Testing the performance of a spatial consistency test for data quality control

Cristian Lussana[(1)], Marta R. Salvati[(1)] and Francesco Uboldi[(2)]

(1) Lombardia's Public Weather Service, ARPA Lombardia, Milan (Italy)
(2) consultant, Novate Milanese, Milan (Italy)

Meteorological network

Automatic stations;

Complex orography;

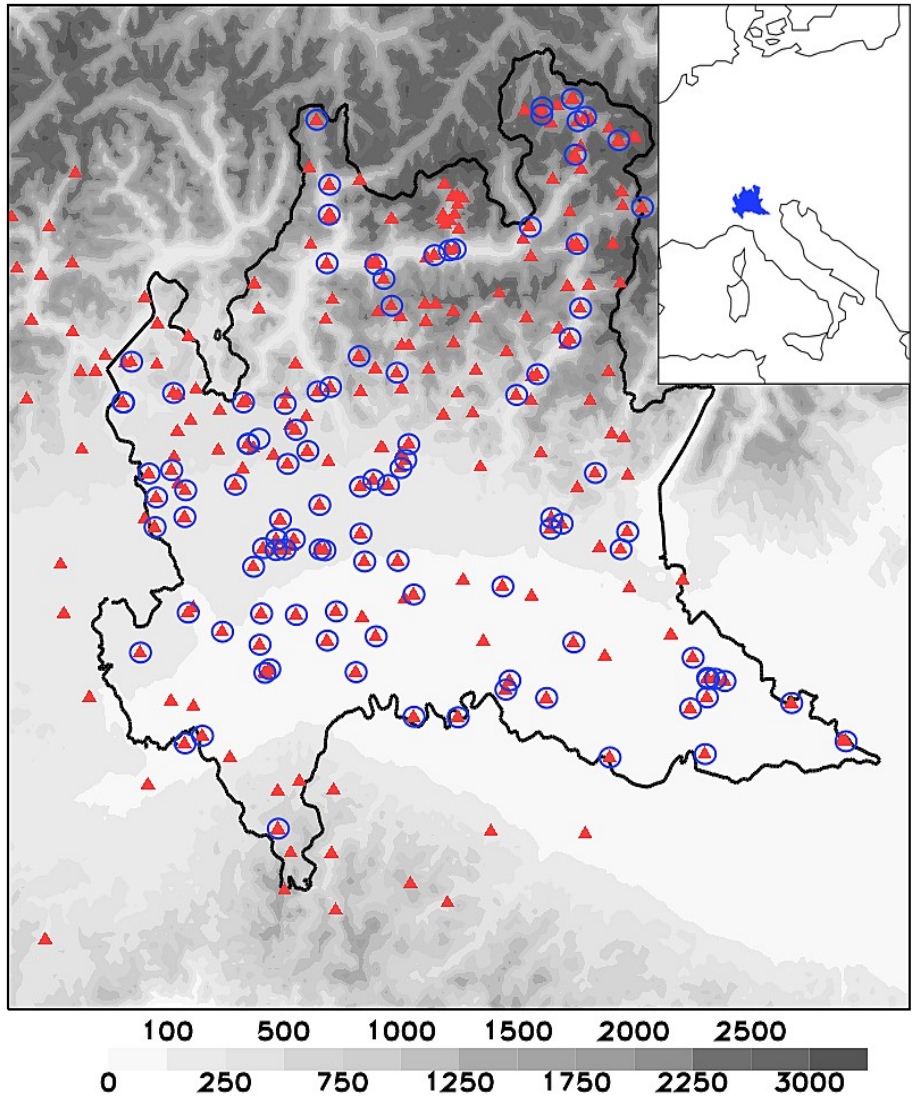
Hourly data;

High station density;

Station altitudes: 10m - >3000m AMSL;

Grid: 1.5 Km (177x174);

Grid orography from a high resolution DEM (250 m) without smoothing;



meteo ARPA

Meteorological network

Automatic stations;

