

Guidelines on Forecast Verification

To trigger action based on a forecast...

We need to know if the forecast is any good. Specifically, we need some sort of evidence that the forecast can give an indication about the likelihood of an event that we are worried about.

In a perfect world...

We would assess this using a lot of data. There are three critical types of data needed:

1. Historical forecasts
2. Historical observations (what actually happened)
3. Historical data on disaster impacts

The more data that we have for the forecast evaluation, the more confident we can be that we understand the skill of the forecast.

The problem is that disasters are rare, by definition, so we need many years of data to have a record of enough disasters that we can evaluate whether they can be forecasted.

We compare data to understand forecast skill

First step: compare historical hydrometeorological observations (e.g. river levels) with historical impact data (e.g. flood events), and identify a reasonable threshold of the hydrometeorological event that can be considered a disaster. In many cases, this is just a simple threshold, but it can also be a set of probabilities, as in Figure 1.

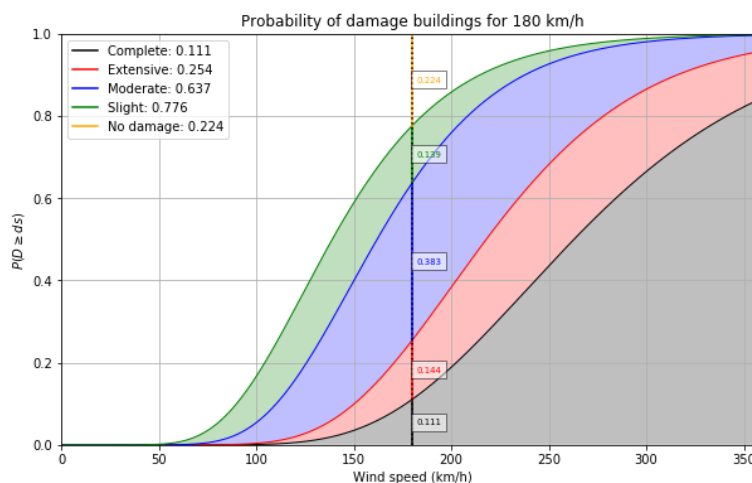
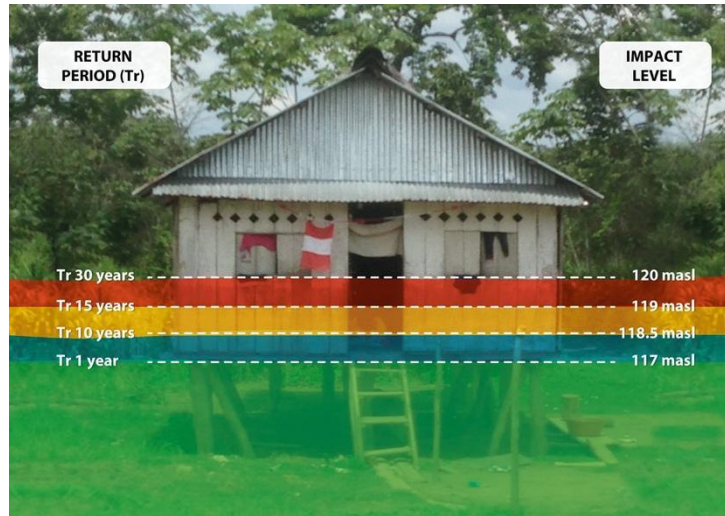


Figure 1. The colour-coded numbers adjacent to the vertical line indicate the probability of this type of building being in the given damage state. At 180 km/h, there is a 22.4% probability of no damage to buildings of this type. There is a 13.9% probability of only slight damage, 38.3% probability of moderate damage, 14.4% probability of extensive damage and finally, just over 11% probability that the building type will be completely destroyed.

This “disaster” event should not happen very frequently. It should have a return period of about 1 in 5 years. A sufficient explanation would be for example to demonstrate that there were approximately 3 events in the last 15 years in an area of approximately 50,000km².



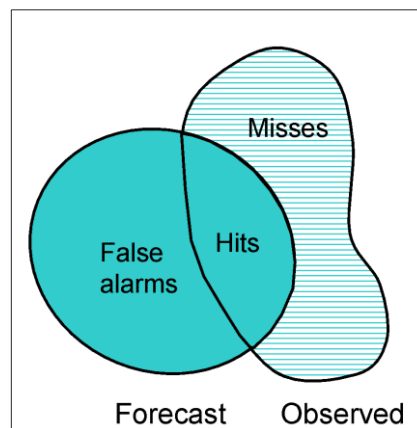
Source : FbF Peru

Second step: compare the historical forecasts with the historical observations of events, and make a contingency table for how the forecast performs at different lead times.

	Event	No event
Forecast yes	Hit	False alarms
Forecast no	Misses	Correct negatives

From this table, you can see how frequently we would expect to trigger action (hits plus false alarms), and understand the frequency of false alarms. You can calculate the False Alarm Ratio from this.

There are of course more complicated and interesting ways to evaluate a forecast, but for the purposes of an EAP, this simple contingency table is sufficient.



Source: 4th Int'l Verification Methods Workshop, Helsinki, 4-6 June 2009

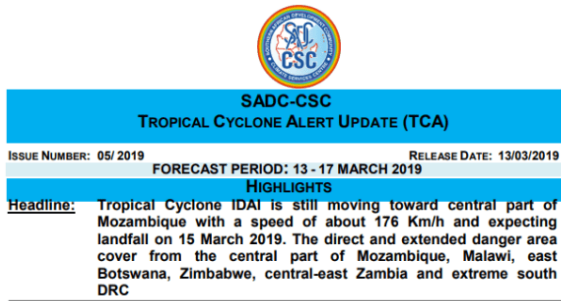
In an imperfect world...

