Global Flash Flood Forecasting from the ECMWF Ensemble

Calum~~n~~ Baugh, Toni Jurlina, Christel Prudhomme, Florian Pappenberger

calum.baugh@ecmwf.int

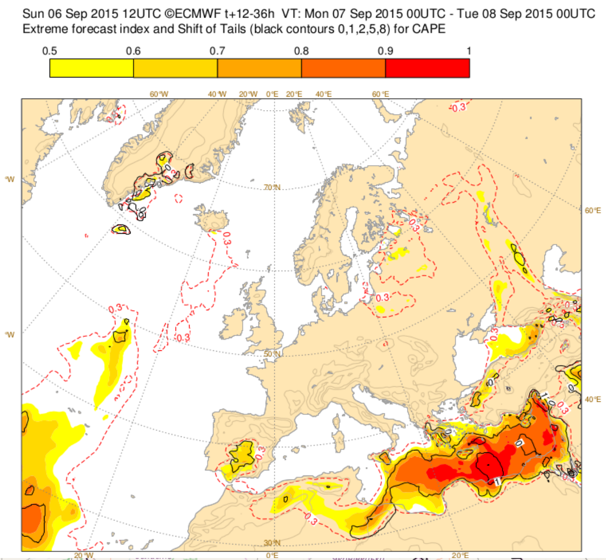
Building a Global FF System

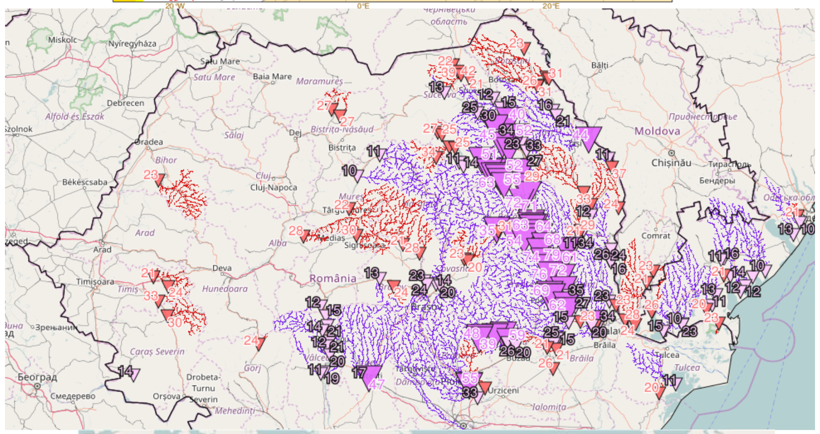
1. **Current flash flood forecasting globally**

2. **Flash flood forecasting from ECMWF ENS**
   - **Extreme Forecast Index (EFI)**

3. **Verification**

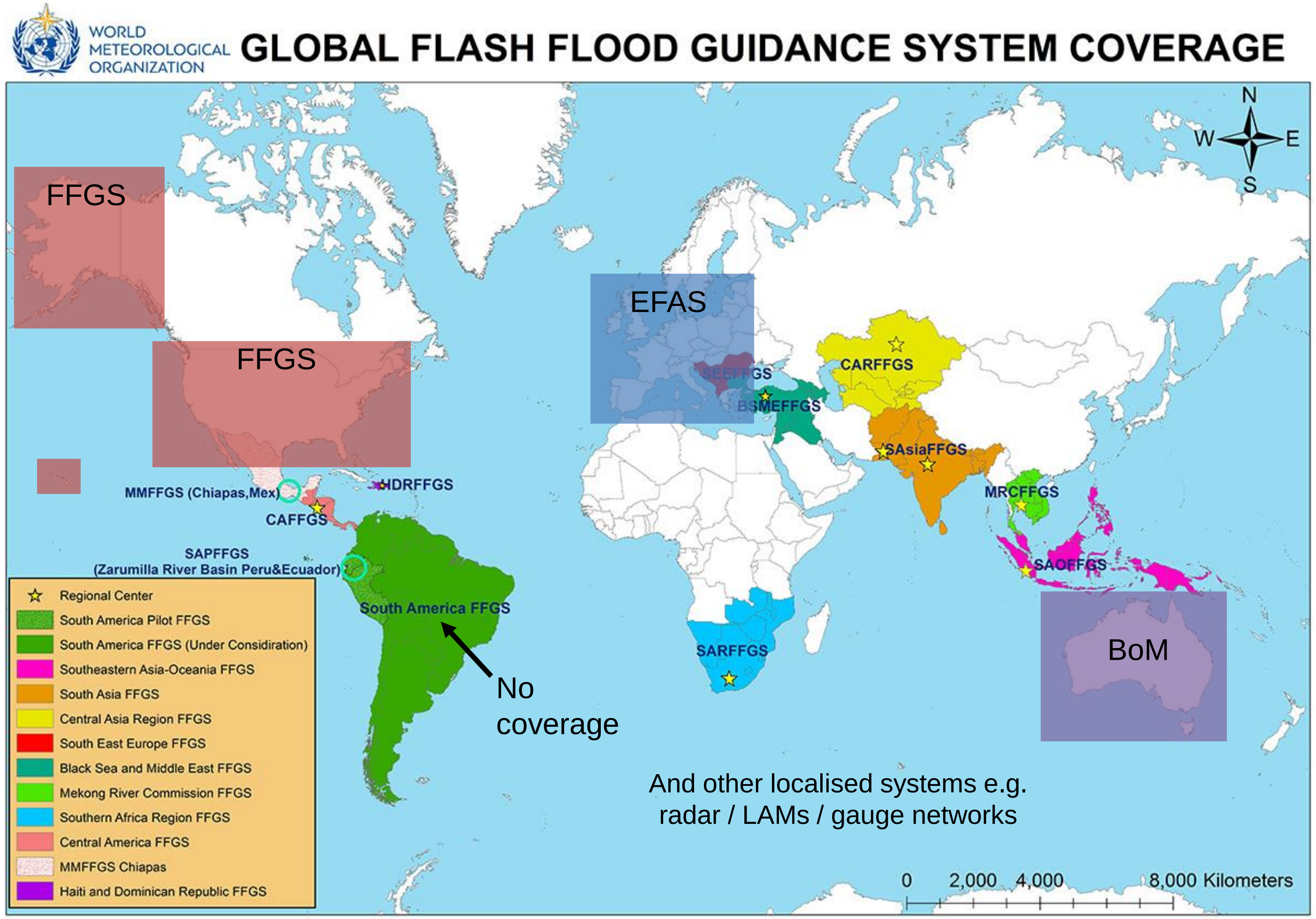
4. **Enhancing forecasts with population exposure**





Current Flash Flood Systems Globally

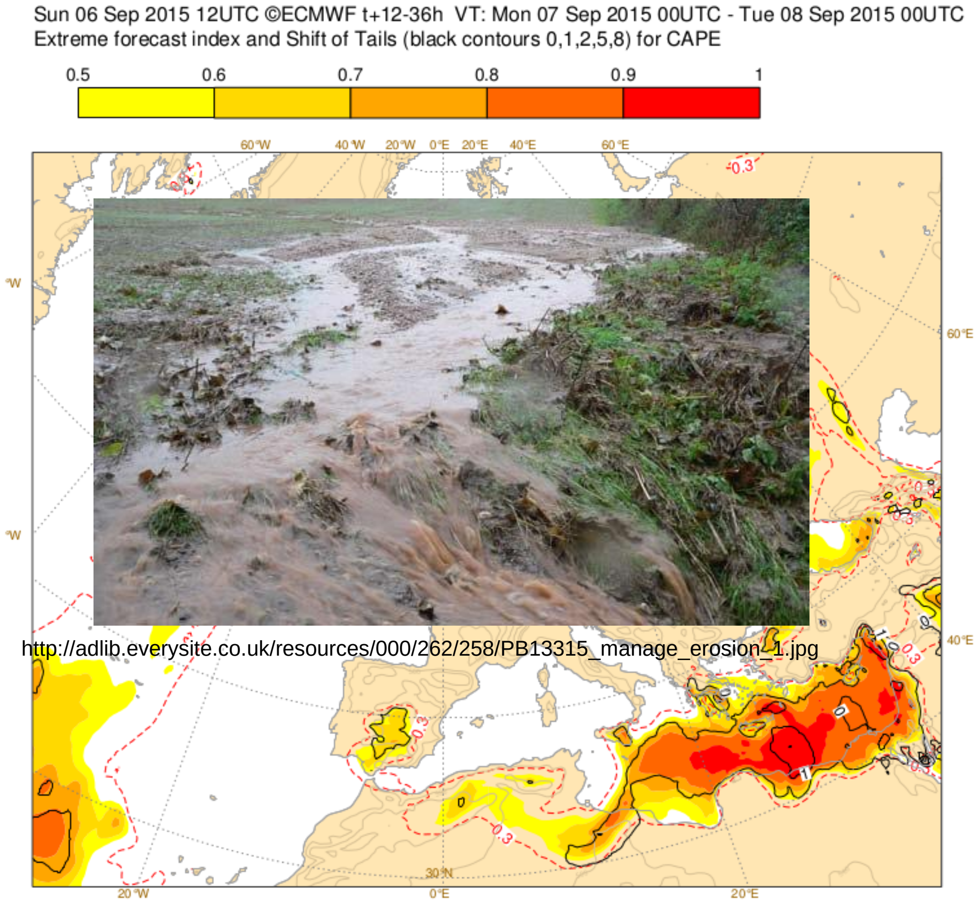
- Single global system not present
- Instead a piecemeal approach regionally / locally – important?
  - Opportunity for users to tailor their own systems
  - Or contributes to high risk associated with flash floods?



