Event-based verification

HEPEX, Melbourne, 2018

Schalk Jan van Andel

IHE Delft Institute for Water Education

s.vanandel@un-ihe.org

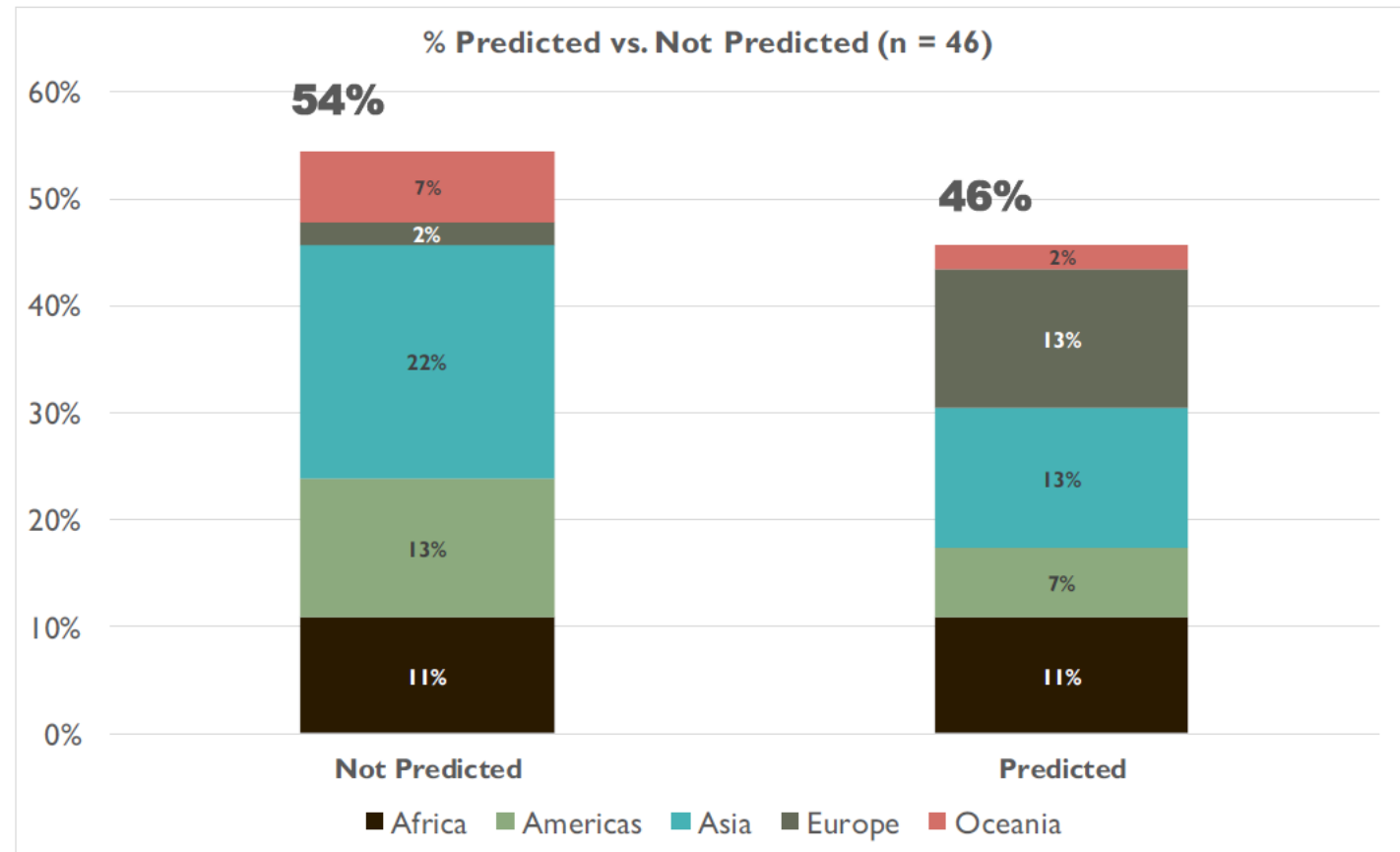# Flood early warning: How are we doing today?

- Youth Challenge: organized by UNESCO-IHP secretariat of the Netherlands, Sandra de Vries

- Case submitted to challenge participants to find/select 50 flood events in the passed 5 years, and then

- Find out whether these floods were predicted and/or warned for

Participants (and case owners) in their free time, young professionals, 6 weeks to complete the challenge

For about 50% of the events, a warning could be found online

• Note, language bias

Many thanks to
Anouk van Stokkum
Gerben Dekker
Margot Leicher

% Predicted vs. Not Predicted (n = 46)

54%

46%

	Not Predicted	Predicted
Oceania	7%	2%
Europe	2%	13%
Asia	22%	13%
Americas	13%	7%
Africa	11%	11%

■ Africa ■ Americas ■ Asia ■ Europe ■ Oceania

Do we feel pleasantly surprised, or do we feel challenged?

