# Entity-based Verification

## and Uncertainty Issues

**Pertti Nurmi & Christoph Zingerle**

Meteorological Research & Development
Finnish Meteorological Institute

# Outline

1) Introduction – Verification of categorical events
2) Issue of forecast resolution
   - ✓ “Double penalty” in verification
3) The entity-based verification system
4) Concluding remarks
   - ✓ Relation to hydrological forecasting…

# Categorical Events

## A piece of forecast verification history, 1884; Tornados in the USA, so-called "Finlay case"

Rare event - might be e.g. a flood event with a 50 year return period, as well !

| Tornado forecast | Tornado observed | | |
|---|---|---|---|
| | Yes | No | fc Σ |
| Yes | 30 | 70 | 100 |
| No | 20 | 2680 | 2700 |
| obs Σ | 50 | 2750 | 2800 |

- 30 hits
- 70 false alarms
- 20 missed events
- 2680 "correct rejections"

$$\frac{2680 + 30}{2800} = 96{,}8\ \%$$

| Tornado forecast | Tornado observed | | |
|---|---|---|---|
| | Yes | No | fc Σ |
| Yes | 0 | 0 | 0 |
| No | 50 | 2750 | 2800 |
| obs Σ | 50 | 2750 | 2800 |

**Hypothesis:**

**Never forecast a tornado:**

- **NO** hits
- **NO** false alarms
- 50 missed events
- 2750 "correct rejections"

$$\frac{2750 + 0}{2800} = 98{,}2\ \%$$

# Categorical Events

| Event forecast | Event observed | | |
|---|---|---|---|
| | Yes | No | Marginal total |
| Yes | Hit | False alarm | Fc Yes |
| No | Miss | Corr. rejection | Fc No |
| Marginal total | Obs Yes | Obs No | Sum total |

| Event forecast | Event observed | | |
|---|---|---|---|
| | Yes | No | Marginal total |
| Yes | a | b | a + b |
| No | c | d | c + d |
| Marginal total | a + c | b + d | a + b + c + d =n |

| Tornado forecast | Tornado observed | | |
|---|---|---|---|
| | Yes | No | fc Σ |
| Yes | 30 | 70 | 100 |
| No | 20 | 2680 | 2700 |
| obs Σ | 50 | 2750 | 2800 |

# Categorical Events

## *Measures* 1

| Event forecast | Event observed | | |
|---|---|---|---|
| | Yes | No | |
| Yes | a | b | a+b |
| No | c | d | c+d |
| | a+c | b+d | n |

| Tornado forecast | Tornado observed | | |
|---|---|---|---|
| | Yes | No | fc Σ |
| Yes | 30 | 70 | 100 |
| No | 20 | 2680 | 2700 |
| obs Σ | 50 | 2750 | 2800 |

**Base Rate** aka **Event Probability**

aka **Sample Climatology**

$r = ( a + c ) / n$

**Probability of a Forecast of Occurence**

$s = ( a + b ) / n$

**Bias** aka **Frequency Bias Index**

$B = FBI = ( a + b ) / ( a + c )$

**Proportion Correct**

$PC = ( a + d ) / n$

**Probability Of Detection, Hit Rate ( H )** aka **Prefigurance**

$POD = a / ( a + c )$

**False Alarm Ratio**

$FAR = b / ( a + b )$

**False Alarm Rate** aka **Probability of False Detection ( POFD )**

$F = b / ( b + d )$

**B = 2.00 / 1.00**
**PC = 0.97 / 1.00**
**POD = 0.60 / 1.00**
**FAR = 0.70 / 0.00**
**F = 0.03 / 0.00**

# Categorical Events

## *Measures* 2

| Event forecast | Event observed | | |
|---|---|---|---|
| | Yes | No | |
| Yes | a | b | a+b |
| No | c | d | c+d |
| | a+c | b+d | n |

| Tornado forecast | Tornado observed | | |
|---|---|---|---|
| | Yes | No | fc Σ |
| Yes | 30 | 70 | 100 |
| No | 20 | 2680 | 2700 |
| obs Σ | 50 | 2750 | 2800 |

**Hanssen & Kuiper's Skill Score**

aka **True Skill Statistics**

$$KSS = TSS = POD - F = (ad - bc) / [(a+c)(b+d)]$$

**Threat Score**

aka **Critical Success Index**

$$TS = CSI = a / (a + b + c)$$

**Equitable Threat Score**

$$ETS = (a - a_r) / (a + b + c - a_r)$$

where $a_r = (a + b)(a + c) / n$

**Heidke Skill Score**

$$HSS = 2(ad - bc) / [(a + c)(c + d) + (a + b)(b + d)]$$

**Odds ratio**

$$OR = a\,d / b\,c$$

$$ORSS = (ad - bc) / (ad + bc) = (OR - 1) / (OR + 1)$$

**KSS = 0.57 / 1.00**

**TS = 0.25 / 1.00**

**ETS = 0.24 / 1.00**

**HSS = 0.39 / 1.00**

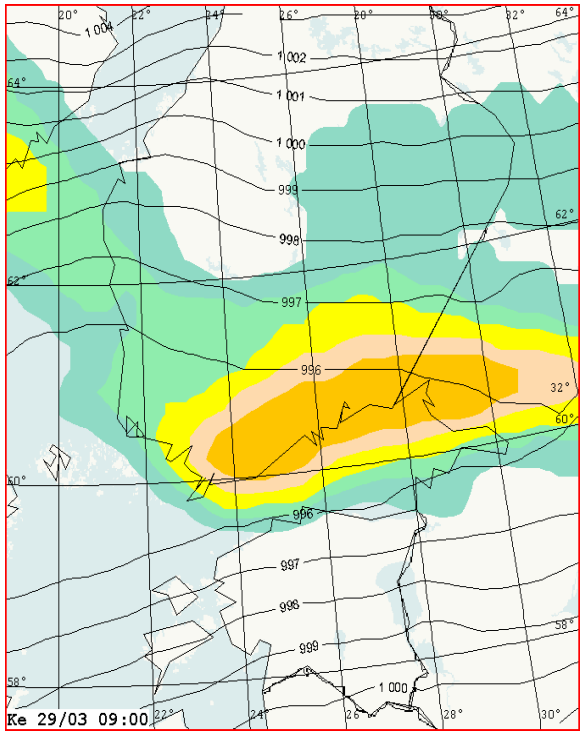
**OR = 57.43 / oo**

**ORSS = 0.97 / 1.00**

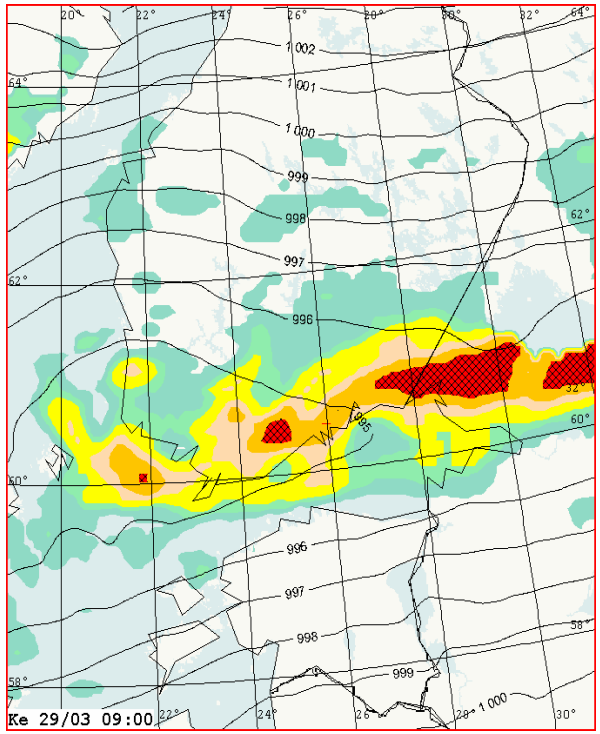
# "Eyeball" Verification

## Which one would you choose?

ECMWF global 25 km




HIRLAM MBE 9 km




**Observed**
**"TRUTH"**

?

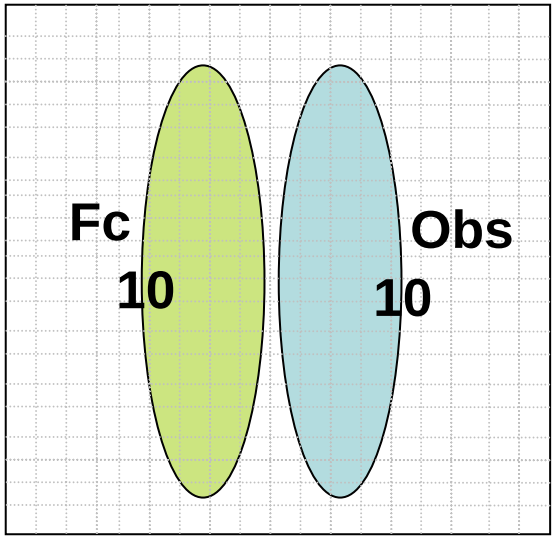
Uncertainty in verification

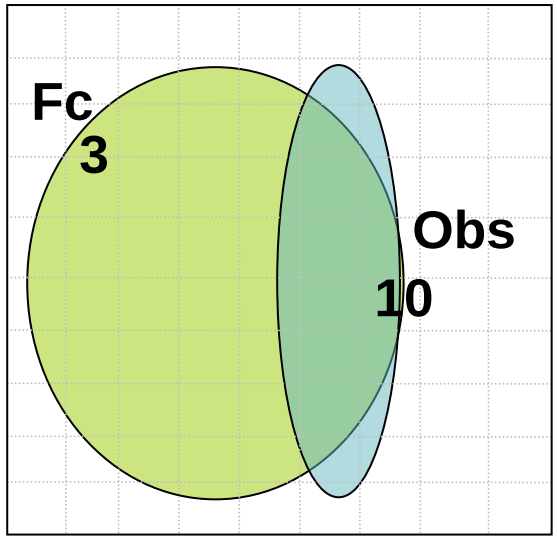
# Verification requirements

**…for *optimal* spatial forecasts like precipitation field**

- ✓ **Occurence vs. non-occurence predicted**
- ✓ **Location predicted precisely**
- ✓ **Amplitude predicted correctly**
- ✓ **Size predicted exactly**
- ✓ **Shape predicted accurately**
- ✓ **Spatial variability and structure predicted properly**

# "Double penalty"




**Hi-resolution forecast**

RMS ~ 5

POD= 0, FAR= 1, TS= 0

| FORECAST | OBSERVATION | | |
|---|---|---|---|
| | YES | NO | TOTAL |
| YES | 0 | 38 | 38 |
| NO | 38 | 285 | 323 |
| TOTAL | 38 | 323 | 361 |