Adapted from: http://www.wmo.int/pages/prog/hwrp/flood/ffgs/images/FFGS-global-coverage14_12_2016-full.jpg

Forecasting Flash Floods

- What type of flash flood?
  - Pluvial/rapid response rivers/’nuisance’
  - Want to represent as many as possible
- What variable should we forecast?
  - River discharge
  - Atmospheric – Precipitation/CAPE
    - Precip’ principal driver?
  - Surface runoff
    - Precip’ conditioned by land surface
- Build system using precipitation and surface runoff

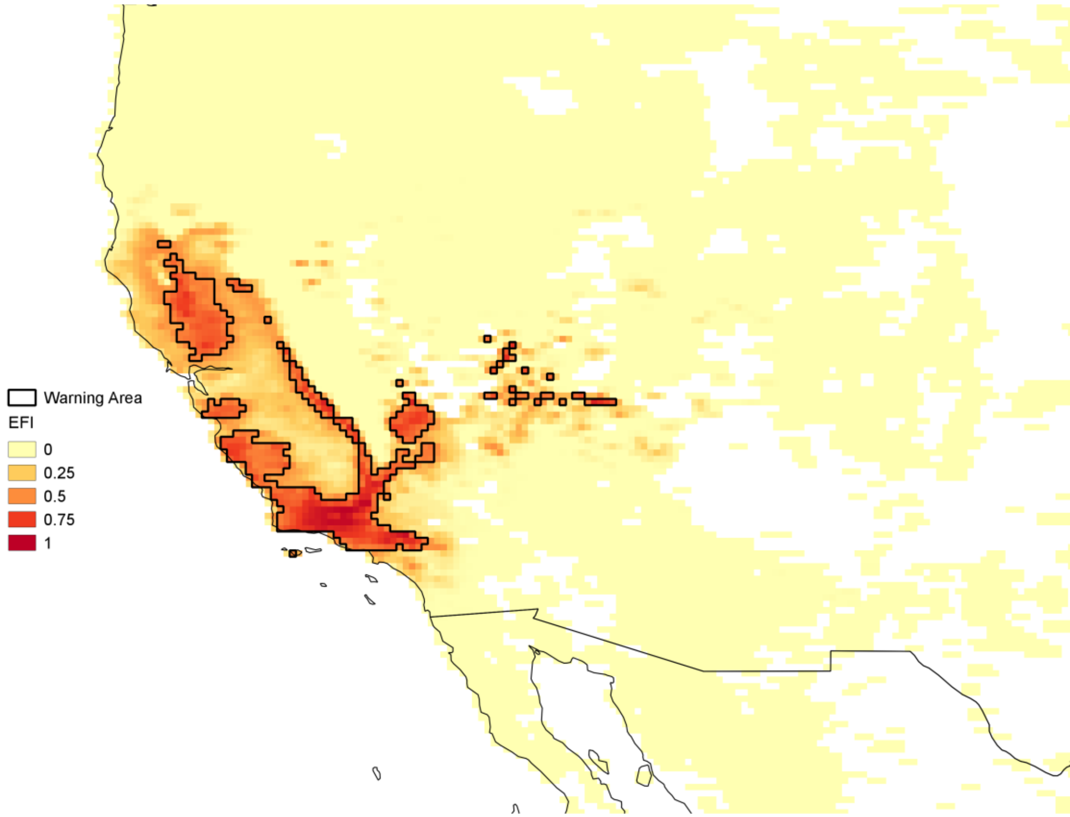


http://adlib.everysite.co.uk/resources/000/262/258/PB13315_manage_erosion_1.jpg

Converting Forecasts to Warnings

Extreme Forecast Index (EFI):

- Integration of difference between model forecast and model climatology (20 years reforecasts)
  - EFI>~0.5 = severe event
- Convert to warning areas based on:
  - Minimum EFI threshold

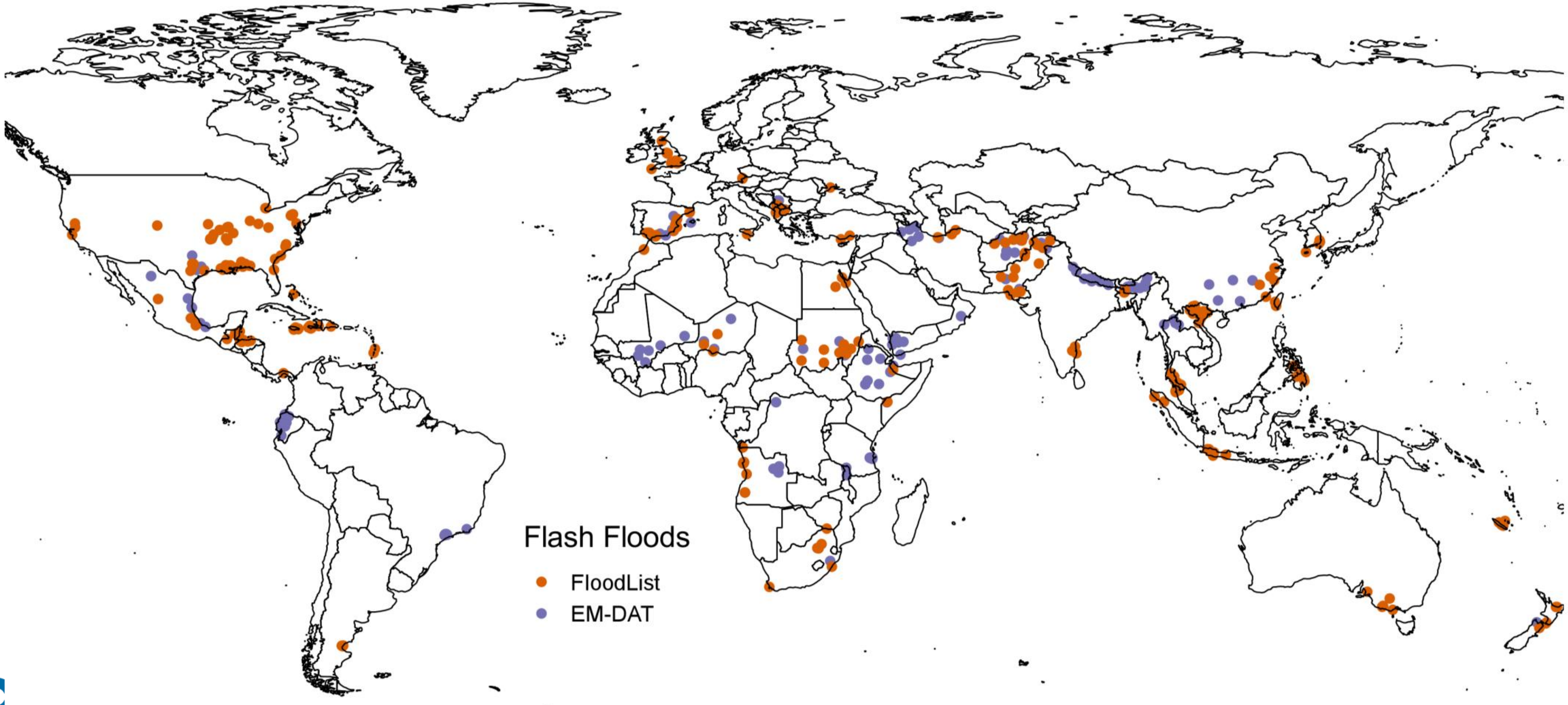


$\pi \int_0 \sqrt{p(1-p)}$

-1 ≤ EFI ≤ 1

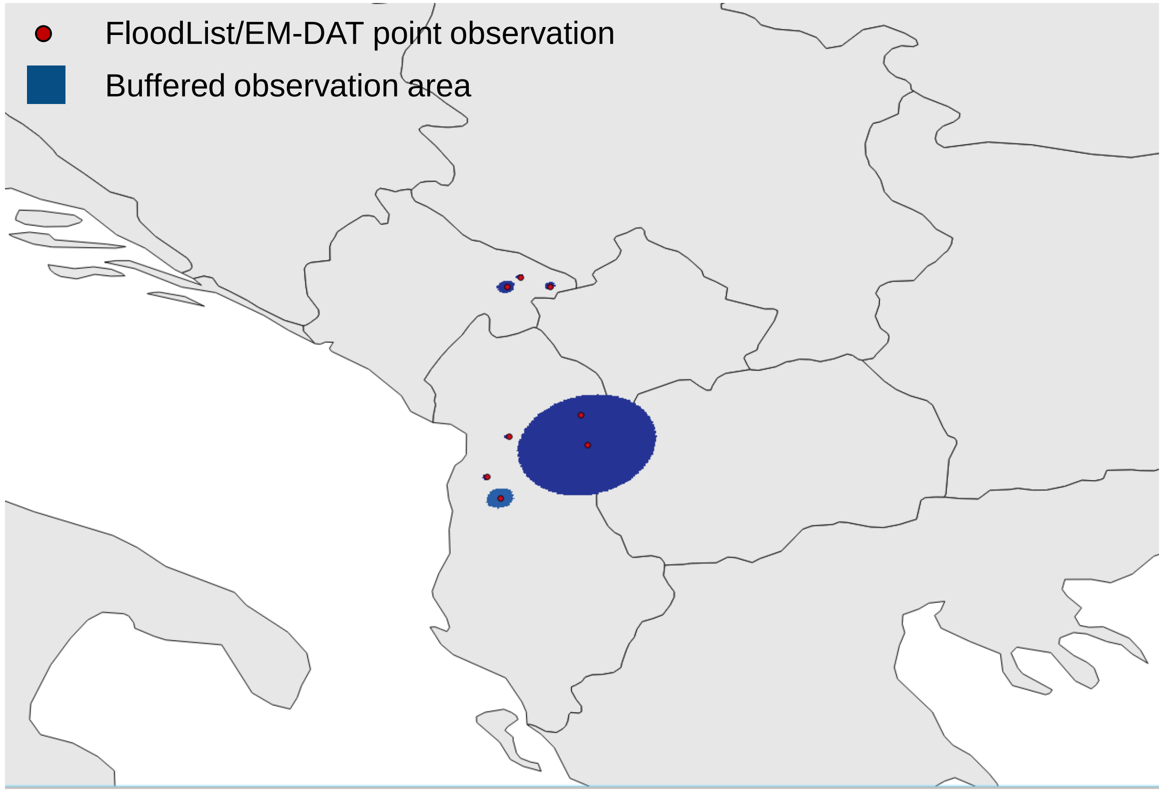
Verification Experiment: March 2016 – March 2017