Back to verification..

Discussion on event-based verification, from what I think would be a user's perspective..

Verification

- Equal interval (e.g. daily, hourly)
- Event-based (not analysis of one event, but events being the unit to score a continuous series of forecasts)

Verification purposes

- Research (measuring progress, comparison)
- Selection of forecast products / methods
- Considering for use in water management, and how (e.g. decision guidance/rules)

Focus on the third purpose, where any difference between fixed-interval and event-based verification would matter most (I assume).

(for verification purposes 1. and 2. what matters most is to consistently apply the same verification method)

Flood early warning

Metrics for end user to go/no go ensemble (or any forecasting product)

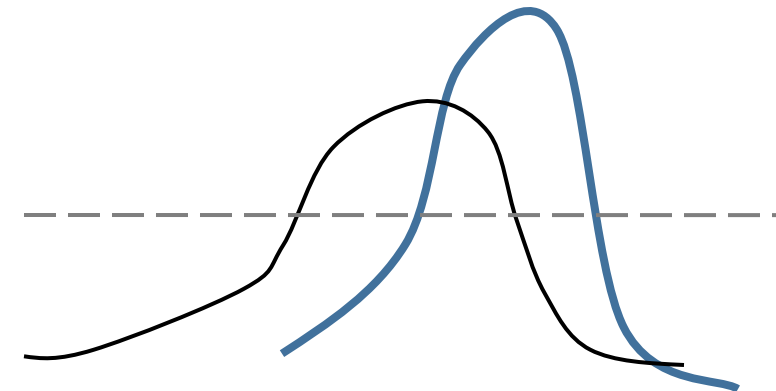- Is it going to help me make more winning decisions
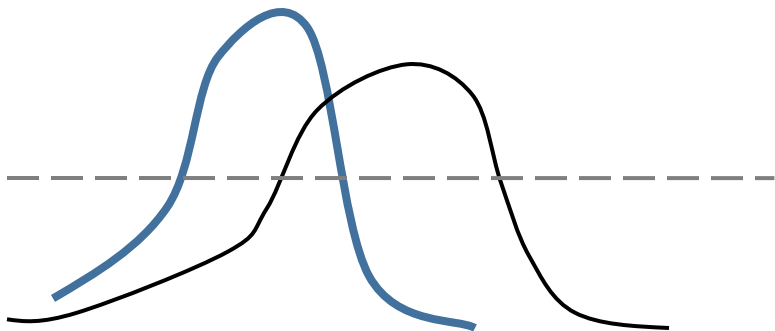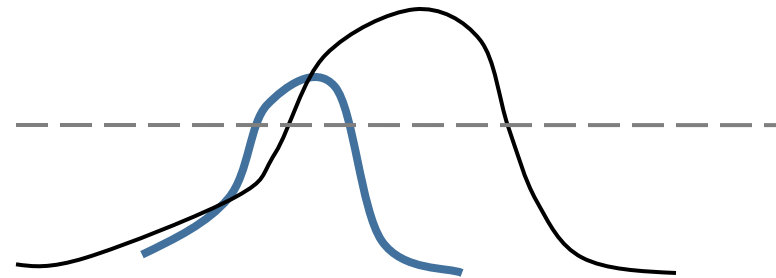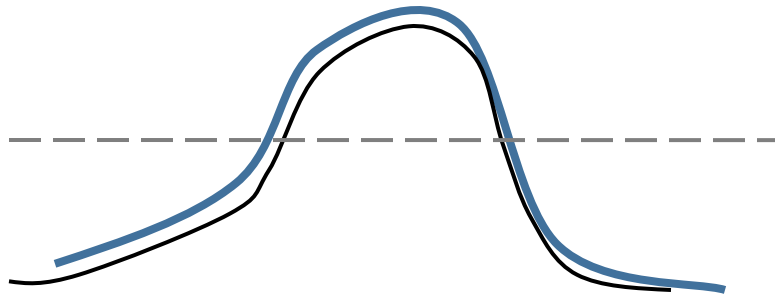- Hit rate
- False alarm ratio