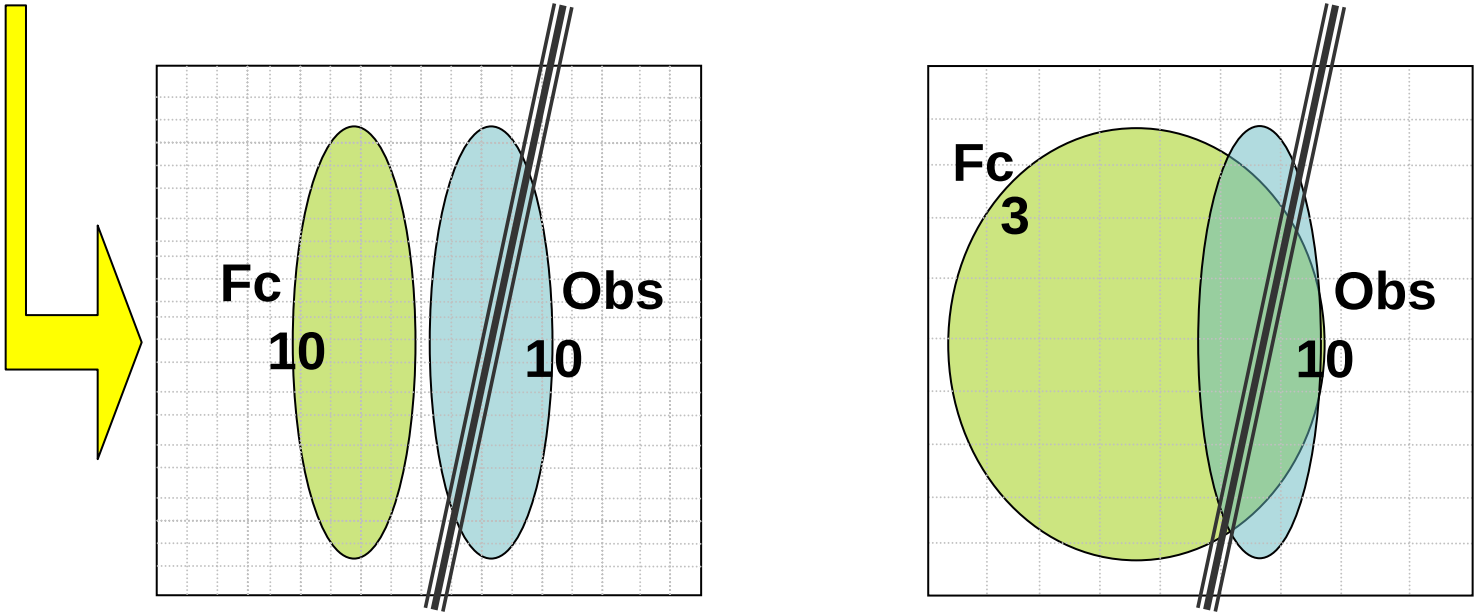
**Lo-resolution forecast**

RMS ~ 2.5

POD ~ 0.8, FAR ~ 0.7, TS ~ 0.3

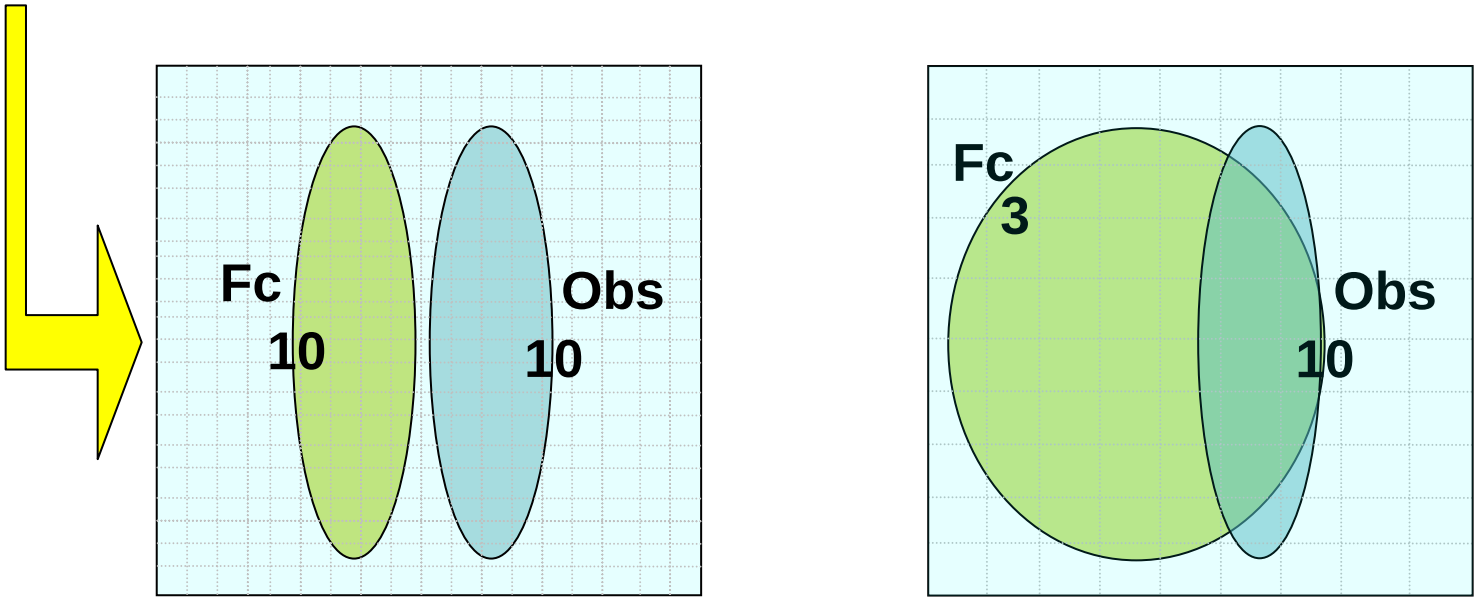
| FORECAST | OBSERVATION | | |
|---|---|---|---|
| | YES | NO | TOTAL |
| YES | 8 | 20 | 28 |
| NO | 2 | 91 | 93 |
| TOTAL | 10 | 111 | 121 |

# "Double penalty"




How about in the case of e.g. a road stretch thru the area ?

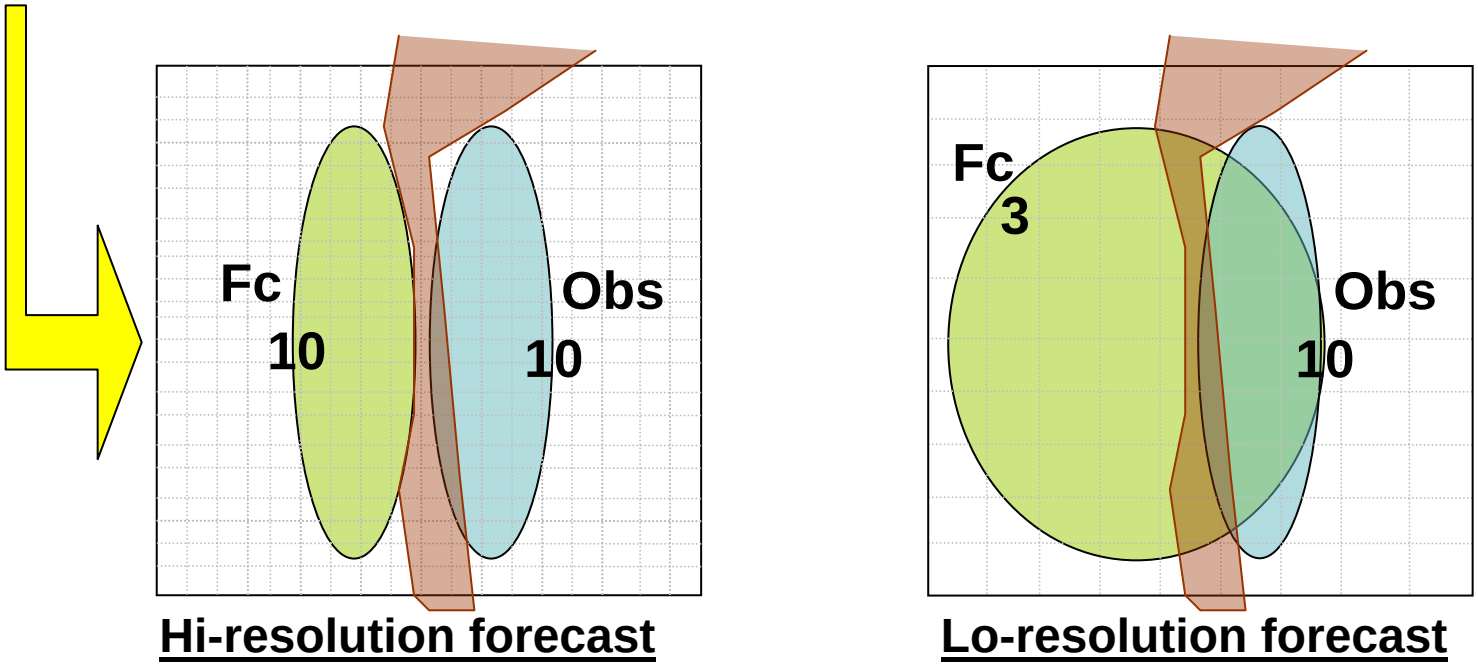
# "Double penalty"




**Hi-resolution forecast** **Lo-resolution forecast**

How about in the case if the area is a distinct catcment area ?

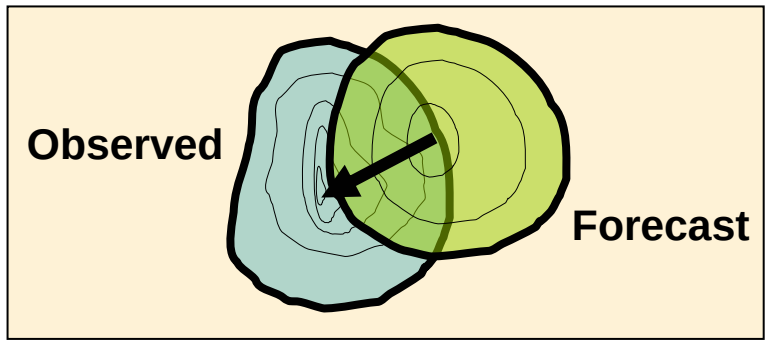
# "Double penalty"




How about in the case if e.g. a mountain ridge separates the area into two distinct catchment areas ?

# “CRA“

- Entity-based *error decomposition* method
  - ⇔ *Ebert & McBride, 2000*
- Entity (~“blob”) is defined using a specific threshold
  - ⇨ “CRA” ~ Contiguous Rain Area (e.g. R >100 mm /12hr)
- Forecast “blob” is moved around horizontally in a sub-domain to best fit the observed “blob”, i.e. until a *pattern matching* criterion is met:
  - *Minimize the MSE* between forecast and observed “blobs”
    - ⇔ Maximum overlap
  - Alternatively, *maximize the correlation*



# Error decomposition

1. Displacement (~ position) error (in %)
   - ✓ Vector difference between the original and final locations of the forecast
2. Volume (~ intensity) error (in %)
3. Pattern (~ residual) error (in %)
   - ✓ Accounts for differences in the fine structure of the forecast and observed fields

$$MSE_{tot} = \frac{1}{N} \sum_{i=1}^{N} (f_i - o_i)^2$$

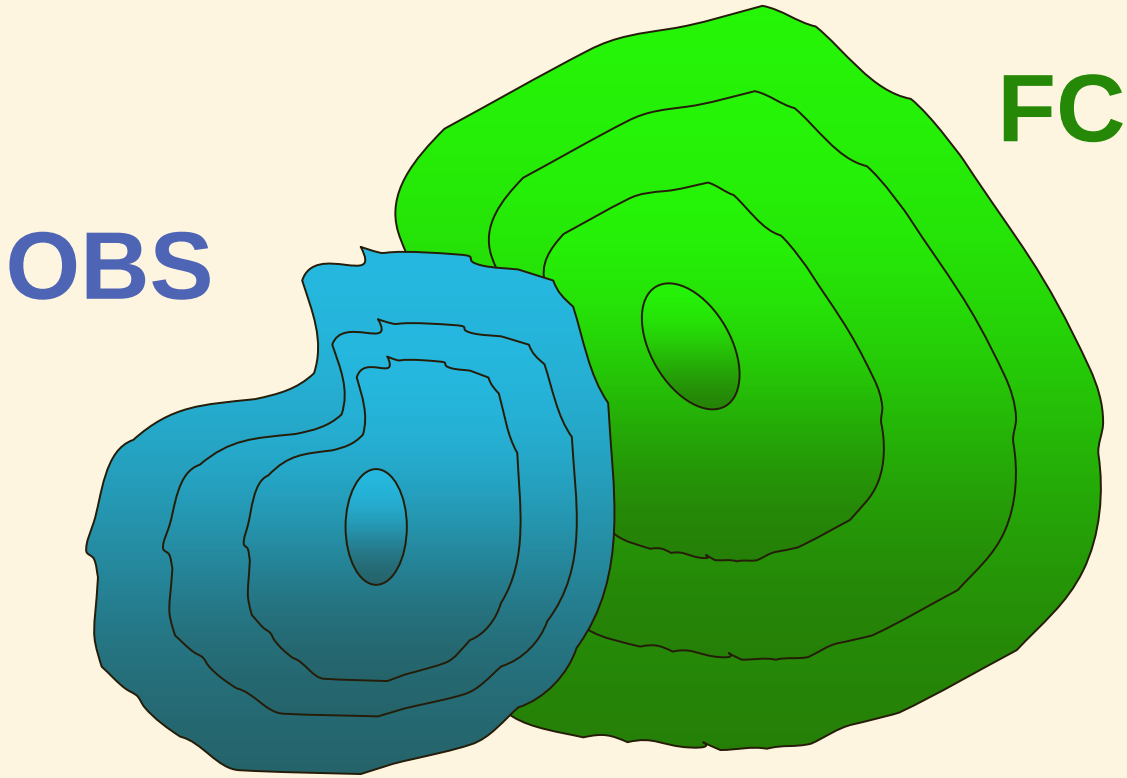
$$MSE_{tot} = MSE_{displ} + MSE_{vol} + MSE_{pat}$$