There is very little data available. Here are some of the creative ways that people have carried out a forecast evaluation with very little data:

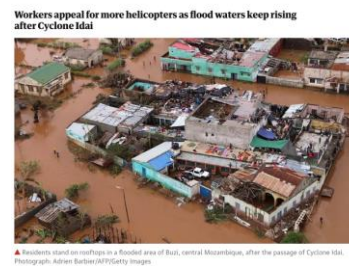
No impact data? A team in Uganda looked through historical newspapers for flood events and used those records to understand when disasters happened in the past.



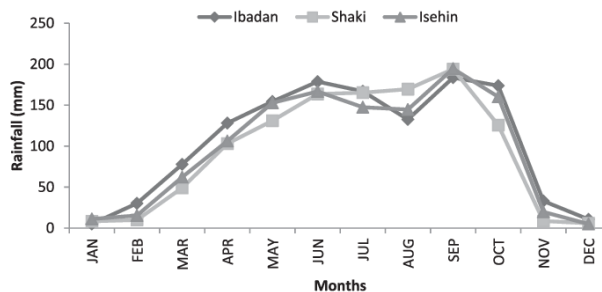
No observations? People have compared the historical forecasts directly with disaster events, to see if they aligned. Even if we don't know whether the forecast predicted the flood levels (because we don't have a record of the river levels themselves), maybe the forecast gave a clear warning of an extreme event before most flood disasters, which could be enough to show that the forecast is skillful.



Mozambique rescue teams struggle to save thousands

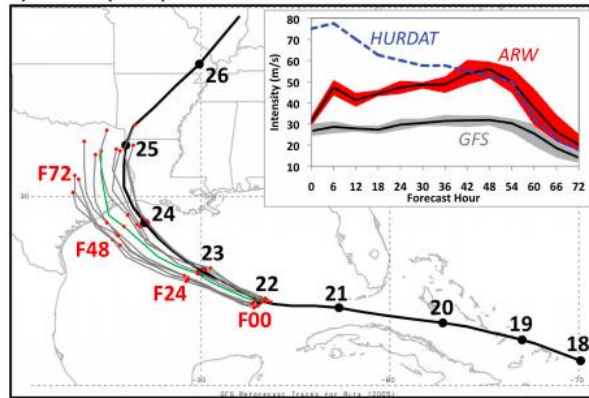


No local data? People have used a nearby location that has similar characteristics to do the forecast evaluation, assuming that the forecast will work similarly in the region with no local data.



No historical forecasts? Scientists can run today's forecast models in the past to create "re-forecasts" and simulate what we might have forecasted in past events.

a) TC Rita (2005) 72-h GFS Ensemble Reforecast



Source: Thomas Galareau NCAR

No reforecasts? People have used archived alert messages that were given to the public to make a record of when the forecast was predicting an extreme event.

Avisos Meteorológicos a nivel nacional

Los Avisos Meteorológicos son pronósticos de carácter preventivo ante eventos severos, indicando las áreas que podrían verse afectadas y el nivel de peligrosidad.

Mostrar 90 registros

Aviso	Nro.	Emisión	Inicio	Fin	Duración	Nivel
PRECIPITACIONES EN LA SIERRA NORTE Y CENTRO	148 (vigente)	2019-10-27	2019-10-28	2019-10-30	57 Hrs.	NIVEL 3
INCREMENTO DE VIENTO EN LA COSTA	147 (vigente)	2019-10-26	2019-10-28	2019-10-30	54 Hrs.	NIVEL 2
PRECIPITACIONES EN LA SIERRA NORTE Y CENTRO	146	2019-10-20	2019-10-22	2019-10-25	72 Hrs.	NIVEL 4
LLUVIA EN LA SELVA	145	2019-10-19	2019-10-21	2019-10-23	54 Hrs.	NIVEL 4
PRECIPITACIONES EN LA SIERRA NORTE Y CENTRO	144	2019-10-19	2019-10-20	2019-10-22	48 Hrs.	NIVEL 3
INCREMENTO DE VIENTO EN LA COSTA	143	2019-10-18	2019-10-19	2019-10-22	48 Hrs.	NIVEL

Source: Peruvian Meteorological and Hydrological Service

Nothing? This is a hard stop. We can't use an untested forecast.



What is not useful:

Generalized “Accuracy” estimates. Many forecasters have a general number that they use to refer to how “accurate” their forecast is. For example, this temperature forecast could be said to have 80% accuracy:

	Mon	Tues	Wed	Thurs	Fri	Sat	Sun	Mon	Tues	Wed
Forecast	20	20	20	20	20	20	20	20	20	20
Actual	20	20	20	20	20	35 (heat wave)	35 (heat wave)	20	20	20

But the forecast completely missed the heatwave on Saturday and Sunday. Perhaps this forecast is only skillful when forecasting low temperatures. We need a specific forecast evaluation for extreme events to know whether the forecast can capture the disasters and give us a warning before they happen.

Forecast evaluation for unrelated weather phenomenon.

For example, imagine we are working on floods, and we have the following forecast of the total rainfall in a city for a whole year.

Year	2015	2016	2017
Forecast of total rainfall in the year	200	300	200
Actual total rainfall in the year	200	300	200

This could be an absolutely perfect forecast, but has nothing to do with floods. We need a forecast of the actual extreme event, not some other weather phenomenon.

For more information on the principles behind forecast evaluation, check out this powerpoint of [“Is this a “good” forecast?”](#)