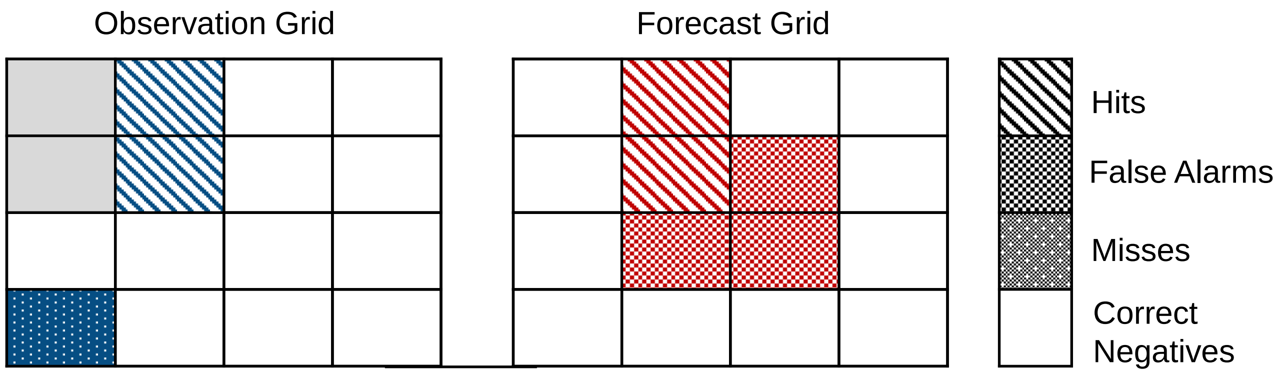
- Calculate total precipitation & surface runoff EFI daily at 00 UTC, 6 hourly out to 120 h lead time
- Create warning areas using EFI thresholds 0.0 – 1.0 (increments = 0.05)
- Compare against ‘flash flood’ observations from EM-DAT (161) & FloodList.com (238)



Verification: Methodology

- Convert point observations to 18km ENS grid:
  - Buffer by spatial uncertainty (1 to 100’s km)
- Create 1 observation grid for each verification date
- Compute the contingency table
- Summarise over the whole verification year

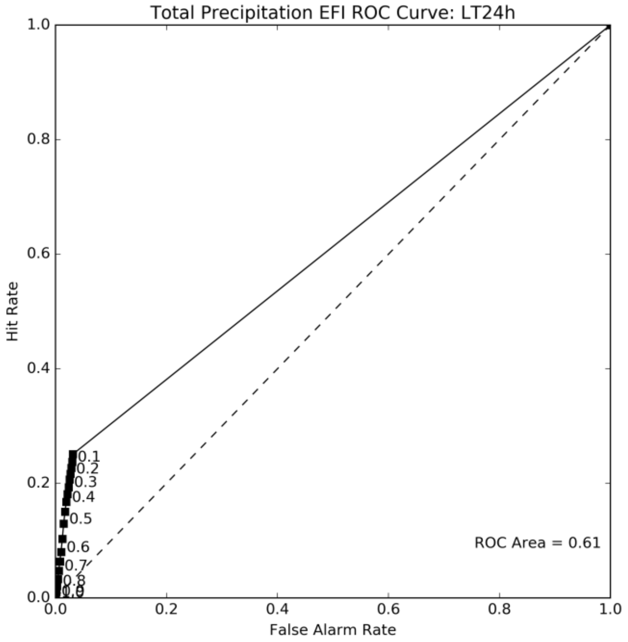


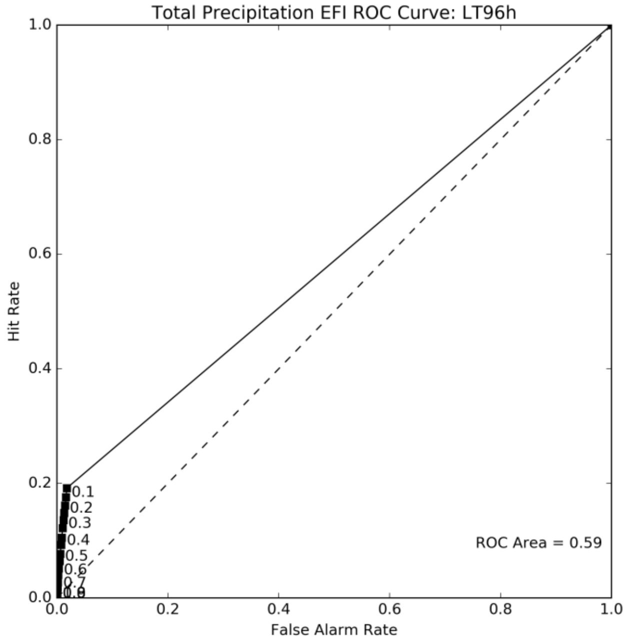


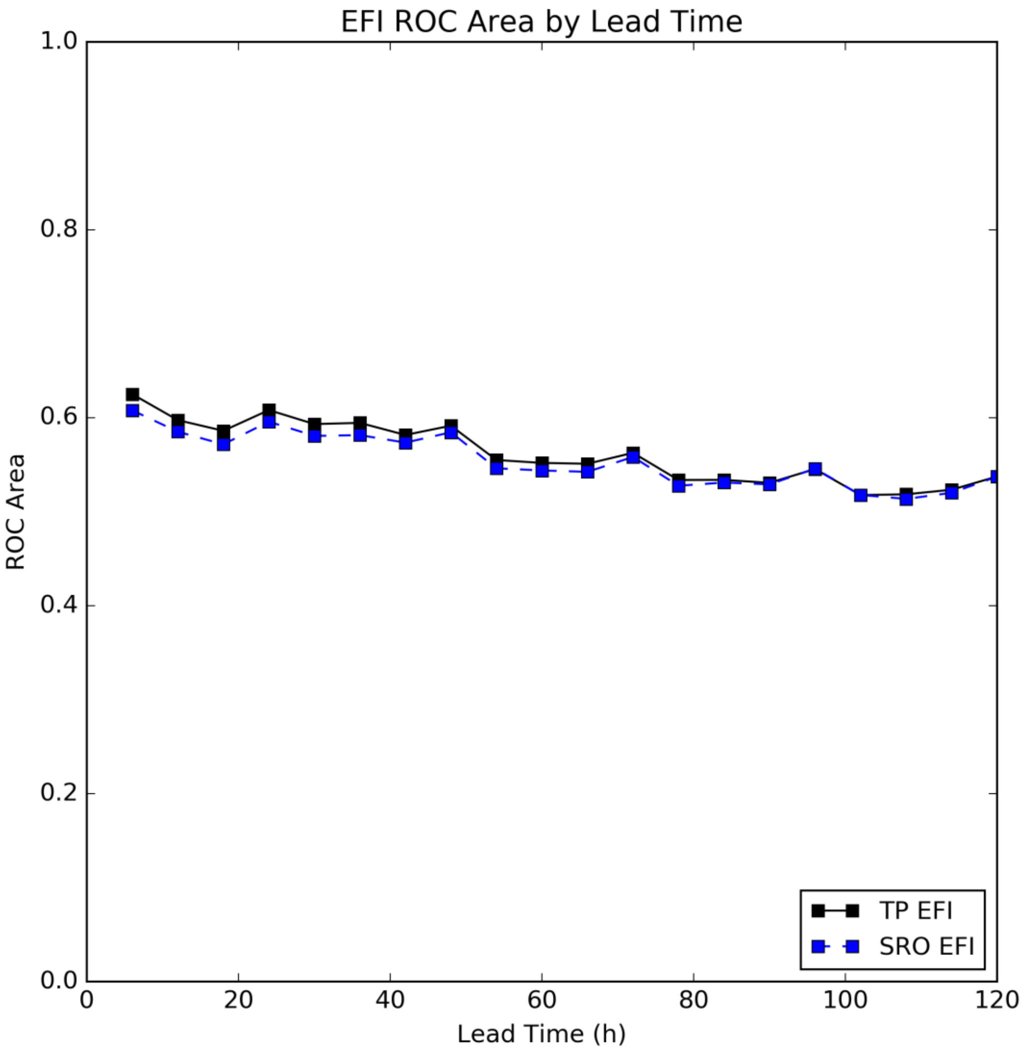
		Observed? Yes	Observed? No
Forecasted?	Yes	2	3
	No	1	8

Verification Results: ROC Area

- ROC Area ~0.6
  - Only shallow gradient with lead time – really??
- ROC curves actually clustered in bottom left corner
  - Really low hit and false alarm rates – the latter doesn't really vary