displacement~position  volume~intensity  pattern~residual

**NB:**
**Relative, not absolute values !**
( ⇔ sum = 100 %)

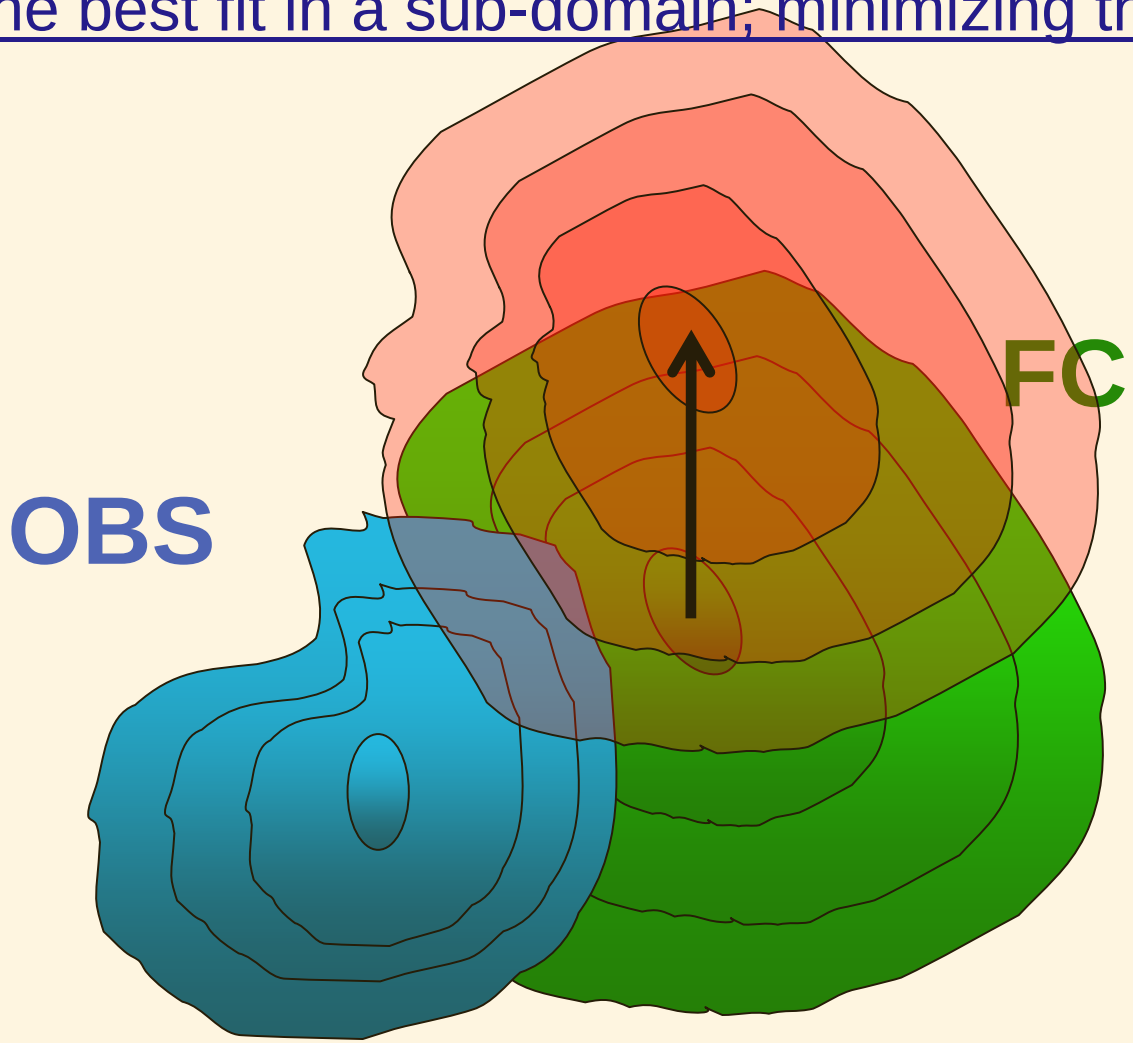
# Pattern Matching

## Searching for the best fit in a sub-domain; minimizing the MSE



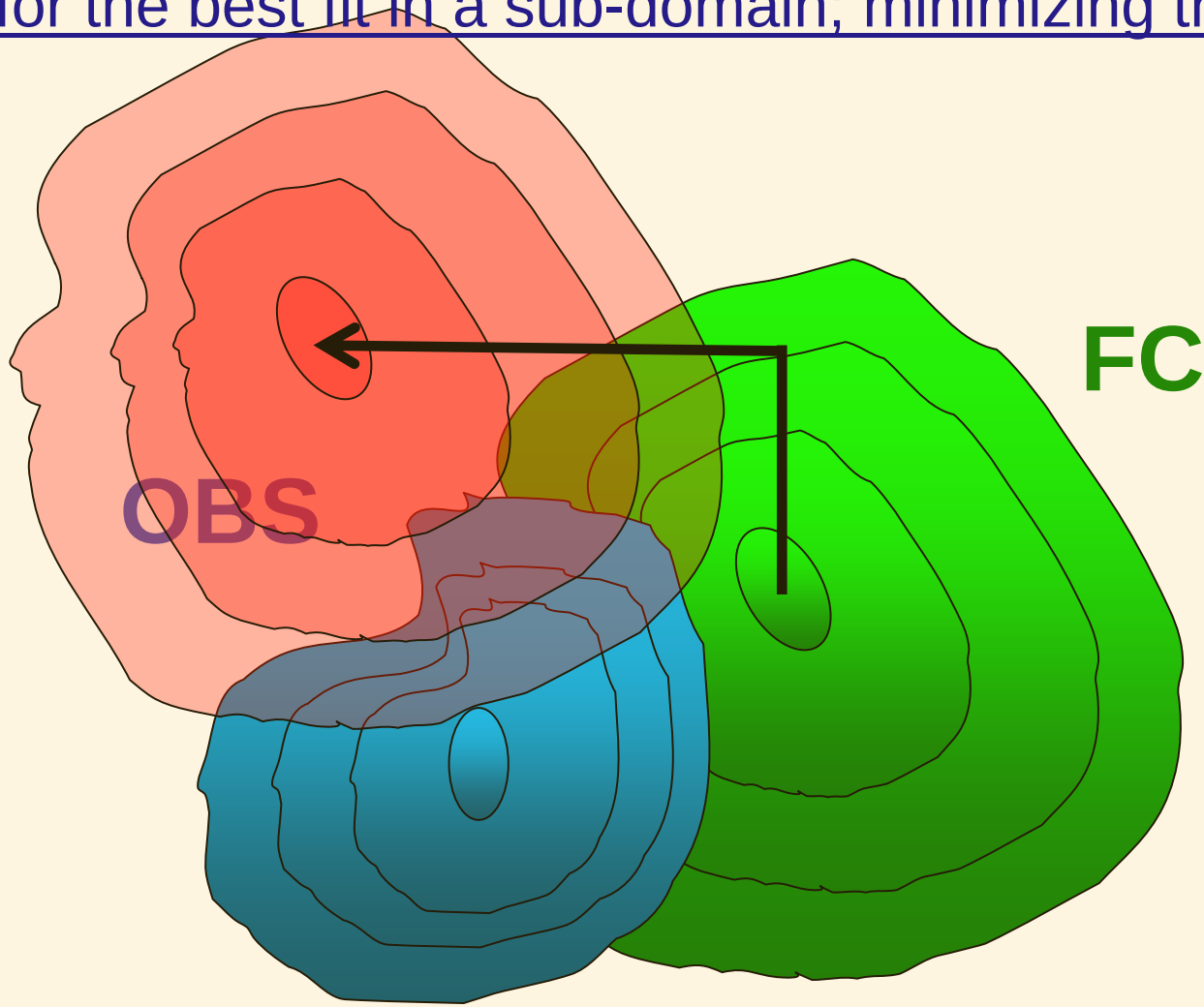
# Pattern Matching

## Searching for the best fit in a sub-domain; minimizing the MSE



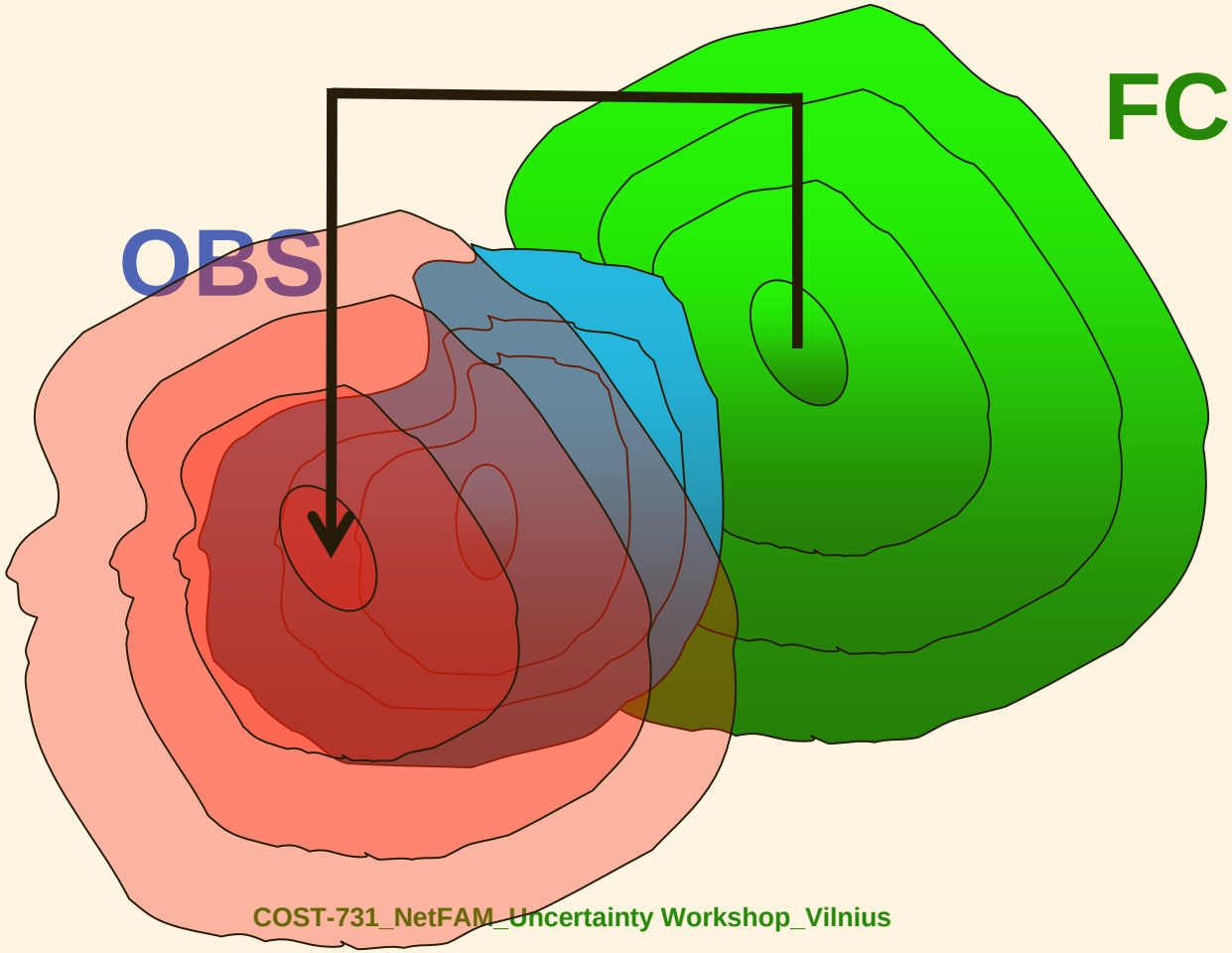
# Pattern Matching

## Searching for the best fit in a sub-domain; minimizing the MSE



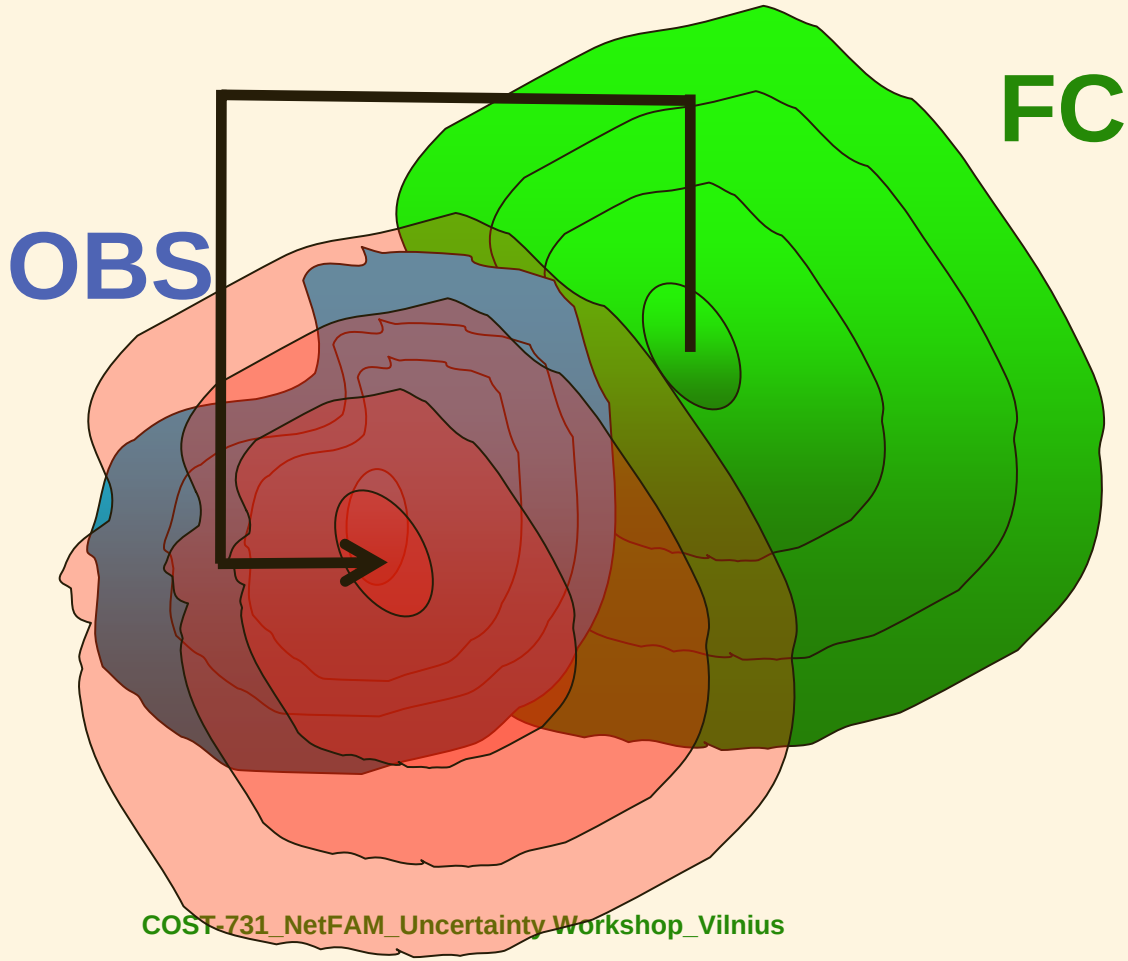
# Pattern Matching

## Searching for the best fit in a sub-domain; minimizing the MSE



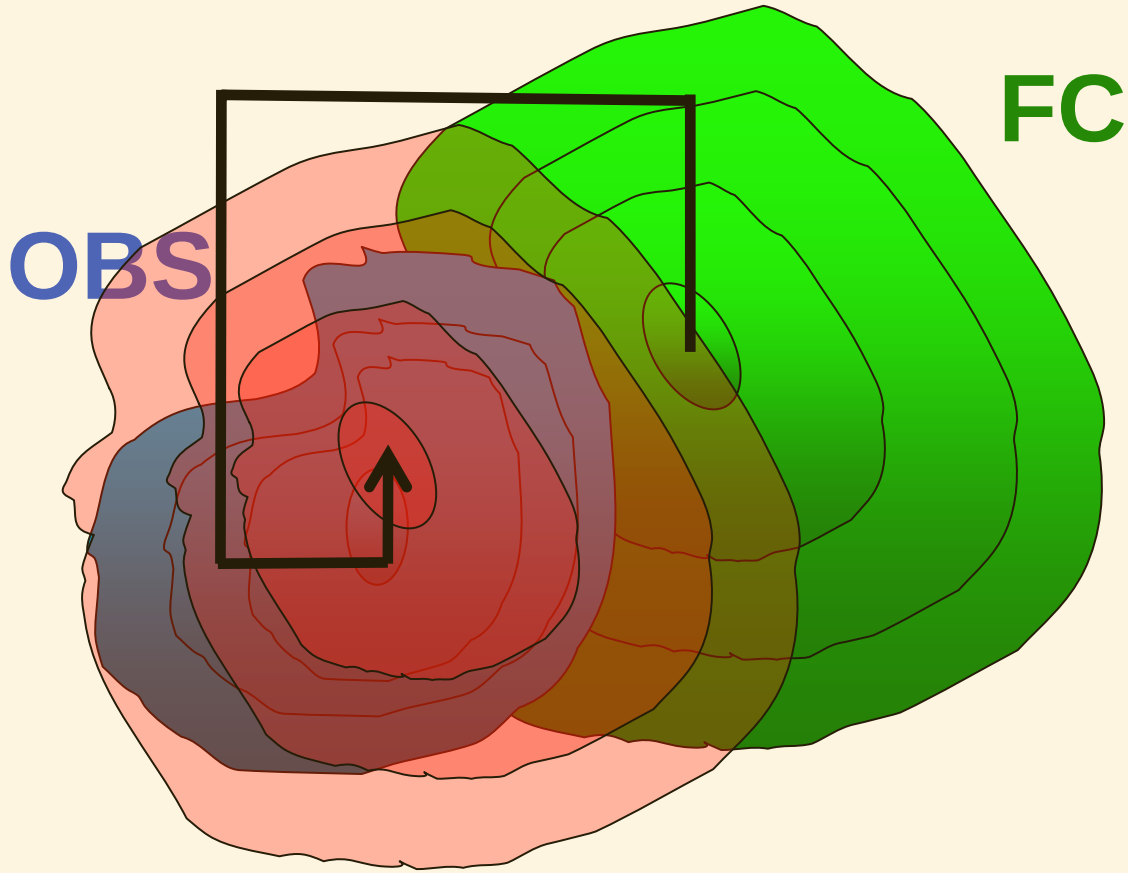
# Pattern Matching

## Searching for the best fit in a sub-domain; minimizing the MSE



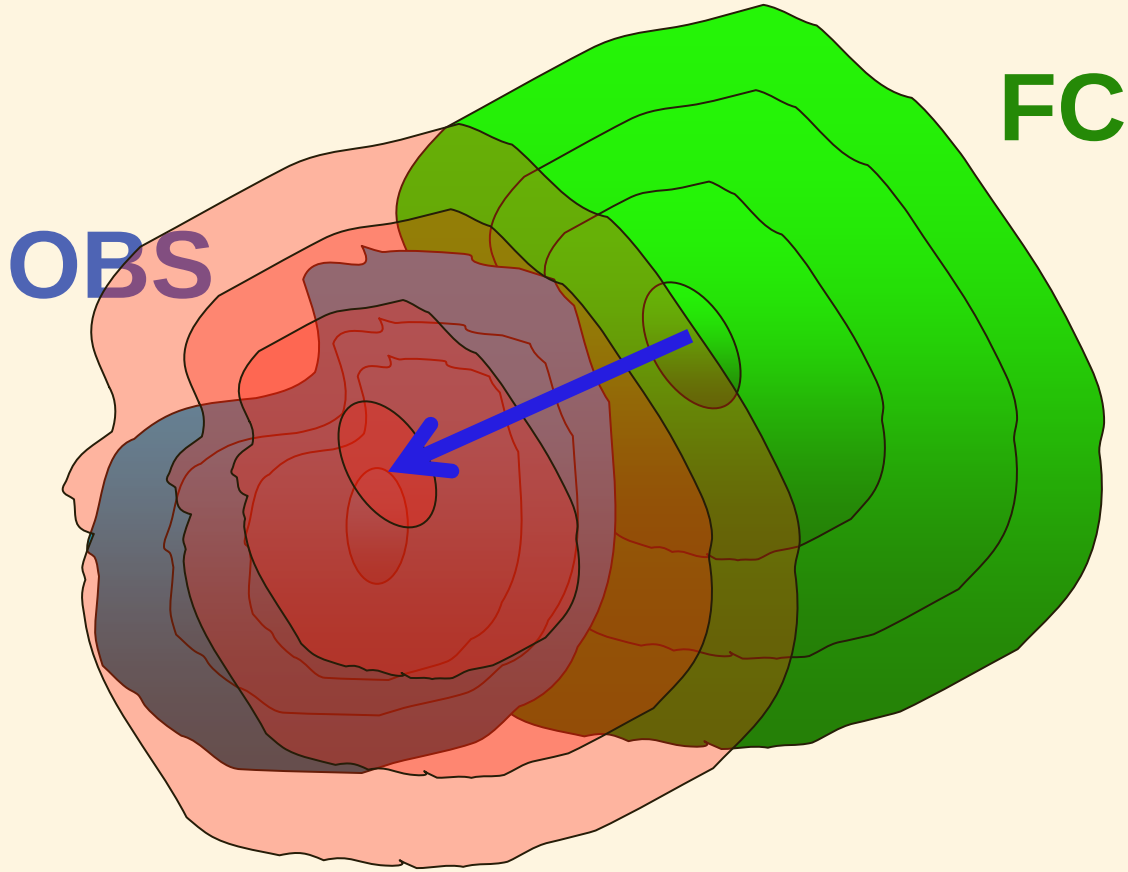