Co

Ho

Hig

Sta

AM

Gri

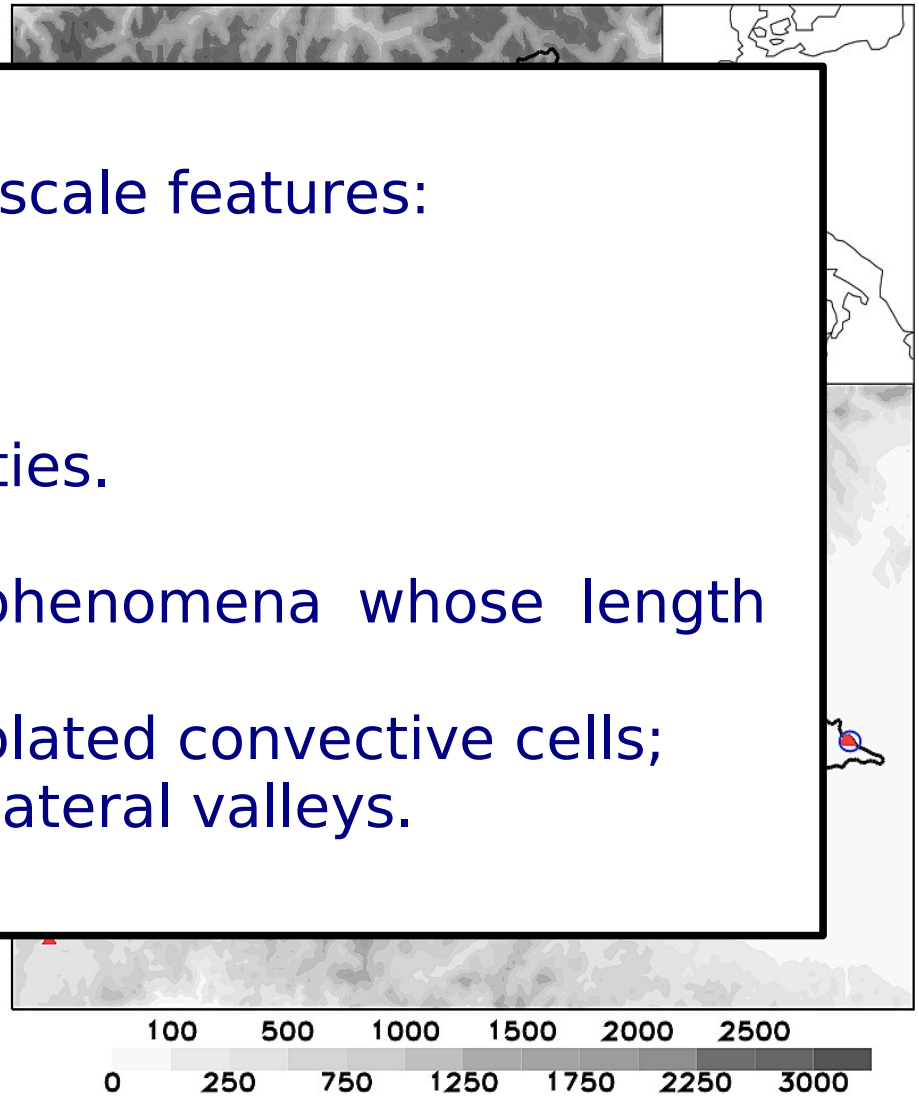
Gri

res

smoothing;

<u>efficiently resolves</u> most mesoscale features:
- major valley cold pools;
- thermal inversion;
- widespread foehn warming;
- heat-island effect of major cities.

<u>unable</u> to correctly resolve phenomena whose length scale < 10 Km:
- cold outflow associated to isolated convective cells;
- katabatic warmings in small lateral valleys.