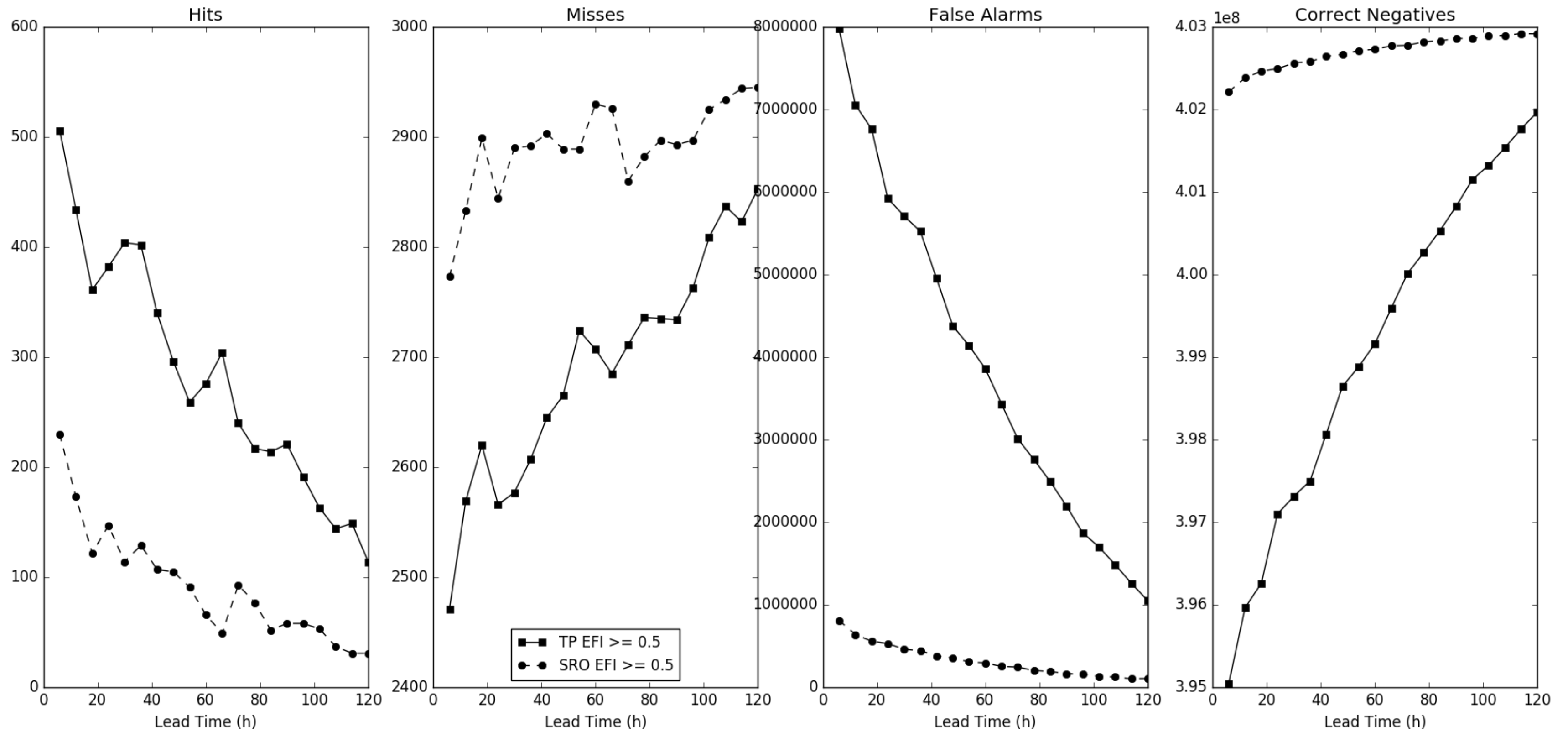






Verification Results: Raw Contingency Table Results

- Surface runoff is less active than total precipitation
- Misses ~x9 greater than hits = low hit rate
- Correct negatives ~x1000 greater than false alarms = v. low false alarm rate
- Skews the subsequent analysis



Verification Results: Problems with Global Analysis

- Only actually have 199 observations (w. spatial uncertainty <= 18km)

  = 995 observation pixels over whole verification period

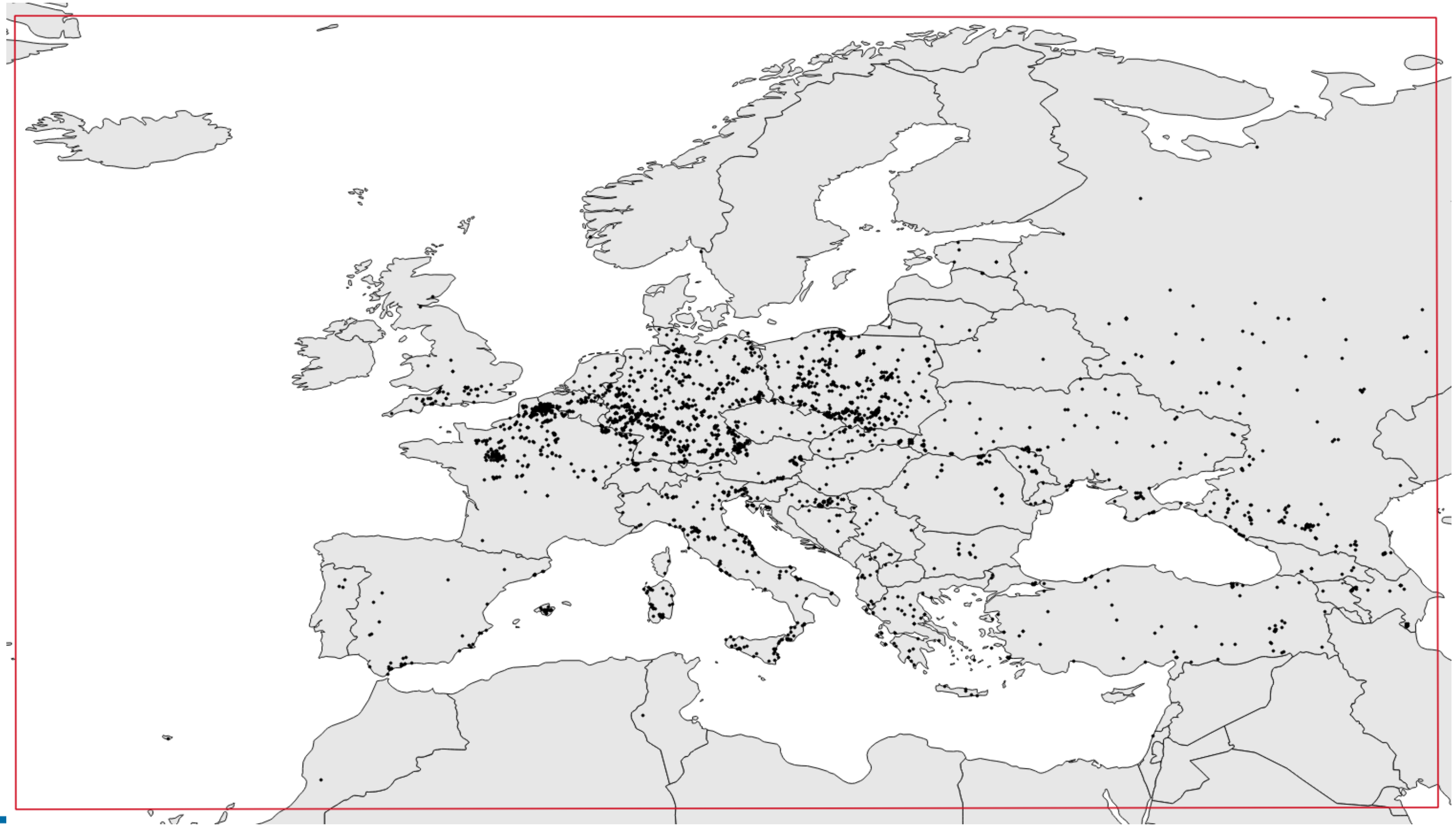
- 1,116,414 land pixels * 365 days
- Consequently the contingency table is unevenly distributed
  - Need to be careful not to repeat the Finley affair (1884)

$$Finley\ skill = \frac{Hits + Correct\ Negatives}{Total} = \frac{0 + \ (1{,}116{,}414 * 365)}{0 + 995 + 0 + (1{,}116{,}414 * 365)} = 99.99\%$$

- Cannot robustly verify the system at global scale
- Need to focus on smaller area with higher density observations >> Europe

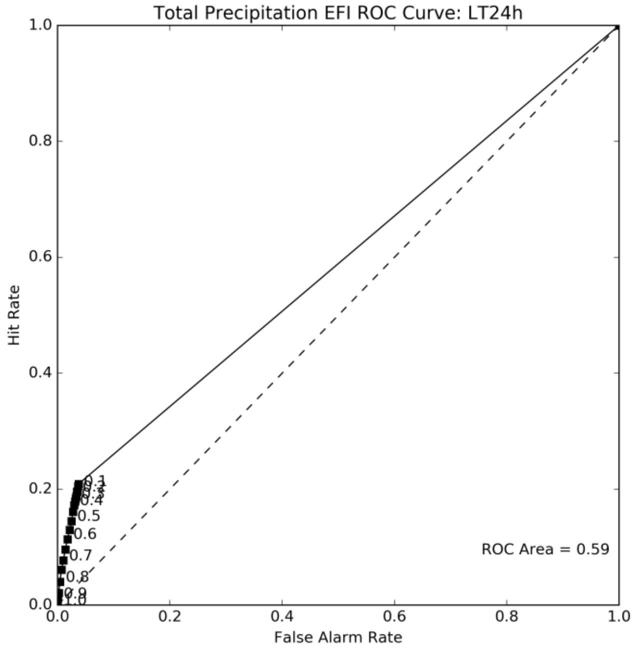
European Verification

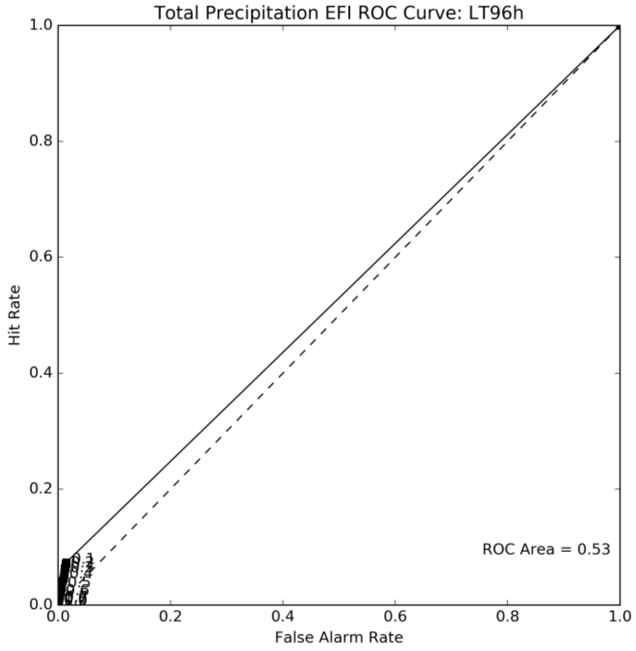
- European Severe Weather Database (ESWD) logs 2,544 heavy rainfall reports which mention flooding
- Repeat previous analysis: spatial uncertainty < 18km, timing uncertainty < 6 hours