# Pattern Matching

## Searching for the best fit in a sub-domain; minimizing the MSE

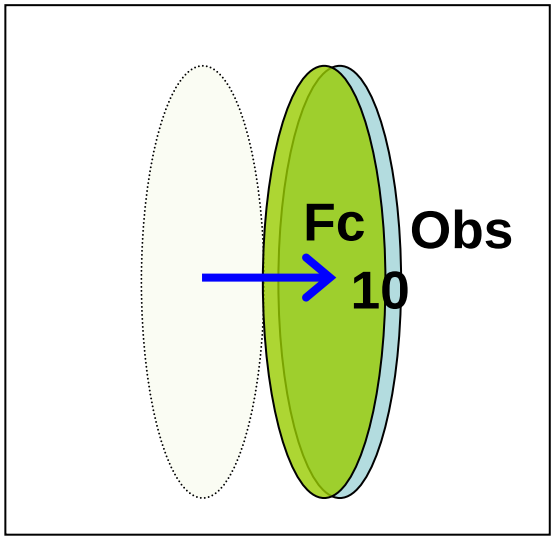


## Searching for the best fit in a sub-domain; minimizing the MSE

- **There you have it !**



# "~~Double penalty~~"




| FORECAST | OBSERVATION | | |
|---|---|---|---|
| | YES | NO | TOTAL |
| YES | 38 | 0 | 38 |
| NO | 0 | 323 | 323 |
| TOTAL | 38 | 323 | 361 |

**Original Hi-resolution forecast**

RMS ~ 5

POD= 0, FAR= 1, TS= 0

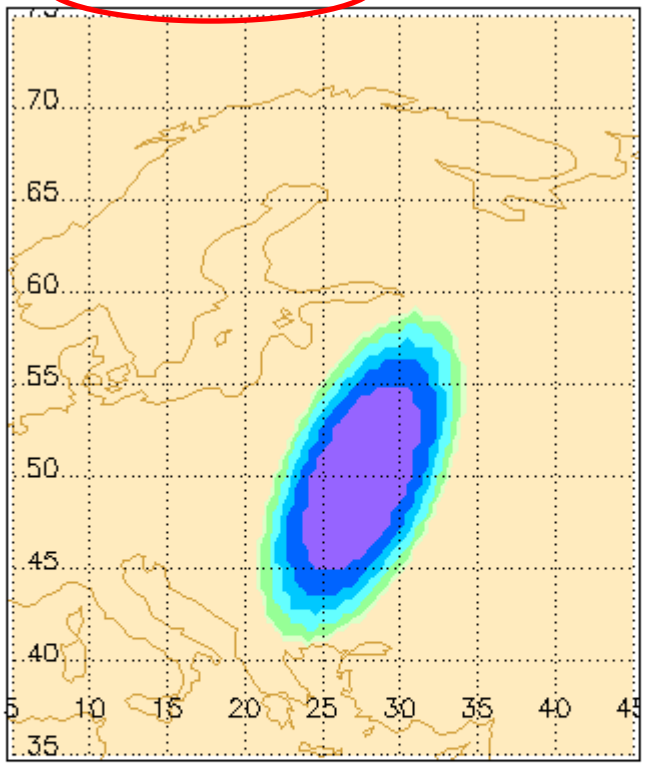
**Shifted Hi-resolution forecast**

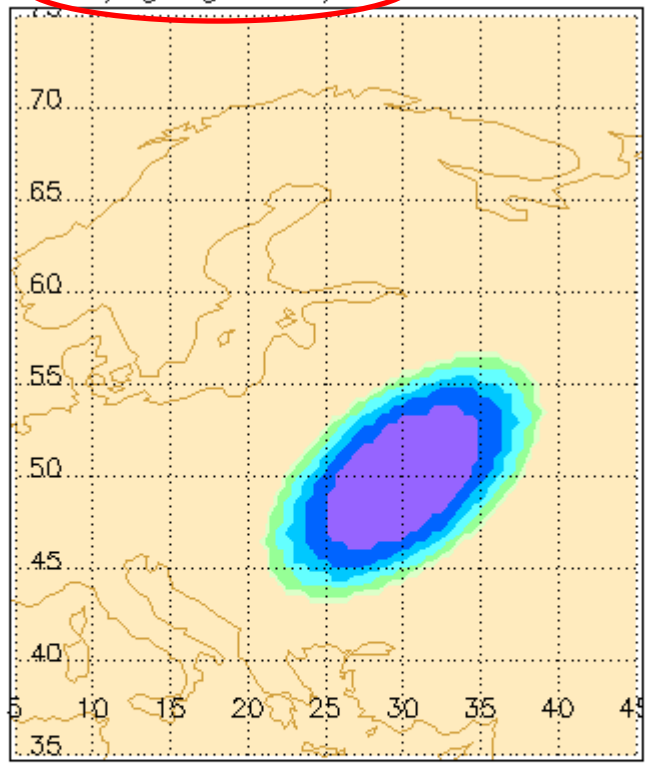
RMS = 0

POD = 1, FAR = 0, TS = 1

# Artificial example…






No data
<0.3 mm
0.3–1 mm
1–2 mm
2–5 mm
5–10 mm
10–20 mm
20–50 mm
50–100 mm
100–150 mm
150–200 mm

| Forecast \ Observed | ≧1 | <1 |
|---|---|---|
| ≧1 | 448 | 185 |
| <1 | 185 | 5743 |