Compared to what I am using now

(Cost-loss-benefit? (minimize risk?))
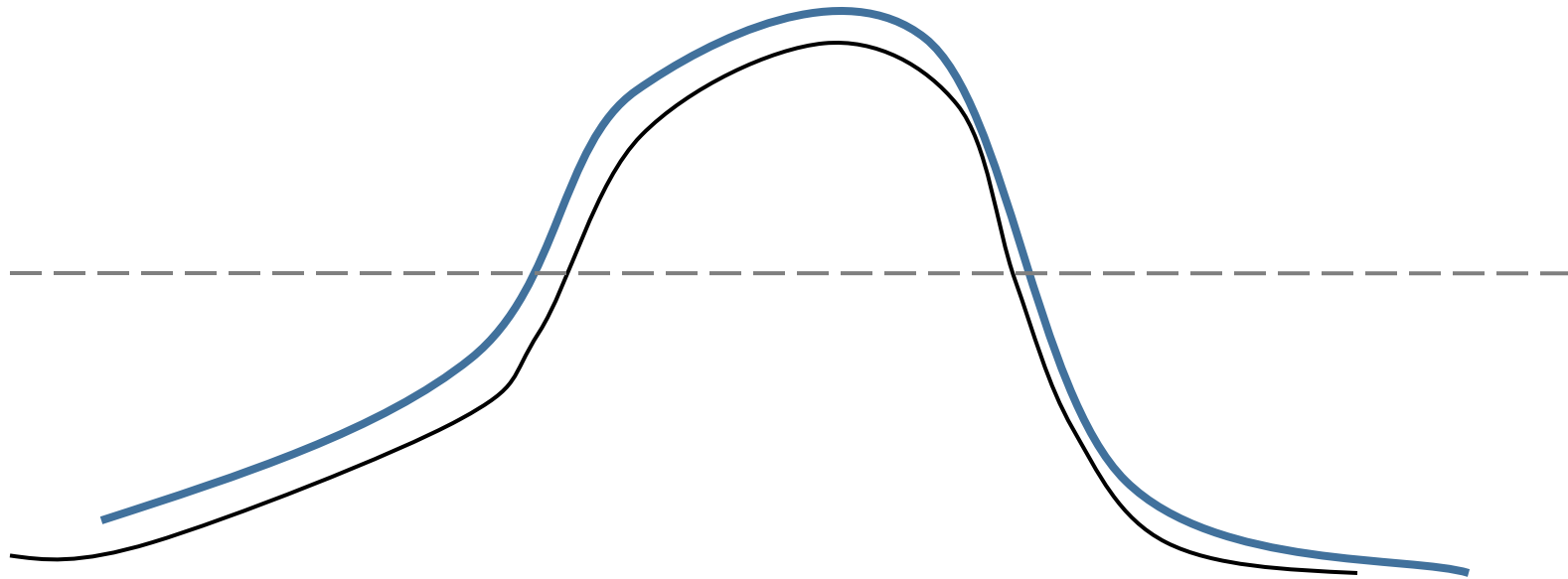
Contingency tables

Defining events and classifying forecasts

- Threshold exceedances: hits, misses, false alarms, correct rejections

Defining events and classifying of forecasts

- But how many hits?

Defining events and classifying of forecasts

- But how many hits?
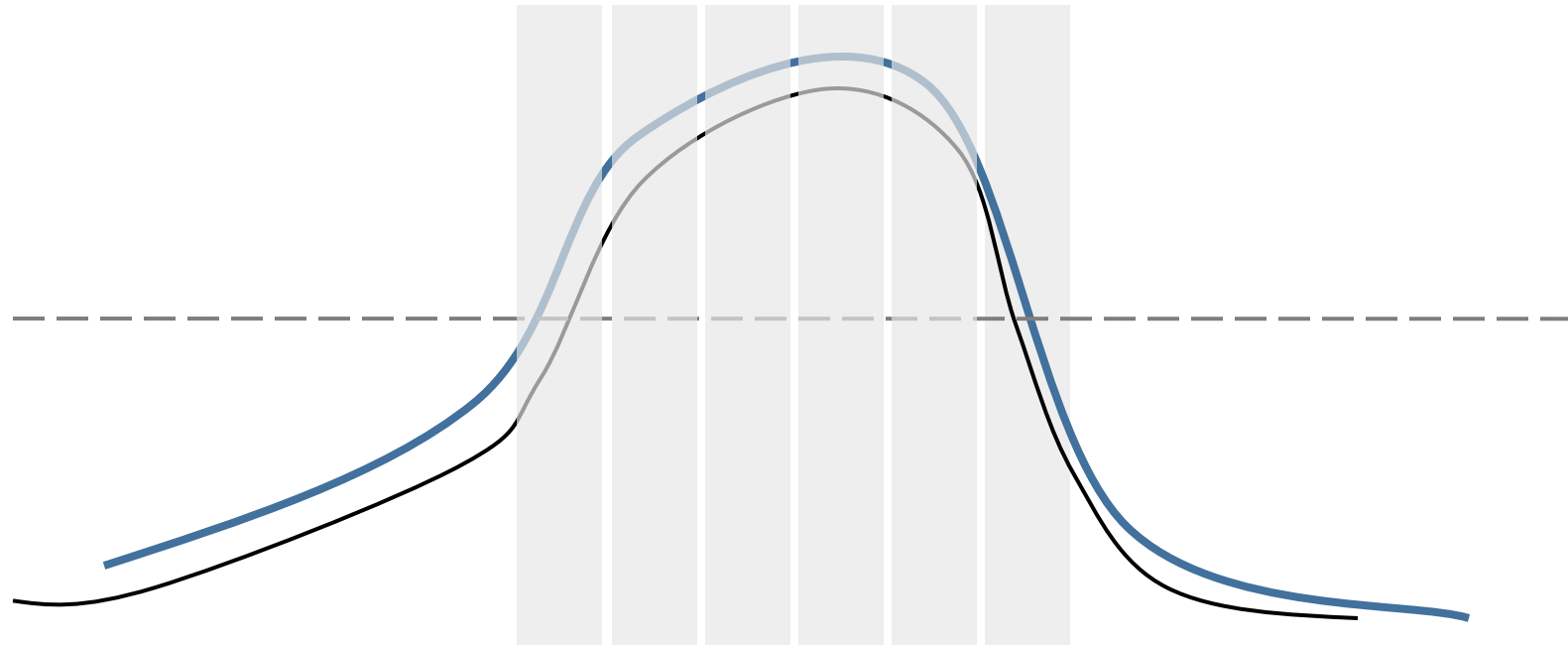
- Equal interval verification: can be multiple
- Event-based: 1