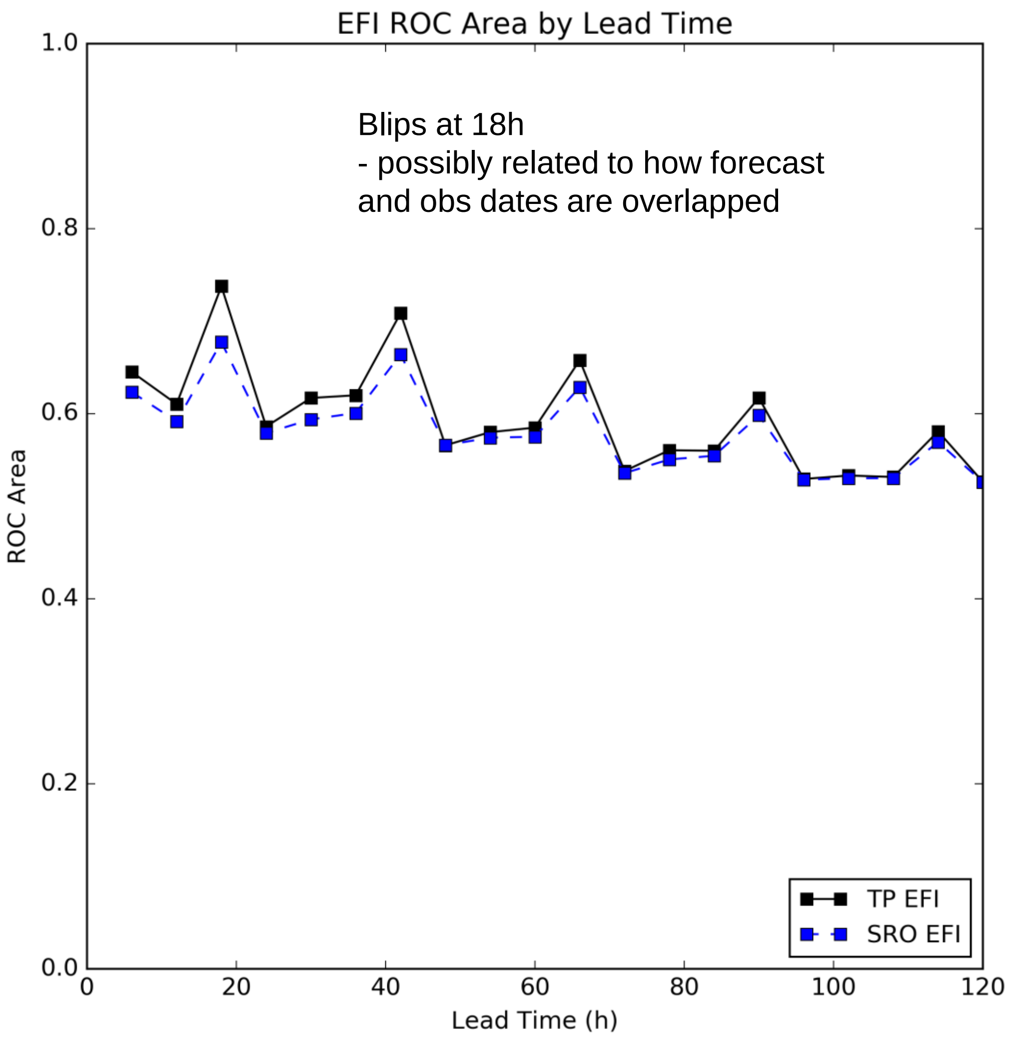


European Verification: Results

- ROC score still ~0.6
- ROC curves clustered in left hand corner
- Correct negatives and false alarms still dominating results

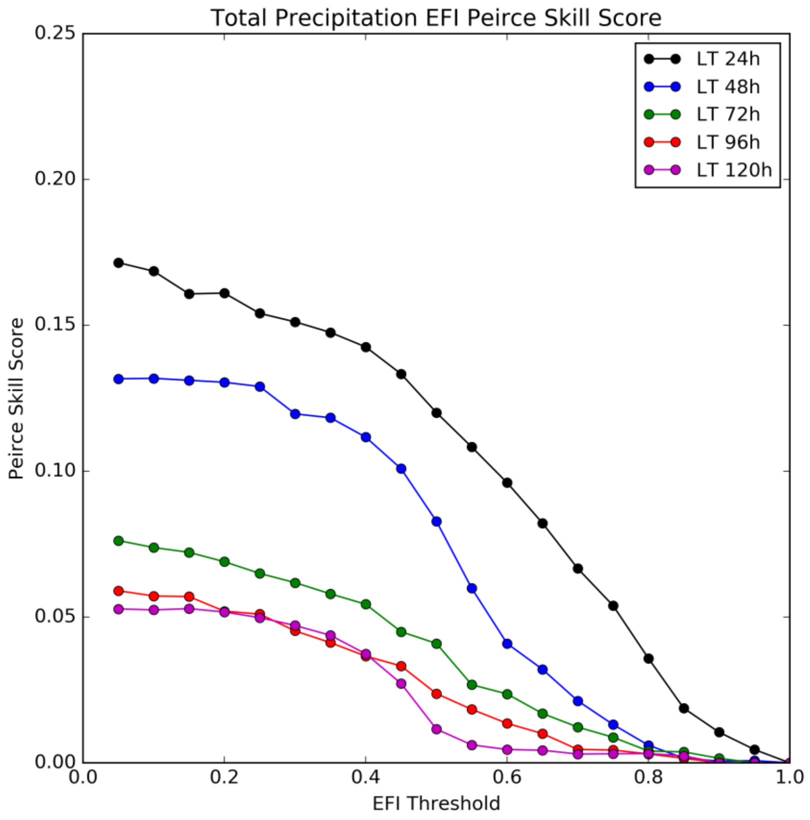


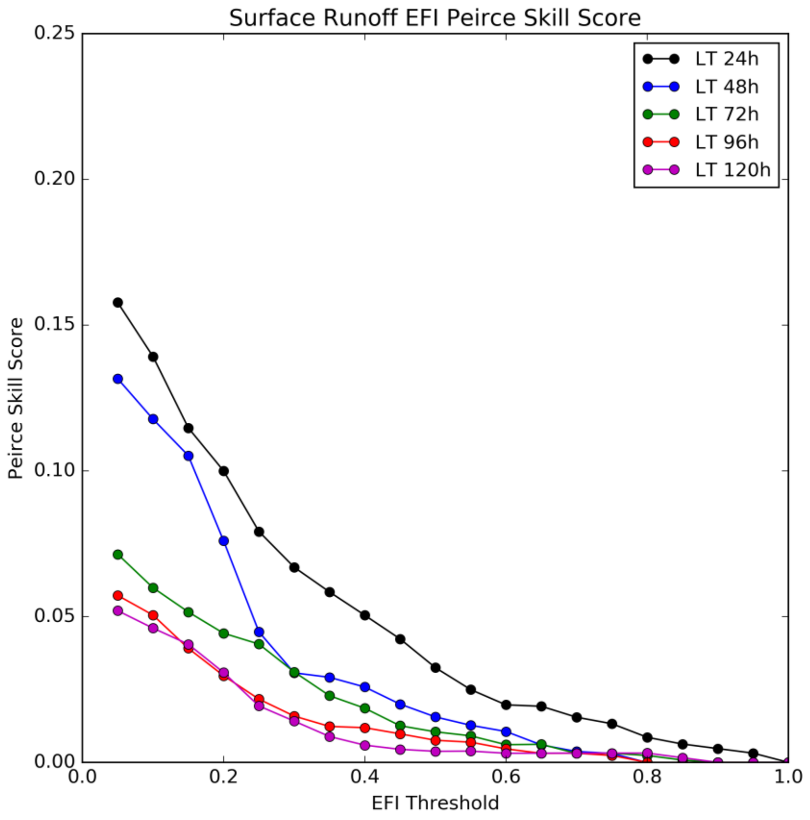




European Verification: Peirce Skill Score

- = Hit Rate – False Alarm Rate
  - How well can the forecast distinguish ‘yes’ events from ‘no’ events
- Hit rate marginally better than false alarm rate
  - Surface runoff EFI skill drops off rapidly with increased EFI threshold