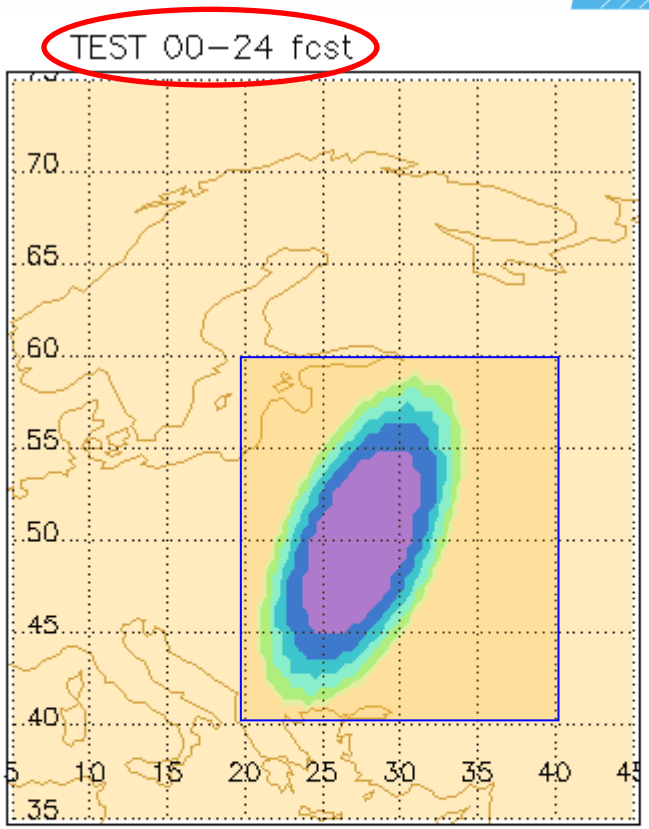
Validation statistics for 0 n=6561 Verif. grid=0.500°

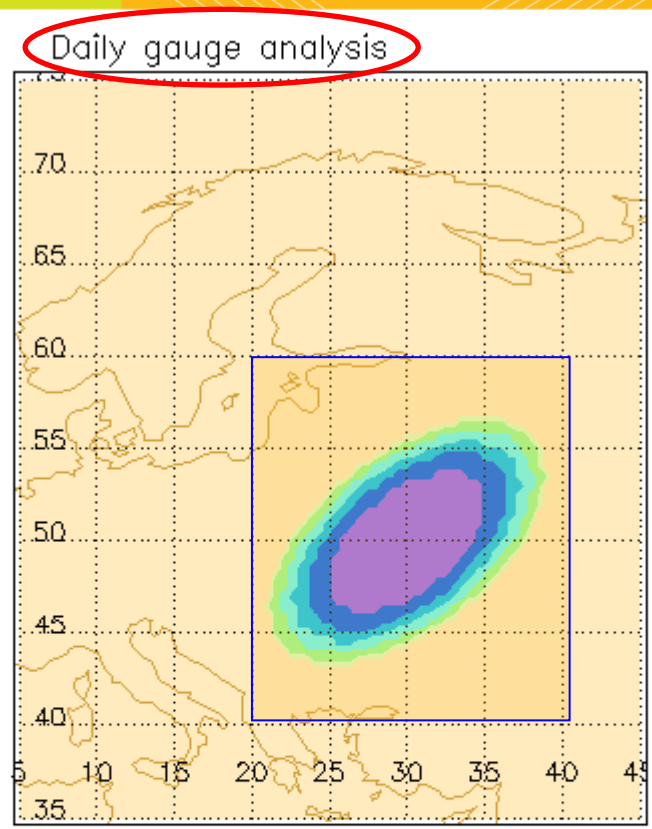
| | Analysed | Forecast |
|---|---|---|
| # gridpoints raining | 633 | 633 |
| Average rainrate (mm/d) | 15.05 | 15.05 |
| Rain volume (km³) | 16.83 | 16.83 |
| Maximum rain (mm/d) | 41.50 | 41.50 |
| Max 0.255° rain (mm/d) | 41.50 | |

Mean abs error = 1.19 mm/d
RMS error = 4.50 mm/d
Correlation coeff = 0.727
Bias score = 1.000
Probability of detection = 0.708
False alarm ratio = 0.292
Hanssen & Kuipers score = 0.677
Equitable threat score= 0.511

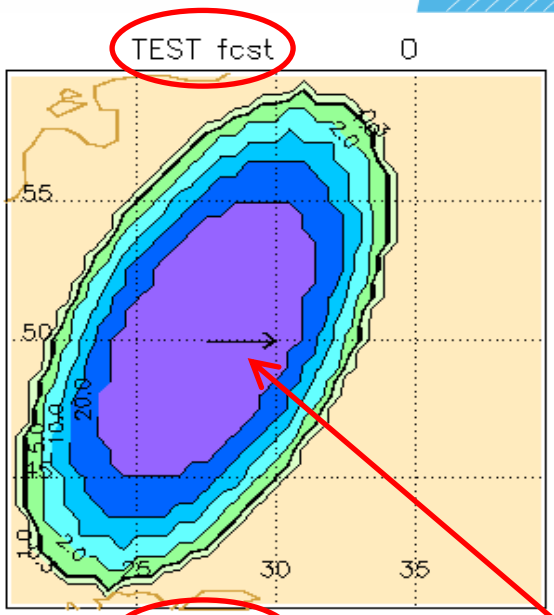
# Artificial example…

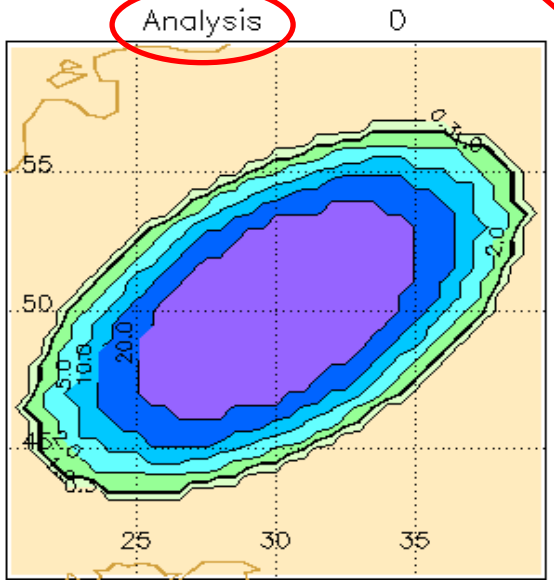


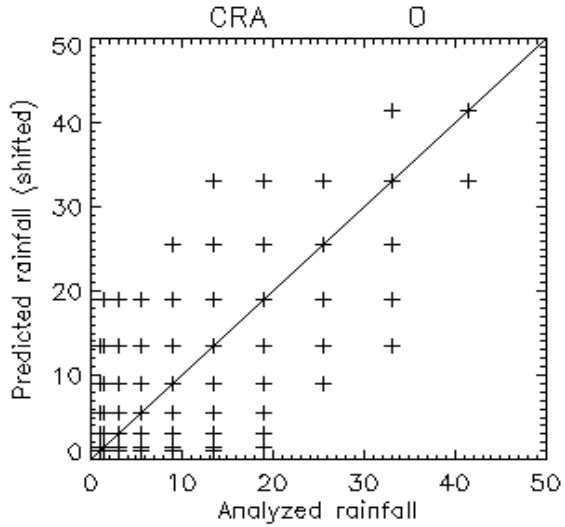


# …Sub-domain




TEST 00–24 fcst 0 n=879
(41.00°,21.00°) to (59.00°,39.00°)
Verif. grid=0.500° CRA threshold=1.0 mm/d

| | Analysed | Forecast |
|---|---|---|
| # gridpoints ≧1 mm/d | 633 | 633 |
| Average rainrate (mm/d) | 15.05 | 15.05 |
| Maximum rain (mm/d) | 41.50 | 41.50 |
| Rain volume ($km^3$) | 18.86 | 18.86 |

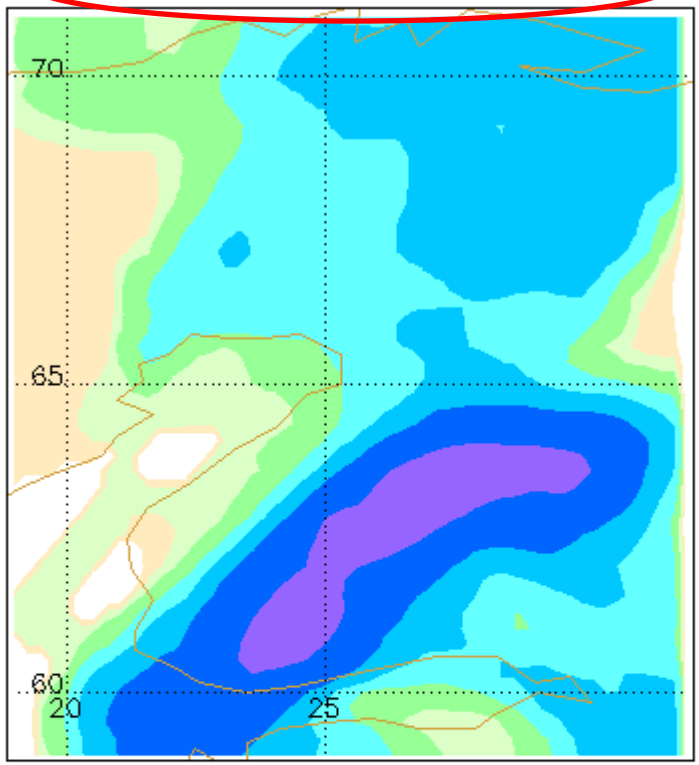
Displacement (E,N) = [−2.50°,0.00°]

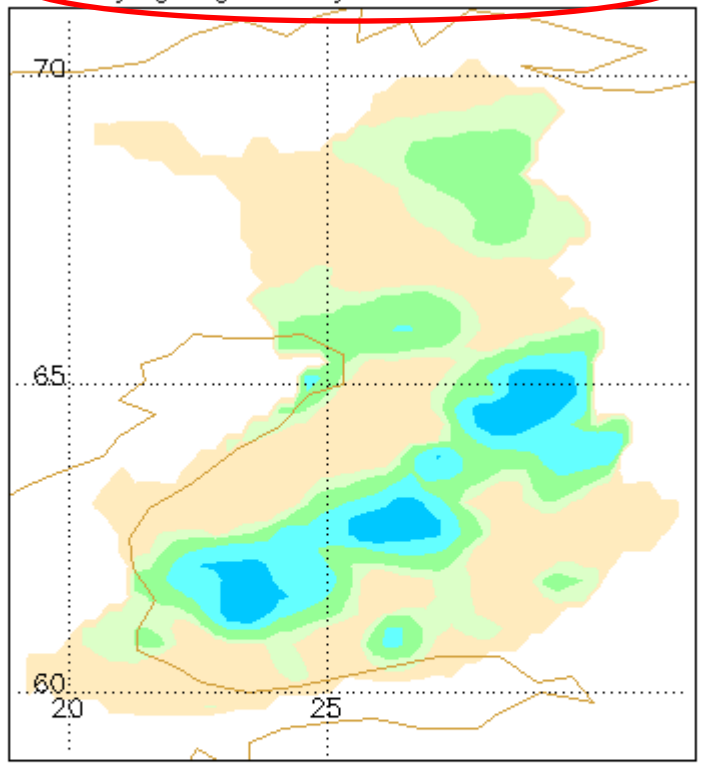
| | Original | Shifted |
|---|---|---|
| RMS error (mm/d) | 12.30 | 6.74 |
| Correlation coefficient | 0.569 | 0.870 |

| Error Decomposition: | |
|---|---|
| Displacement error | 70.0% |
| Volume error | 0.0% |
| Pattern error | 30.0% |

# “True life” example






No data
<1 mm
1–2 mm
2–5 mm
5–10 mm
10–20 mm
20–50 mm
50–100 mm
100–150 mm
150–200 mm