Event-based verification

- Not as straight-forward or 'pure' as fixed interval

- Not used much.. correct me if I'm wrong!

- So good excuse to dive in
  - Proposing methods of event-based verification
  - Case study
  - Discussion

Case: Rijnland

- Hourly observed, 6hourly ECMWF EPS forecasts, through KNMI
- 24h accumulated rainfall thresholds for early alert
- 3-day forecast horizon
- 12h before start of event: 'too early'
- 6h after beginning of event 'too late' for 'hit'
- If overlap between forecast and observed event, then '..too early or too late'
- No overlap: 'missed event'

Defining the non-event

- Fixed interval, easy, take the same?
- Average duration of events
- Daily

Results – fixed interval 6h – 10 mm/24h

Analysis period from 01/10/2014 to 01/01/2017.
That's 823 days, 3292 forecasts (6h interval)

10%	Observed	Not observed
Forecast	196	226
Not forecast	51	2819

Results – fixed interval 6 hours

10 mm / 24 hours

Number of events in passed two years   247

Number of alerts you would have received   422

User would be surprised... did not record so
many events

Results – Event based – 10mm/24h

10%	Observed	Not observed
Forecast	42	37
Not forecast	24	720

Number of events in passed two years: 66

Number of alerts you would have received: 79

..66 events more realistic number

Results compared

Fixed interval

10 mm / 24 hours

Number of events in passed two years: 247

Number of alerts you would have received: 422

10%	Observed	Not observed
Forecast	196	226
Not forecast	51	2819

Hit rate: **79%**

False alarm ratio: 54%

Event-based

10 mm / 24 hours

Number of events in passed two years: 66

Number of alerts you would have received: 79

10%	Observed	Not observed
Forecast	42	37
Not forecast	24	720

Hit rate: **64%**

False alarm ratio: 47%

More extreme events

Fixed interval

20 mm / 24 hours

Number of events in passed two years: **44**

Number of alerts you would have received: 88

5%	Observed	Not observed
Forecast	23	65
Not forecast	21	3183

Hit rate   52%

False alarm ratio   74%

Event-based

20 mm / 24 hours

Number of events in passed two years: **17**

Number of alerts you would have received: 25

5%	Observed	Not observed
Forecast	8	17
Not forecast	9	789

Hit rate   47%

False alarm ratio   68%

Discussion

- When doing event-based verification, duration of alerts (hits and false alarms) should be checked. In the case presented here, durations were consistent with observed duration (0.5 to 1 day on average)

- In the set-up presented here, event-based verification too optimistic on false alarms

- Classification of events subjective, e.g. rules on hit or miss

- Good thing about this approach is that the subjective choices can and have to be made by / together with the intended user

Discussion

- **Equal interval verification** for flood early warning: **too optimistic on hitrate, so not good to present to end users**

- **Let's do also event-based, continuous, verification**: Subjective, but that's good for discussion with user

- Differences depend heavily on the case study, i.e. Duration of extreme events that occur, decision process

- For flood early warning decision processes, often, persistence from one forecast to the next, with stepwise reduced forecast horizon, will have to be valued

- Presenting contingency tables is the best way of shaking up discussion between end user and provider

Discussion

- Who has done event-based verification?
- Event-based verification also relevant for product/method selection? and for R&D?
- Put event-based verification in verification software, e.g. the EVS?
    - Tom, would it fit in the toy box?

HEPEX has to keep pushing the ongoing paradigm shift from deterministic to ensemble prediction – e.g. by organising more activities by, for, and with end users