Verification so far…

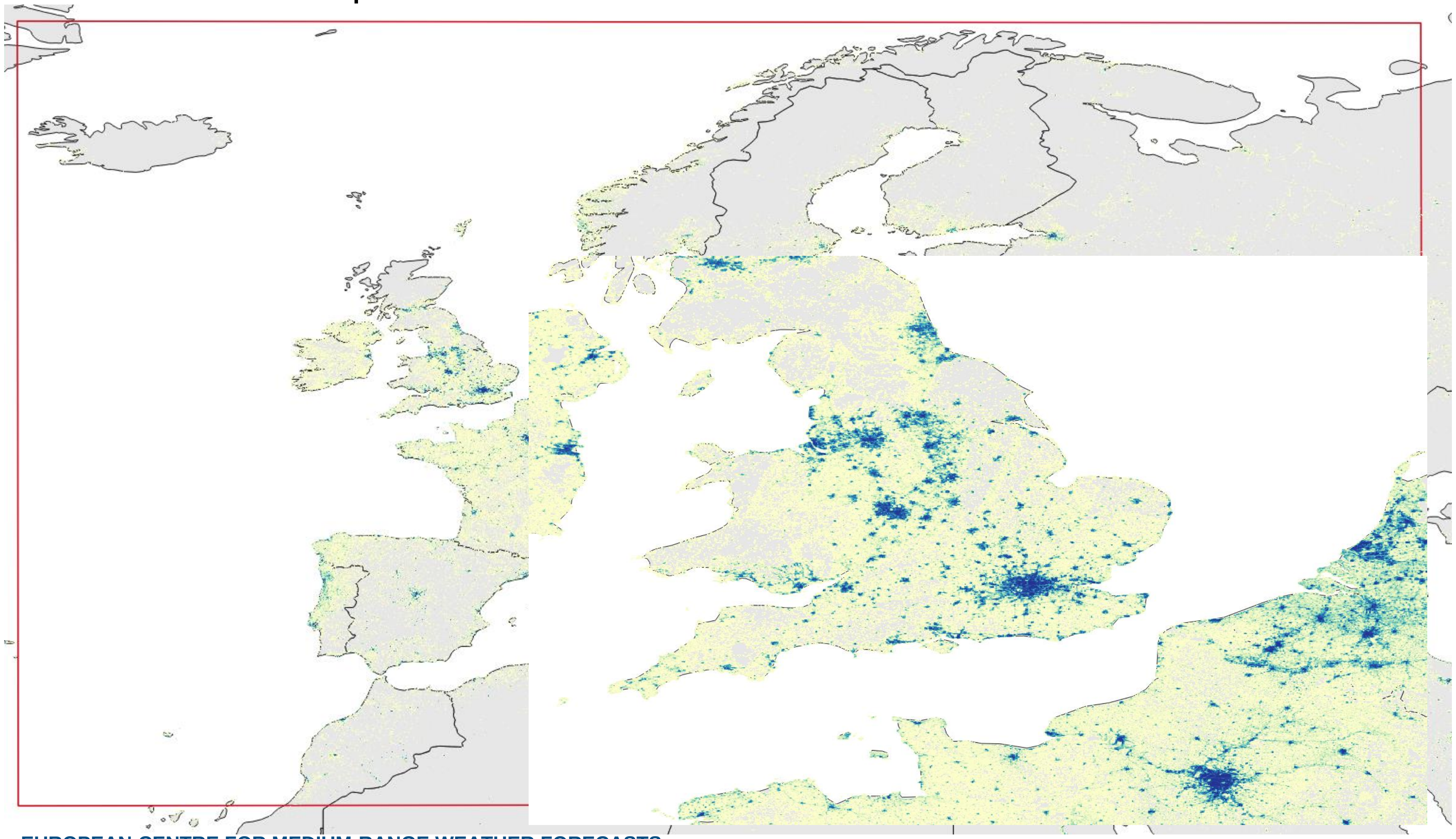
- On balance have a slight semblance of skill at shorter lead times
  - Total precipitation edges surface runoff but at cost of more false alarms

- Metrics are being hindered by uneven distribution of contingency table values
  - Correct negatives are being awarded where there could never be a (reported) event e.g. Sahara desert
  - Therefore refine verification only to areas where we can expect reports i.e. populated areas…

Inclusion of Population Exposure: GHSL

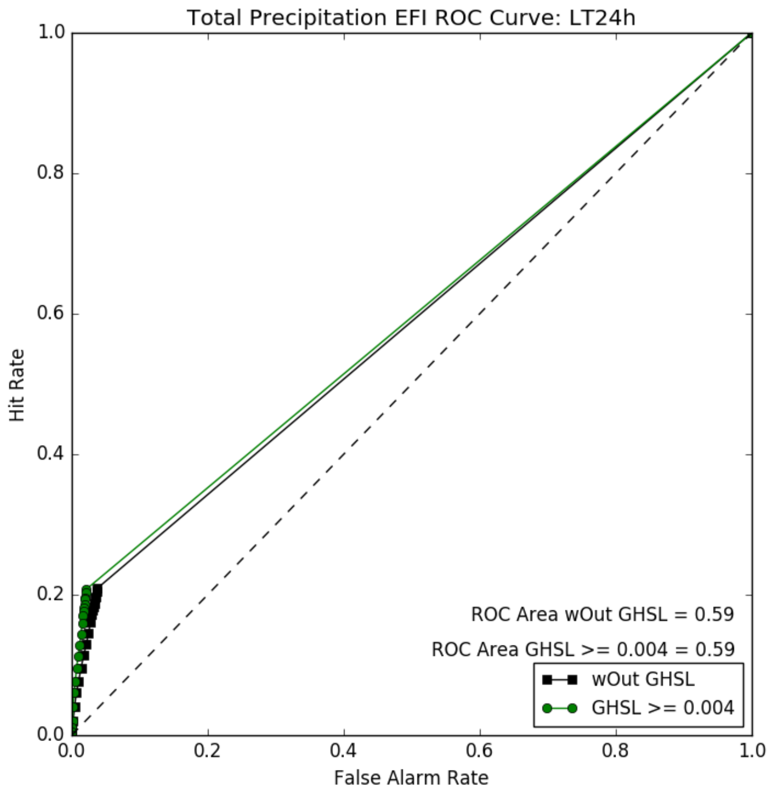
- Global Human Settlement Layer – developed at JRC
- % urbanised cover in each 1km pixel

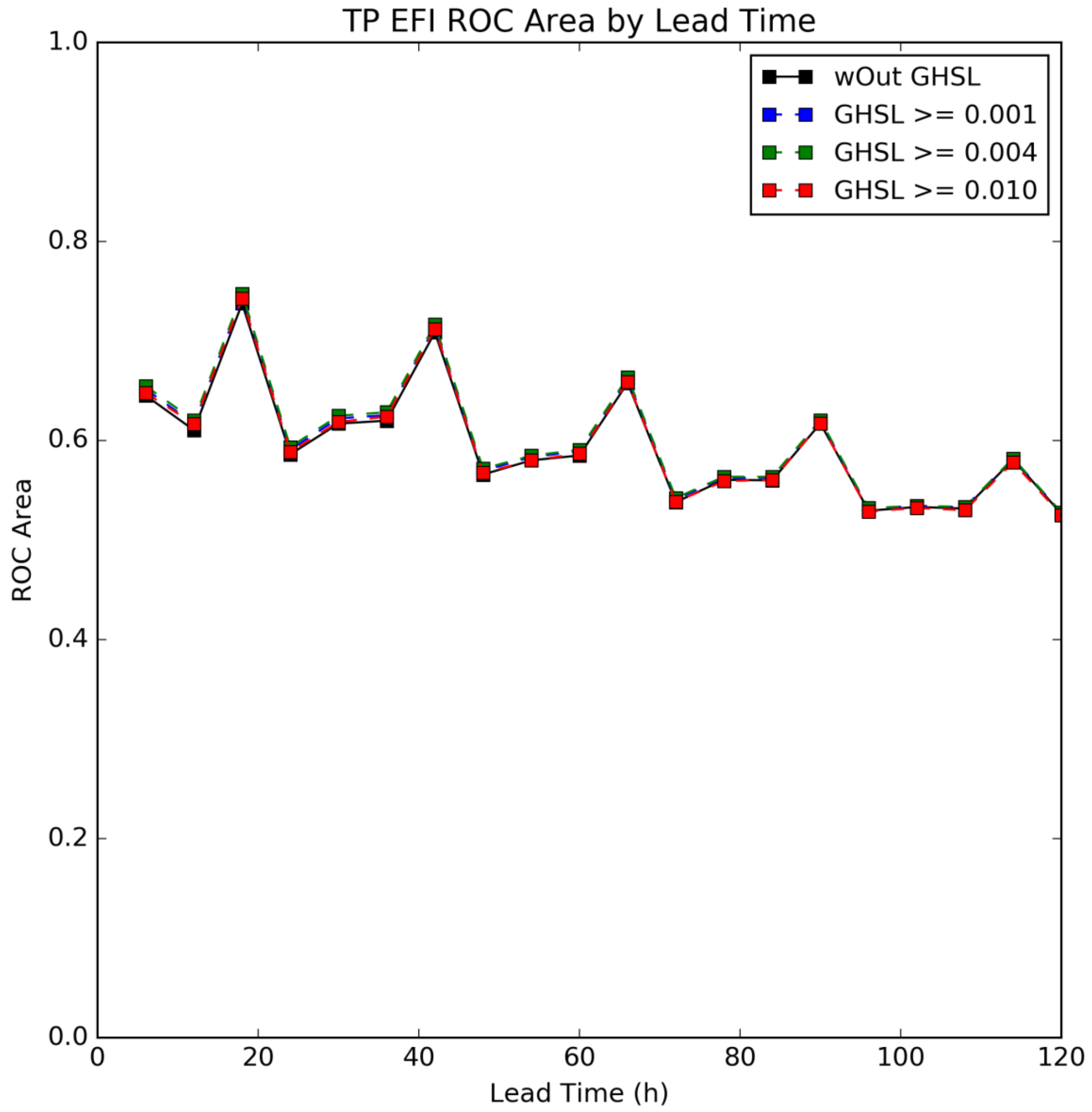
Apply threshold to remove unpopulated areas from analysis