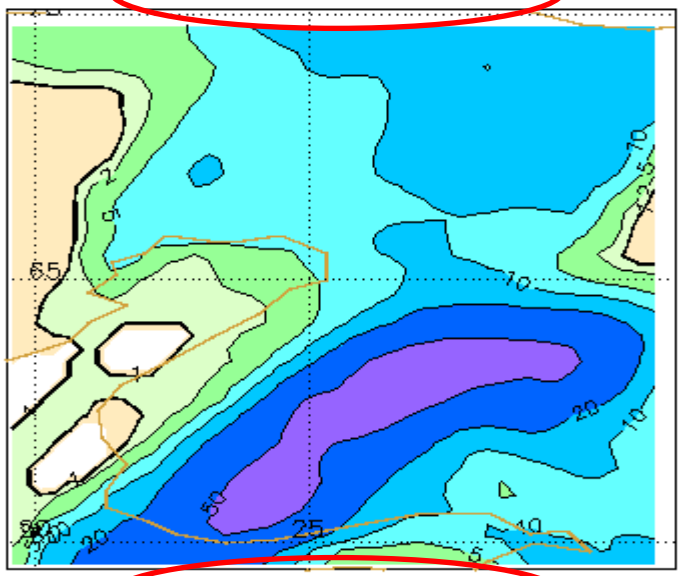
| Forecast \ Observed | ≥1 | <1 |
|---|---|---|
| ≥1 | 224 | 188 |
| <1 | 2 | 6 |

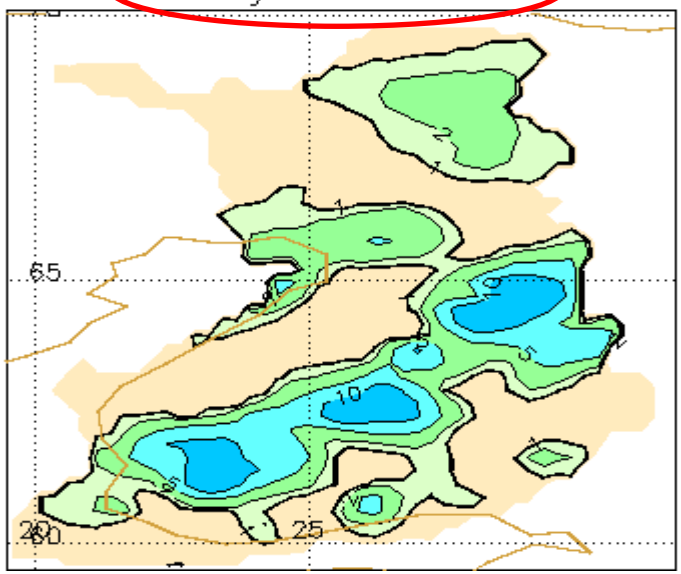
Validation statistics for 20030728 n=420 Verif. grid=0.400°

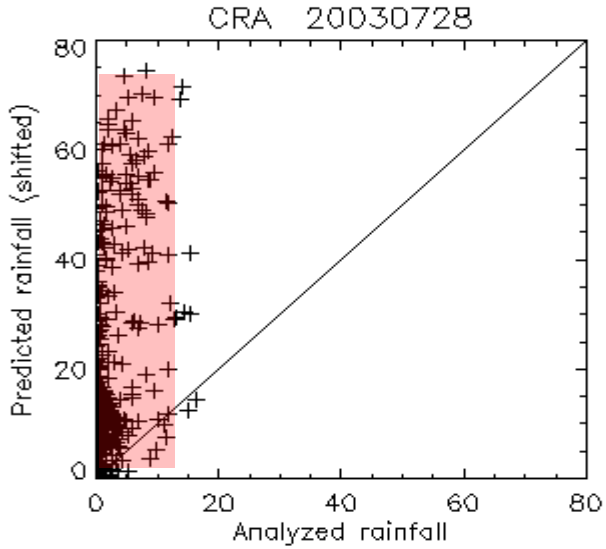
| | Analysed | Forecast |
|---|---|---|
| # gridpoints raining | 226 | 412 |
| Average rainrate (mm/d) | 4.51 | 20.46 |
| Rain volume (km³) | 0.85 | 7.02 |
| Maximum rain (mm/d) | 16.24 | 74.46 |
| Max 0.25° rain (mm/d) | 18.07 | |

Mean abs error = 17.80 mm/d
RMS error = 24.88 mm/d
Correlation coeff = 0.430
Bias score = 1.823
Probability of detection = 0.991
False alarm ratio = 0.456
Hanssen & Kuipers score = 0.022
Equitable threat score= 0.012

# …Sub-domain




ECMWF 12–24 fcst 20030728 n=409
(60.00°,20.00°) to (69.60°,31.20°)
Verif. grid=0.400° CRA threshold=1.0 mm/d

| | Analysed | Forecast |
|---|---|---|
| # gridpoints ≧1 mm/d | 226 | 407 |
| Average rainrate (mm/d) | 4.49 | 20.68 |
| Maximum rain (mm/d) | 16.24 | 74.46 |
| Rain volume ($km^3$) | 0.85 | 7.06 |

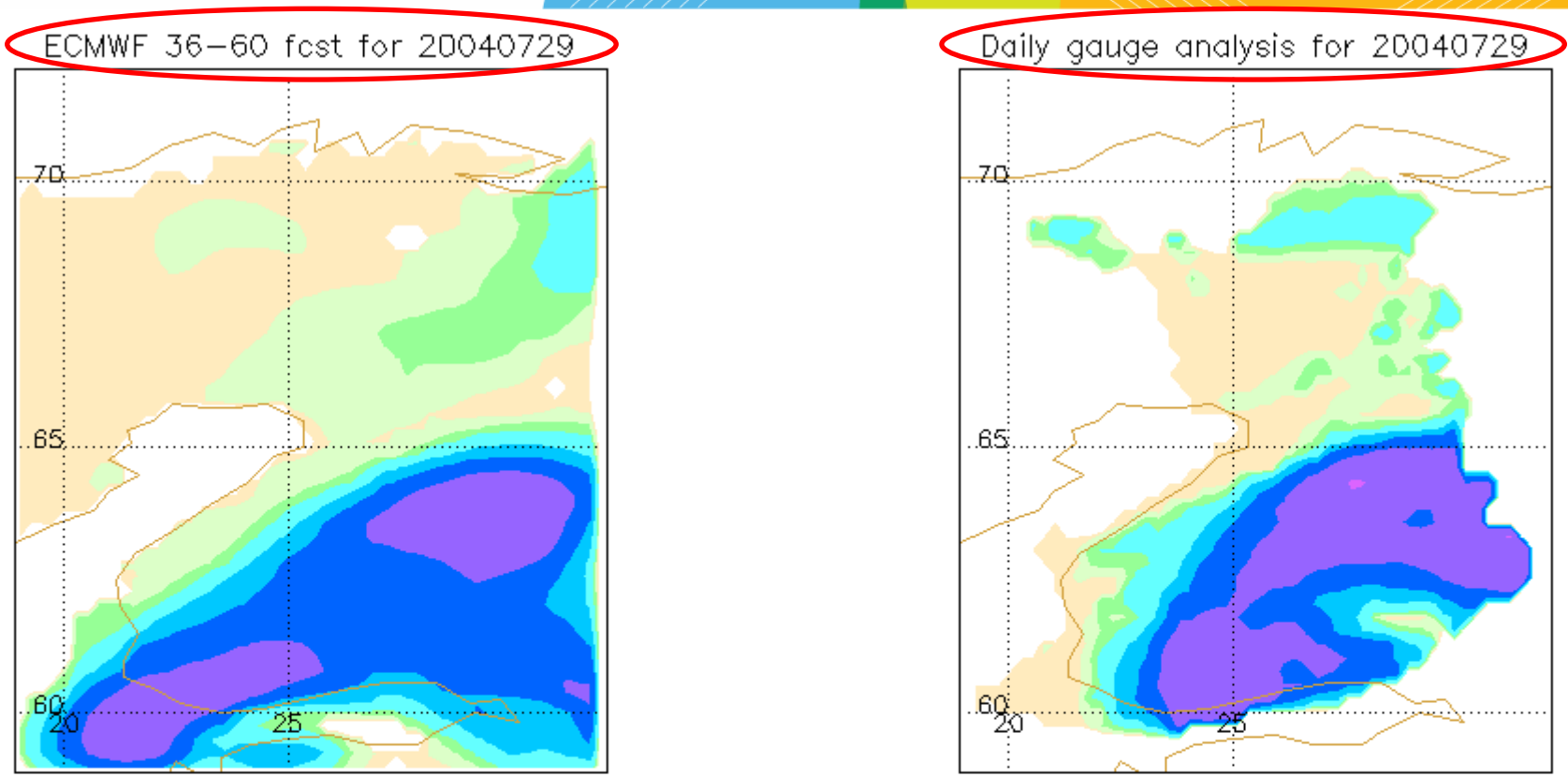
Displacement (E,N) = [0.00°,0.00°]

| | Original | Shifted |
|---|---|---|
| RMS error (mm/d) | 25.21 | 25.21 |
| Correlation coefficient | 0.422 | 0.422 |

| Error Decomposition: | |
|---|---|
| Displacement error | 0.0% |
| Volume error | 51.5% |
| Pattern error | 48.5% |

# “Flooding” example






| Forecast \ Observed | ≧10 | <10 |
|---|---|---|
| ≧10 | 134 | 39 |
| <10 | 35 | 245 |

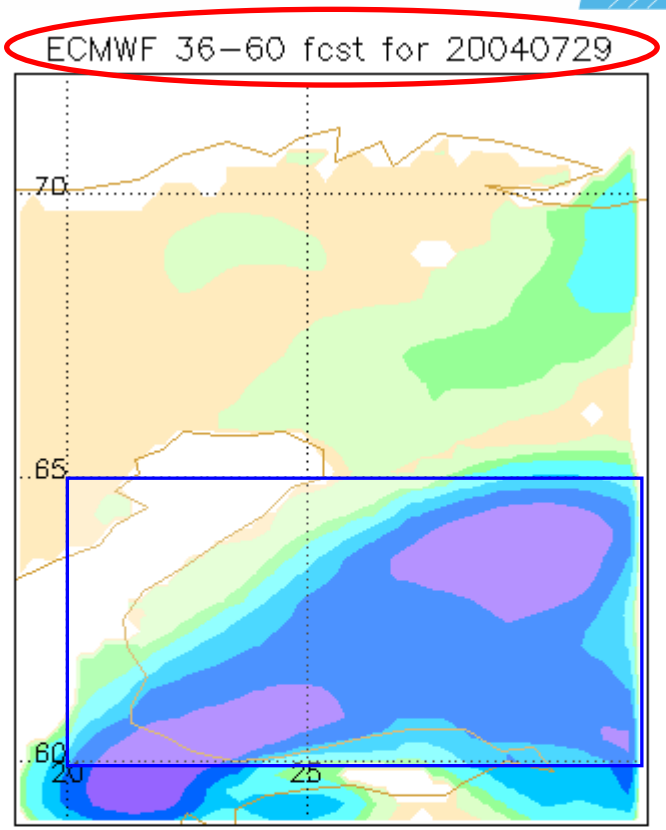
Validation statistics for 20040729 n=453 Verif. grid=0.400°

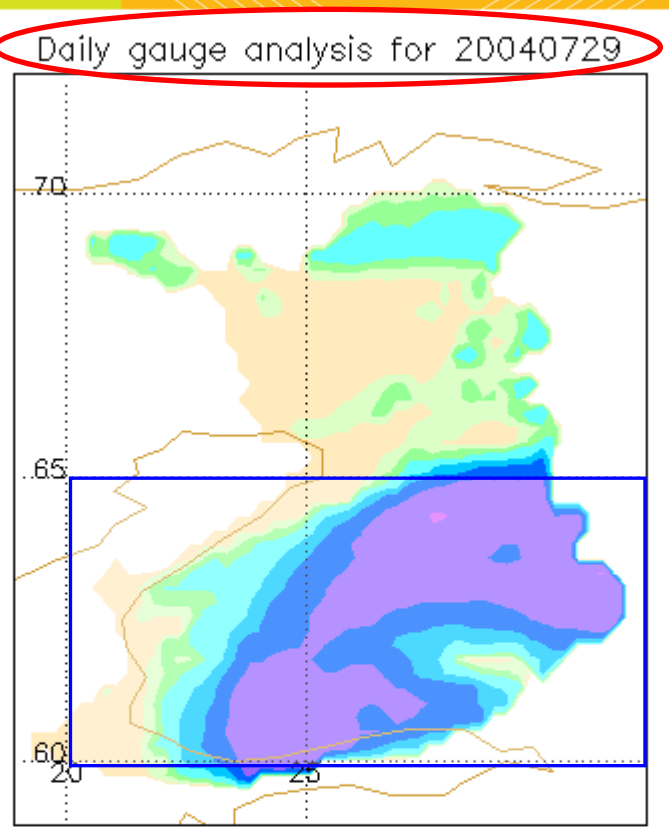
| | Analysed | Forecast |
|---|---|---|
| # gridpoints raining | 169 | 173 |
| Average rainrate (mm/d) | 27.44 | 22.89 |
| Rain volume (km³) | 3.79 | 3.24 |
| Maximum rain (mm/d) | 48.25 | 49.64 |
| Max 0.255° rain (mm/d) | 53.80 | |

