ideal**:
  - **too deterministic** (whilst they are actually probabilities, warnings should be probabilistic)
  - **precise matching** in space and time, no credit given for "close" forecasts
  - either incomplete contingency tables (those **missing d's**) or an **overwhelming number of non-events**

General conclusions 2

- The answer probably lies in using:
  - **greater spatial uncertainty and/or eliminating the trivial non-events**
  - **changing the way we define the unit of comparison**
  - e.g. presentations by Michael Sharpe illustrating our new warnings verification framework and David Stephenson's talk
- We have not attempted to address how we would routinely (and objectively) verify the **occurrence of surface water flooding** in conjunction with the rainfall thresholds having been exceeded:
  - this is very difficult and thus far has been very subjective, based on anecdotal evidence from the public;
  - in the end the occurrence of **surface water flooding is not absolutely tied to the rainfall exceedance thresholds** (but also dependent on local conditions) which is why an element of uncertainty must form part of the verification strategy.

Questions and answers