Verification Results: with GHSL

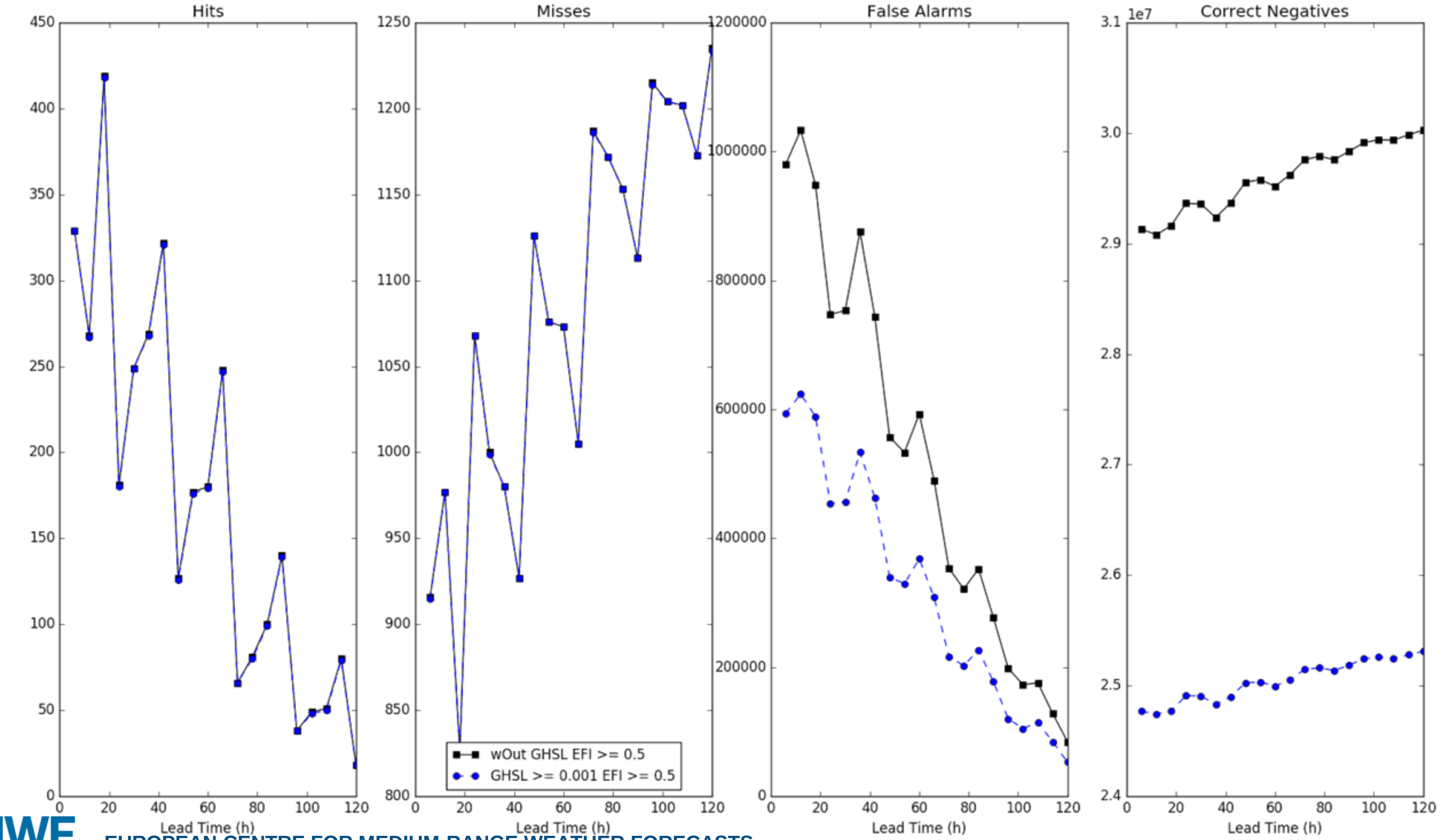
- ROC scores show little difference …





Verification Results: with GHSL

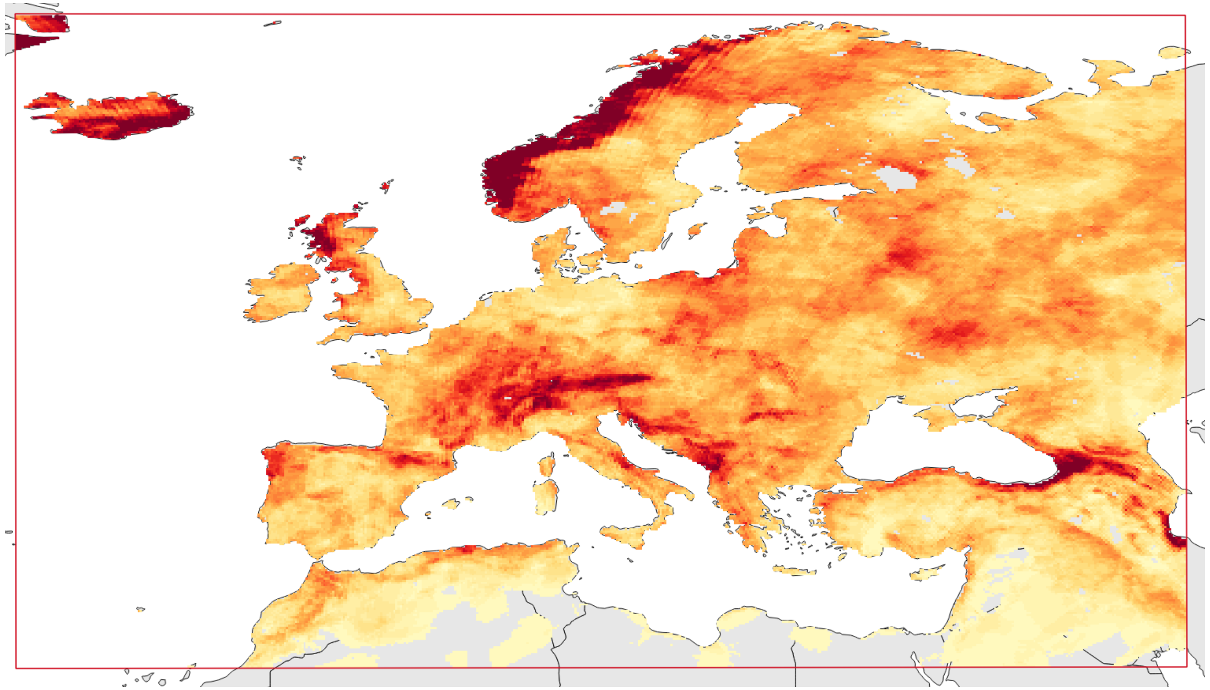
- …large reduction in number of false alarms and correct negatives
- No difference in hits and misses



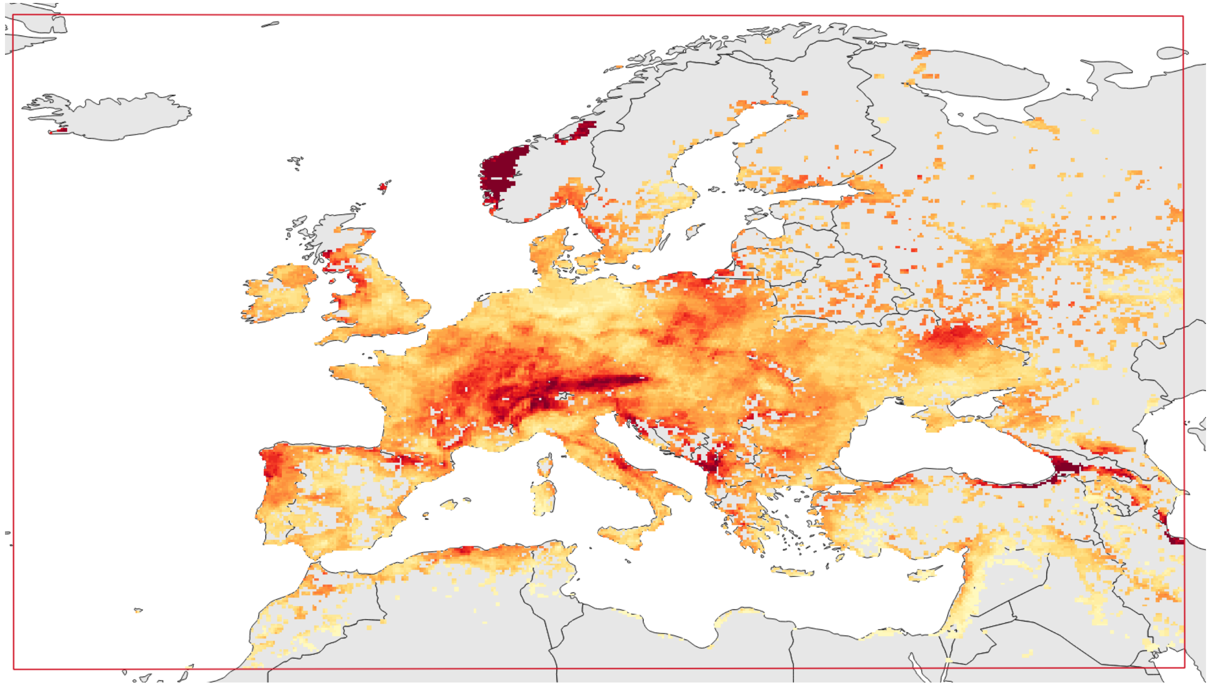
Verification Results: with GHSL

- Total Precipitation EFI: False Alarms
- Reduction in Scandinavia and eastern areas