Mean abs error = 6.02 mm/d
RMS error = 10.00 mm/d
Correlation coeff = 0.662
Bias score = 1.024
Probability of detection = 0.793
False alarm ratio = 0.225
Hanssen & Kuipers score = 0.656
Equitable threat score= 0.484

# “Flooding” example

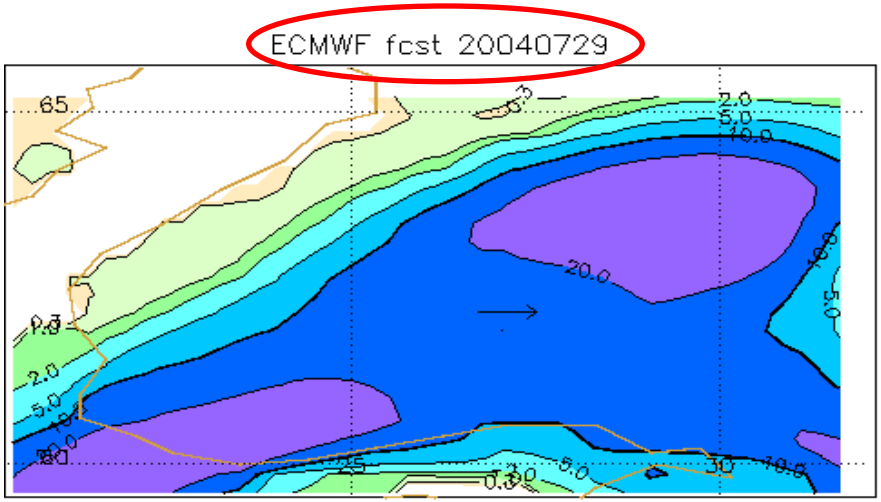


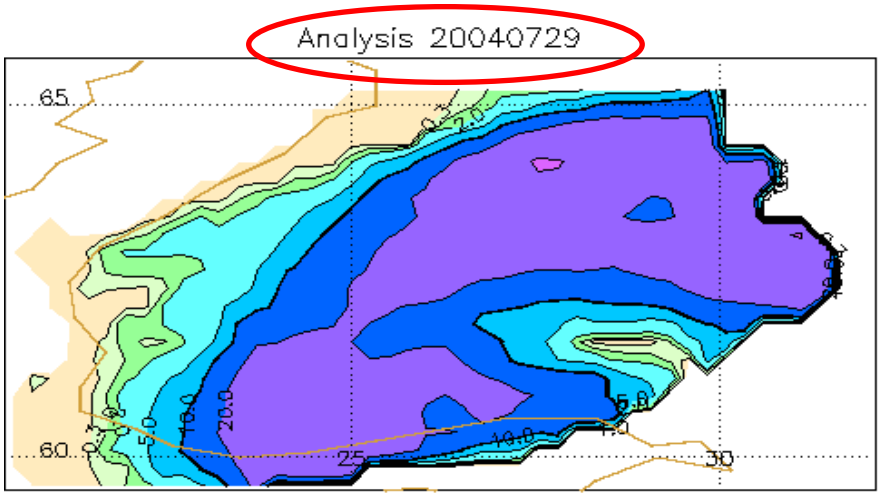


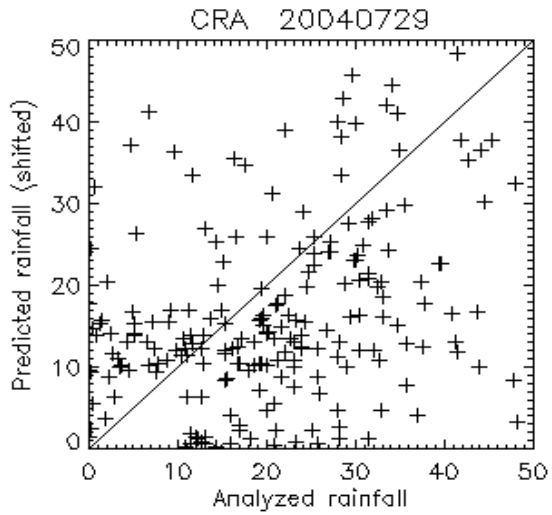



# …Sub-domain






ECMWF 36–60 fcst 20040729 n=210
(60.00°,20.80°) to (65.20°,31.60°)
Verif. grid=0.400° CRA threshold=10.0 mm/d

| | Analysed | Forecast |
|---|---|---|
| # gridpoints ≥10 mm/d | 169 | 173 |
| Average rainrate (mm/d) | 25.65 | 19.74 |
| Maximum rain (mm/d) | 48.25 | 48.46 |
| Rain volume (km³) | 3.93 | 3.10 |

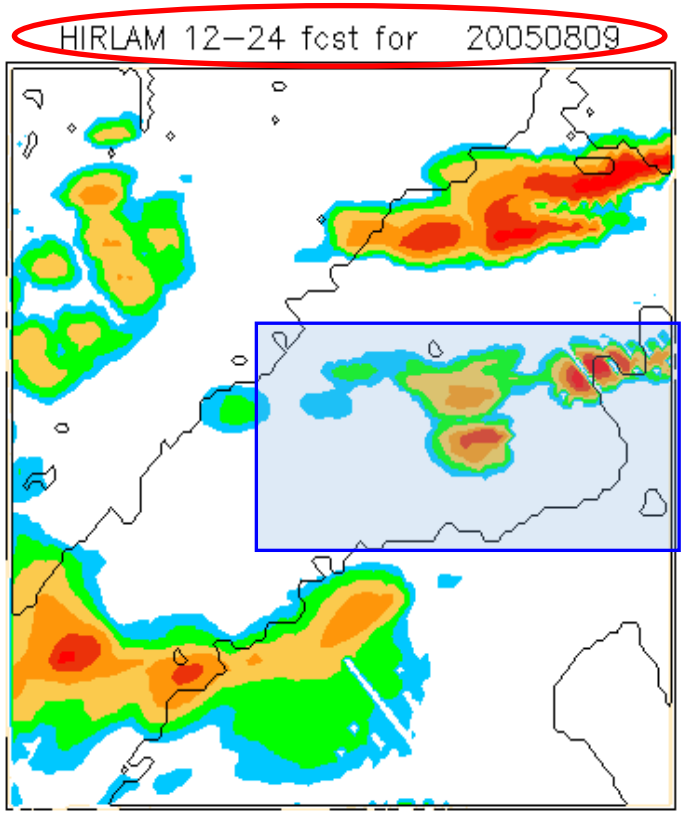
Displacement (E,N) = [−0.80°,0.00°]

| | Original | Shifted |
|---|---|---|
| RMS error (mm/d) | 14.54 | 13.93 |
| Correlation coefficient | 0.197 | 0.313 |

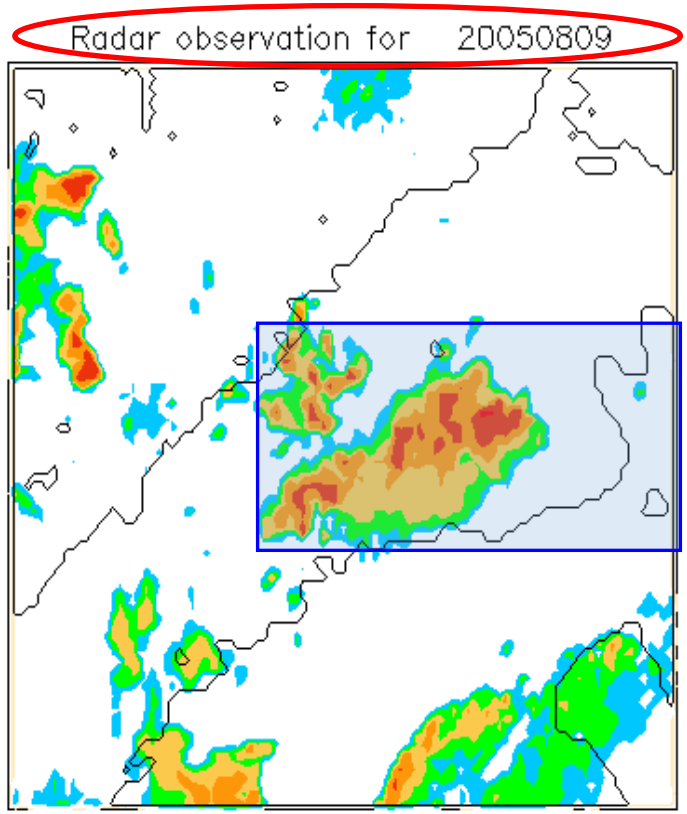
Displacement may be wrong – correlation not signif.

| Error Decomposition: | |
|---|---|
| Displacement error | 8.1% |
| Volume error | 9.1% |
| Pattern error | 82.8% |

# Radar data application…




**Radar simulator**




No data
<8 dbZ
8–16 dbZ
16–24 dbZ
24–32 dbZ
40–48 dbZ
>48 dbZ

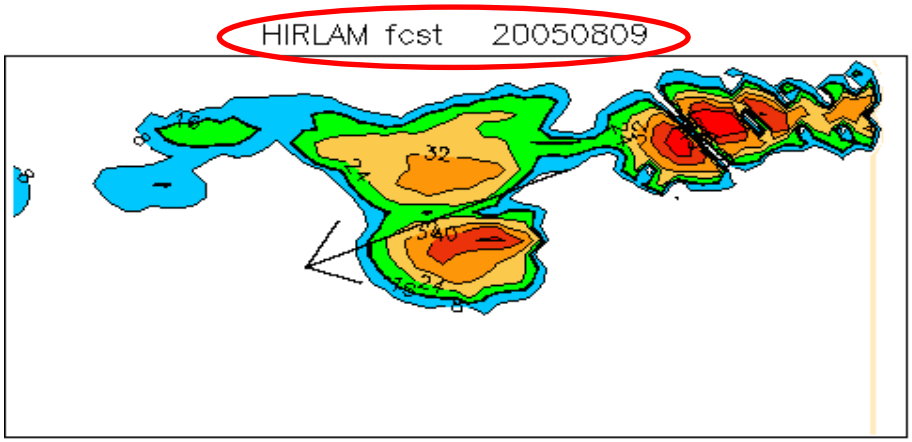
| Forecast \ Observed | ≧16 | <16 |
|---|---|---|
| ≧16 | 310 | 1838 |
| <16 | 1190 | 5651 |

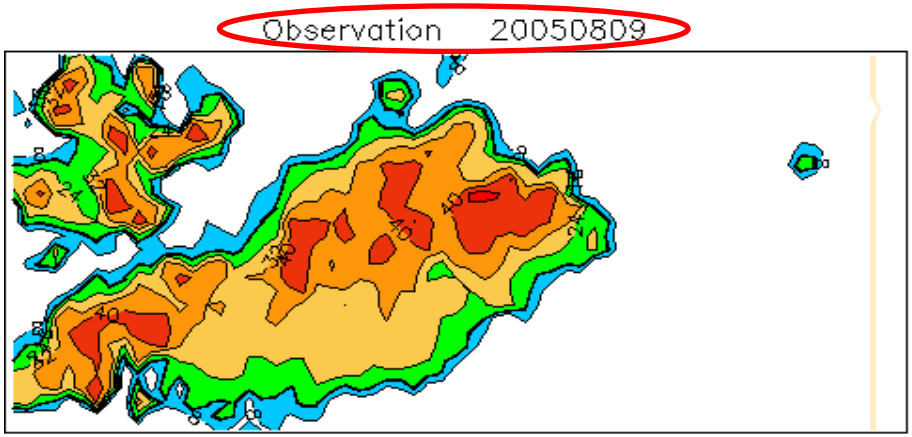
Validation statistics for 20050809 n=8989 Verif. grid=0.025°

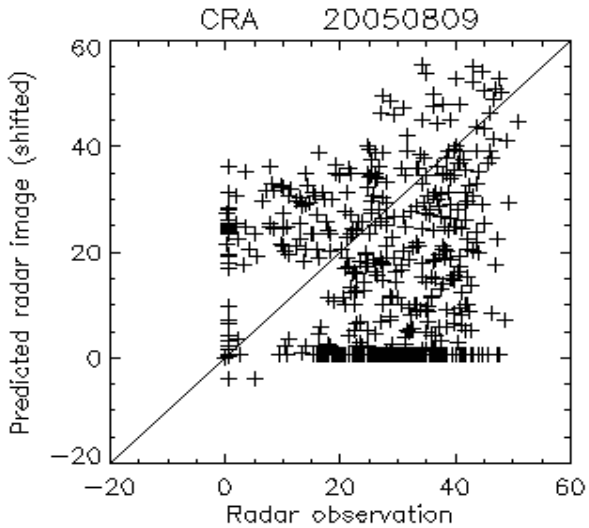
| | Observed | Forecast |
|---|---|---|
| # gridpoints ≧ 16dBZ | 1500 | 2148 |
| Average (dBZ) | 36.49 | 35.17 |
| | ****** | ****** |
| Maximum (dBZ) | 50.88 | 56.88 |
| Max 0.025° Z (dBZ) | 50.88 | |

Mean abs error = 10.77 dBZ
RMS error = 16.42 dBZ
Correlation coeff = −0.022
Bias score = 1.432
Probability of detection = 0.207
False alarm ratio = 0.856
Hanssen & Kuipers score = −0.039
Equitable threat score= −0.016

# …Sub-domain




HIRLAM 12−24 fcst 20050809 n=729
(2.18°,12.45°) to (2.80°,14.00°)
Verif. grid=0.025° CRA threshold=16.0 dBZ

| | Observation | Forecast |
|---|---|---|
| # gridpoints ≧16 dBZ | 528 | 269 |
| Average (dBZ) | 32.04 | 31.80 |
| Maximum (dBZ) | 50.88 | 55.53 |
| ---------- | ****** | ****** |

Displacement (E,N) = [0.47°,0.17°]

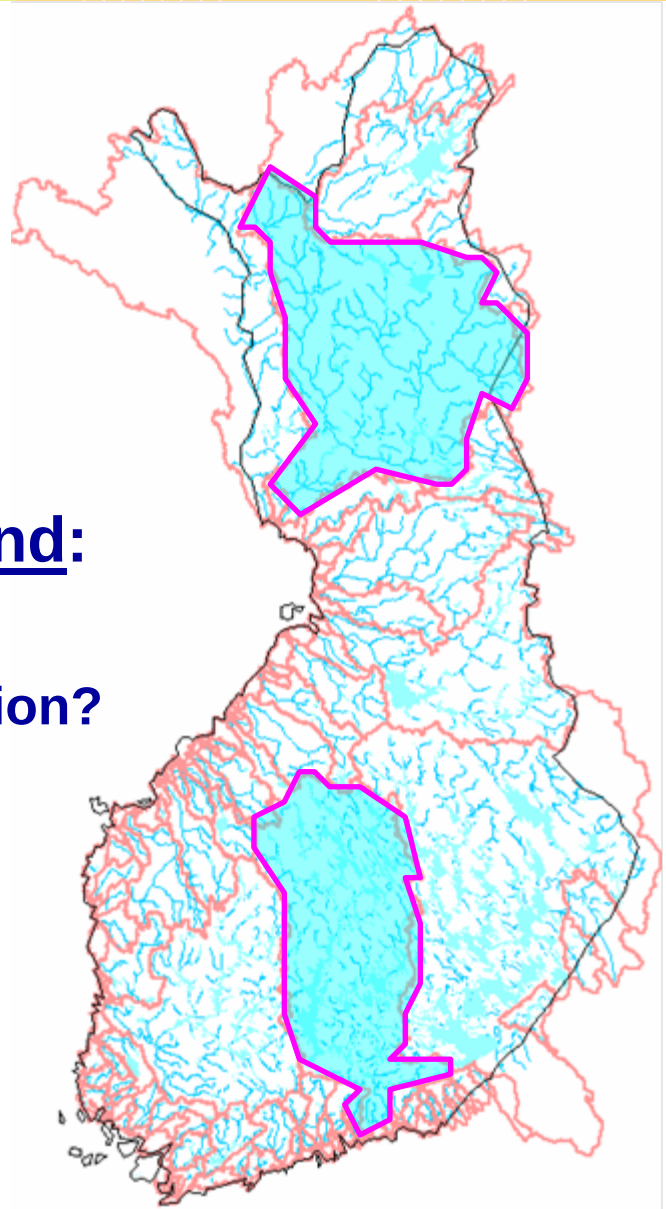
| | Original | Shifted |
|---|---|---|
| RMS error (dBZ) | 25.69 | 20.38 |
| Correlation coefficient | −0.241 | 0.344 |

Displacement may be wrong − correlation not signif.

Error Decomposition:

| | |
|---|---|
| Displacement error | 37.0% |
| Volume error | 19.9% |
| Pattern error | 43.0% |

# Hydrological applicability



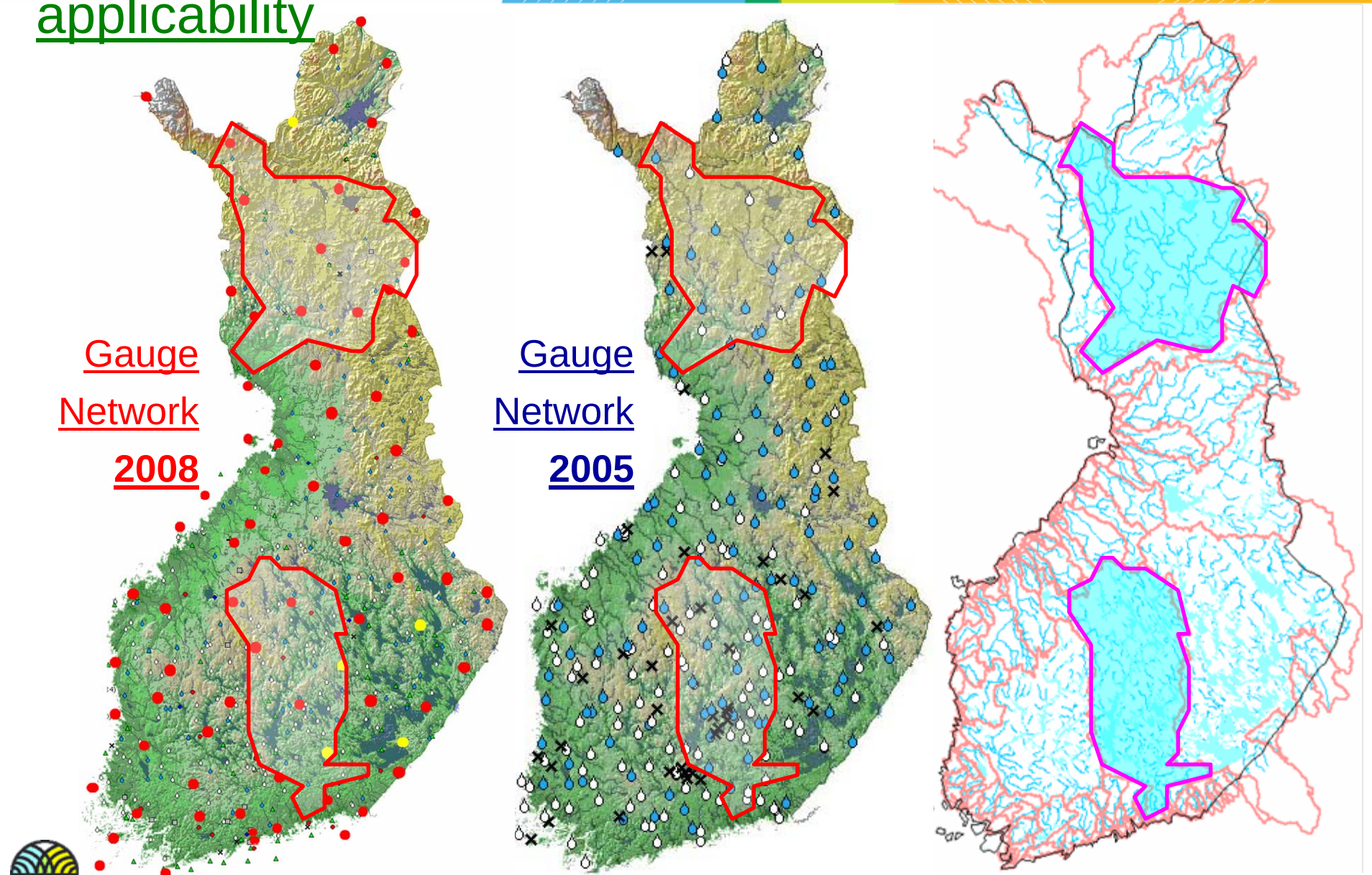
## Catchments areas in Finland:

⇨ **Potential sub-domains for entity-based (CRA) verification?**

# Hydrological applicability

Gauge Network **2008**

Gauge Network **2005**



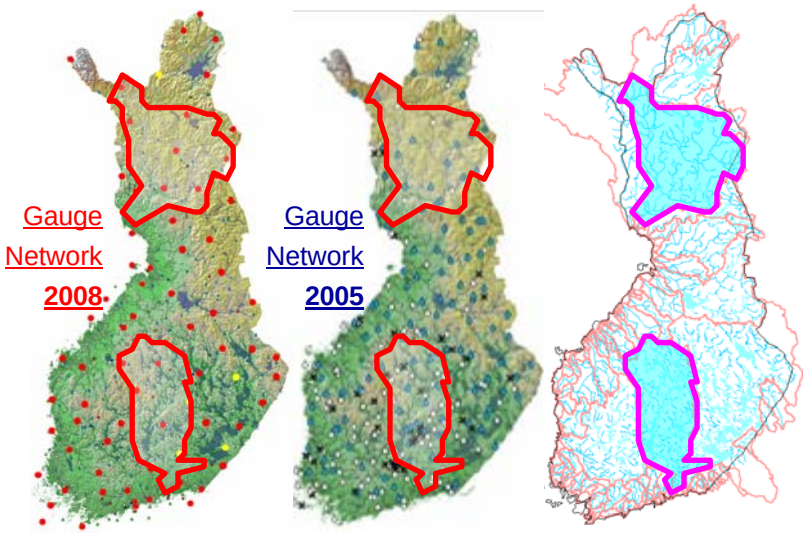
# Hydrological applicability

## Precipitation: From observations to analysis

- ✓ Gauge network ?
- ✓ Radar network ?
- ✓ Combination ?

## What is the observed "truth" (~ *Uncertainty*) ?

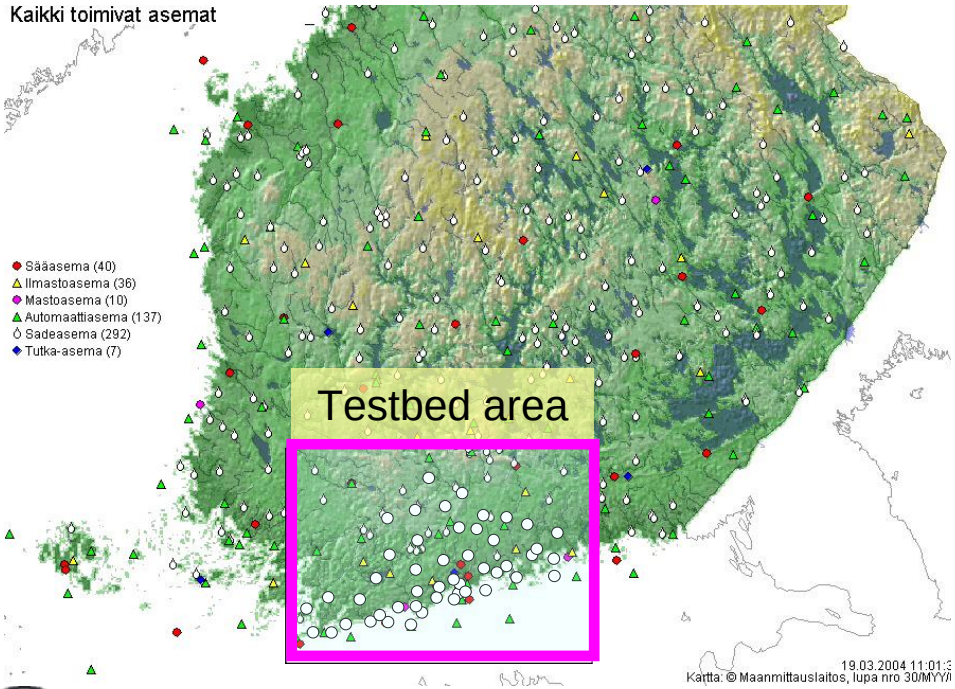
- ❖ Inevitable implications to verification results and their interpretation !



# Hydrological applicability

## Helsinki Testbed

- ✓ Mesoscale meteorology research project
- ✓ Intense meteorological measurement periods
- ⇨ Meteo-hydrological forecasting-verification studies in small catchments within the Testbed area ?




# Conclusion

- Knowledge (~uncertainty) about error components is crucial for model development
- Error decomposition may distinguish systematic errors, e.g.
  - ⇨ Orography induced precipitation
  - ⇨ Systematic deepening of lows
- Knowledge of error origin(s) is useful guidance to forecasters
- CRA method to be applied for hydrological applications
  - ⇨ Catchments defined as ”sub-domains” ⇔ Helsinki Testbed, a.o.
  - ⇨ COST-731 research application !
- CRA methodology still in ”research mode”, but…
  - ⇨ ”Semi-operational” at FMI this spring !
- **Recall always: What is the ”truth” ⇔ Uncertainty ?**

# Thank You !

**Pertti Nurmi**

Meteorological Research & Development

Finnish Meteorological Institute