Without GHSL



GHSL >= 0.004

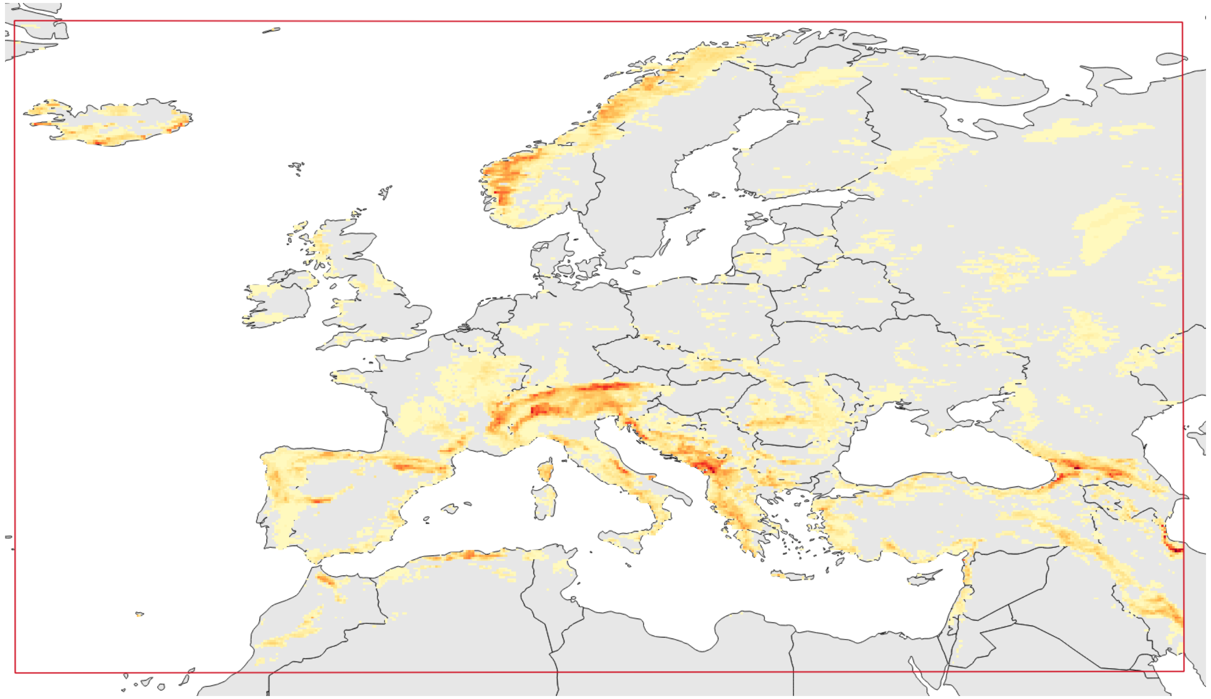


EFI >= 0.50
Lead Time 24h

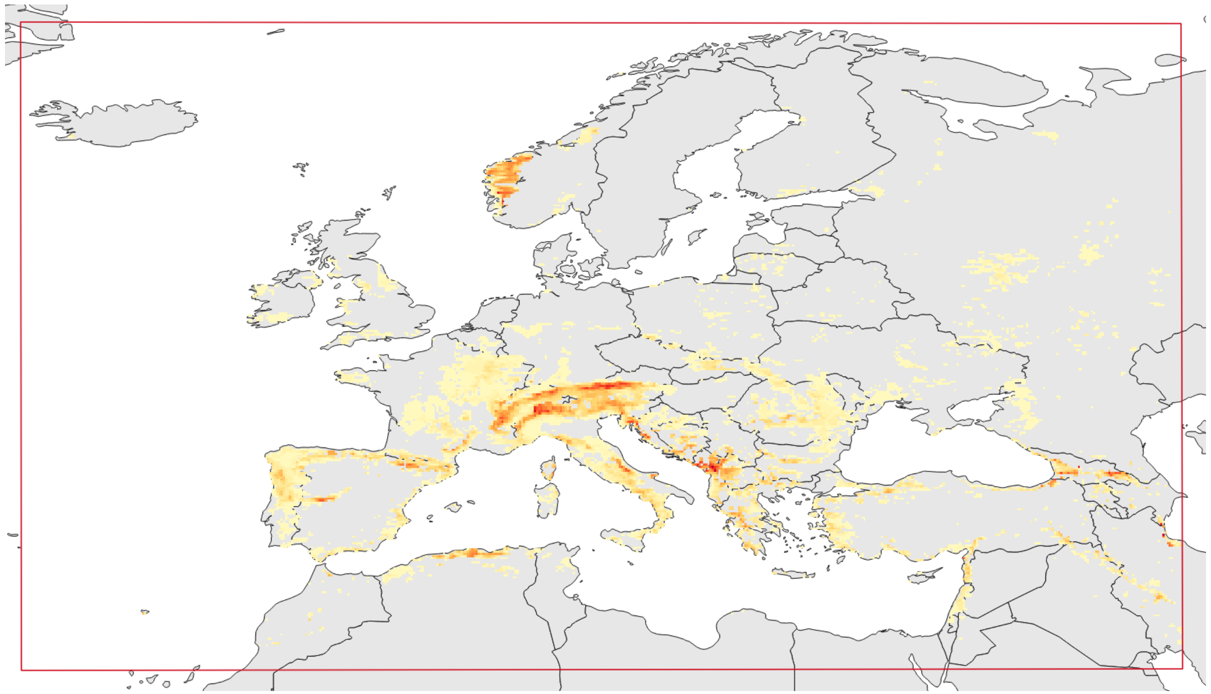
Verification Results: with GHSL

- Surface Runoff EFI: False Alarms

Without GHSL



GHSL >= 0.004



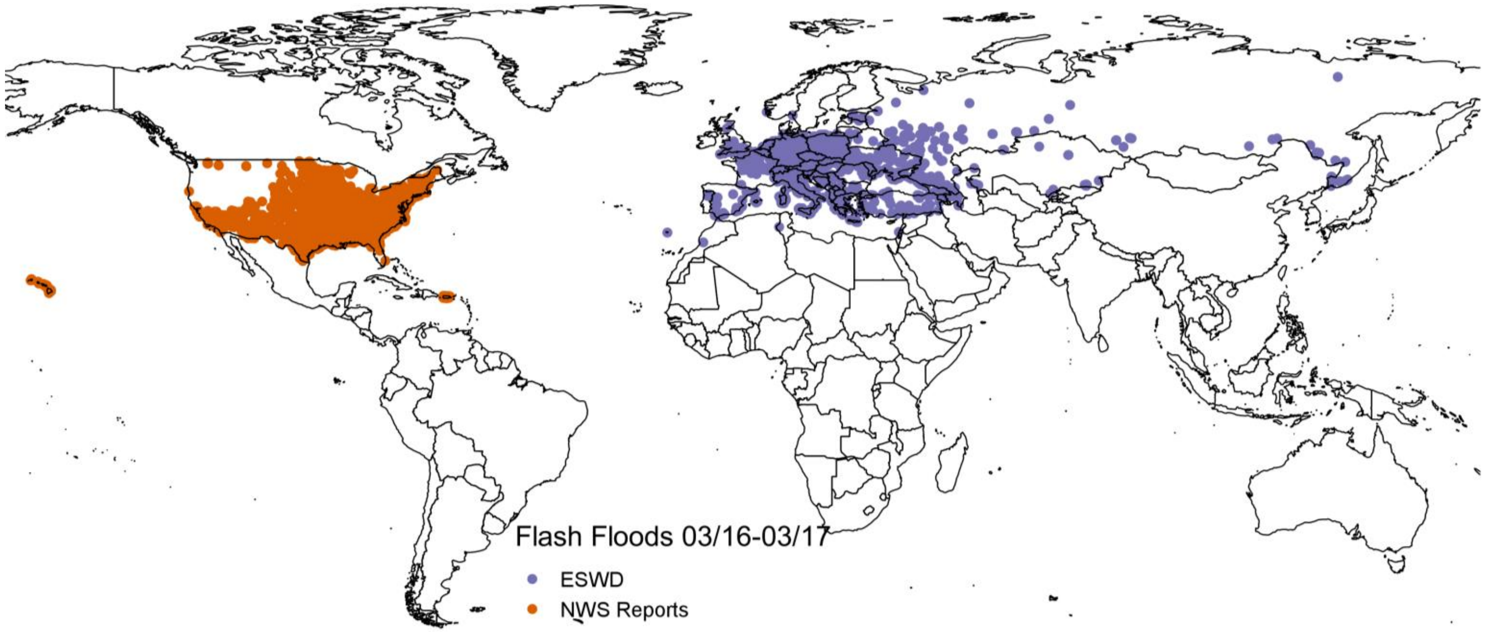
EFI >= 0.50
Lead Time 24h

Future Work

- Repeat analysis in USA using NWS storm reports

- River routing function – simple MC approach?

- Other exposure variables?
  - Transport networks
  - Infrastructure



Final Conclusions

- No global flash flood forecasting system – leaves areas with no capability

- Global ECMWF total precipitation/surface runoff forecasts give limited predictability of flash floods
  - Issue of scale: misses could be localised convection
  - Not enough observations to do robust verification

- Focussing verification upon Europe showed:
  - Large number of false alarms
  - Precipitation (more false alerts) and surface runoff (more misses) forecasts showed similar skill

- Thresholding by population exposure helps to reduce false alarms
  - Though little change in skill score

Thank You

Questions?

calum.baugh@ecmwf.int