Automated Quality Control

- Plausible value check
- Time consistency check 1: **step** (check on a maximum allowed variability of an hourly value)
- Time consistency check 2: **persistence** (check on a minimum required variability in a prescribed time interval)
- **Spatial Consistency Test (SCT)**

Decision Making Algorithms

Data disseminated to the users

Test, parameters and DMA implementation are variable dependent

Automated Quality Control *(details tomorrow in session AW6 !)*

- Plausible value check
- Time consistency check 1: **step** (check on a maximum allowed variability of an hourly value)
- Time consistency check 2: **persistence** (check on a minimum required variability in a prescribed time interval)
- **Spatial Consistency Test (SCT)**

Decision Making Algorithms

Data disseminated to the users

DMA-Temperature	1	2	3	4	5	6	7
Plausibility	P	P	P	P	F	-	-
SCT	P	P	W	W	-	F	-
Step	P	W	P	W	-	-	-
Persistence	P	P	P	P	-	-	F
Result	G	G	G	B	B	B	B

Test, parameters and DMA implementation are variable dependent

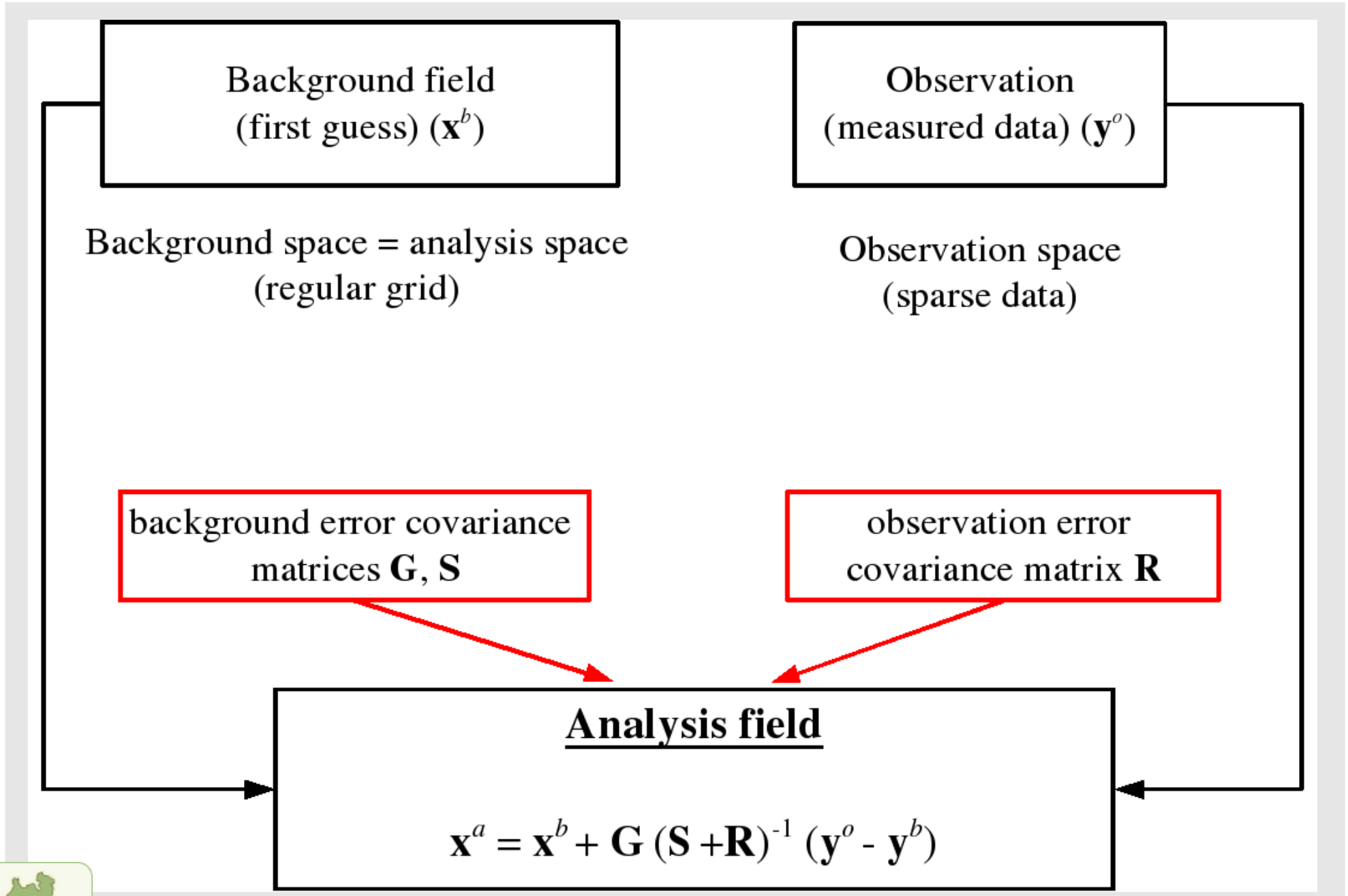
meteo ARPA

Spatial Consistency Test

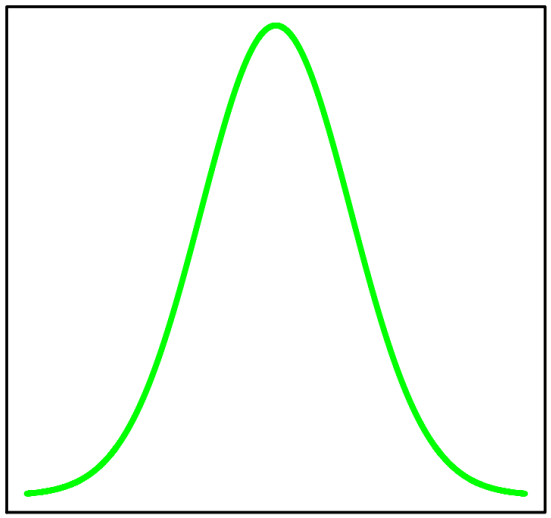
The SCT is based on a spatial analysis scheme (*Uboldi et al., Meteorol. Appl., 2008*), an implementation of Optimal Interpolation

Main features of the analysis scheme:

- Background information derived from observations detrending;
- Background error covariance specified by means of 3D gaussian correlation functions;
- Efficient computation algorithm.


Background field
(first guess) ($\mathbf{x}^b$)
Observation
(measured data) ($\mathbf{y}^o$)
Background space = analysis space
(regular grid)
Observation space
(sparse data)
background error covariance
matrices **G**, **S**
observation error
covariance matrix **R**
Analysis field
$\mathbf{x}^a = \mathbf{x}^b + \mathbf{G}\,(\mathbf{S} + \mathbf{R})^{-1}\,(\mathbf{y}^o - \mathbf{y}^b)$

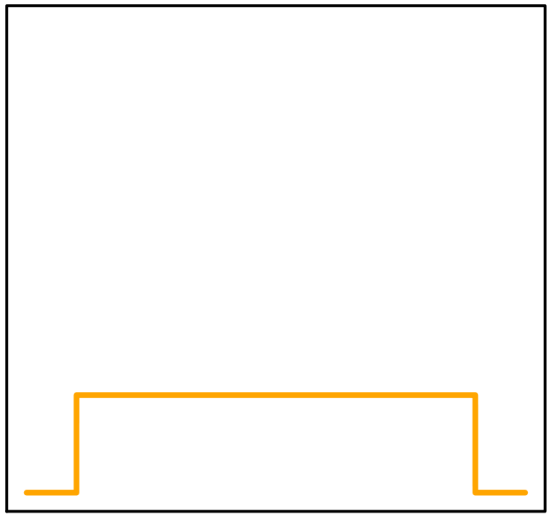
Observational Error Model *(Lorenc and Hammond, qjrms, 1988)*



Observation not affected by **Gross Error (GE)**

$$P(O|\overline{GE})$$

Gaussian *pdf*

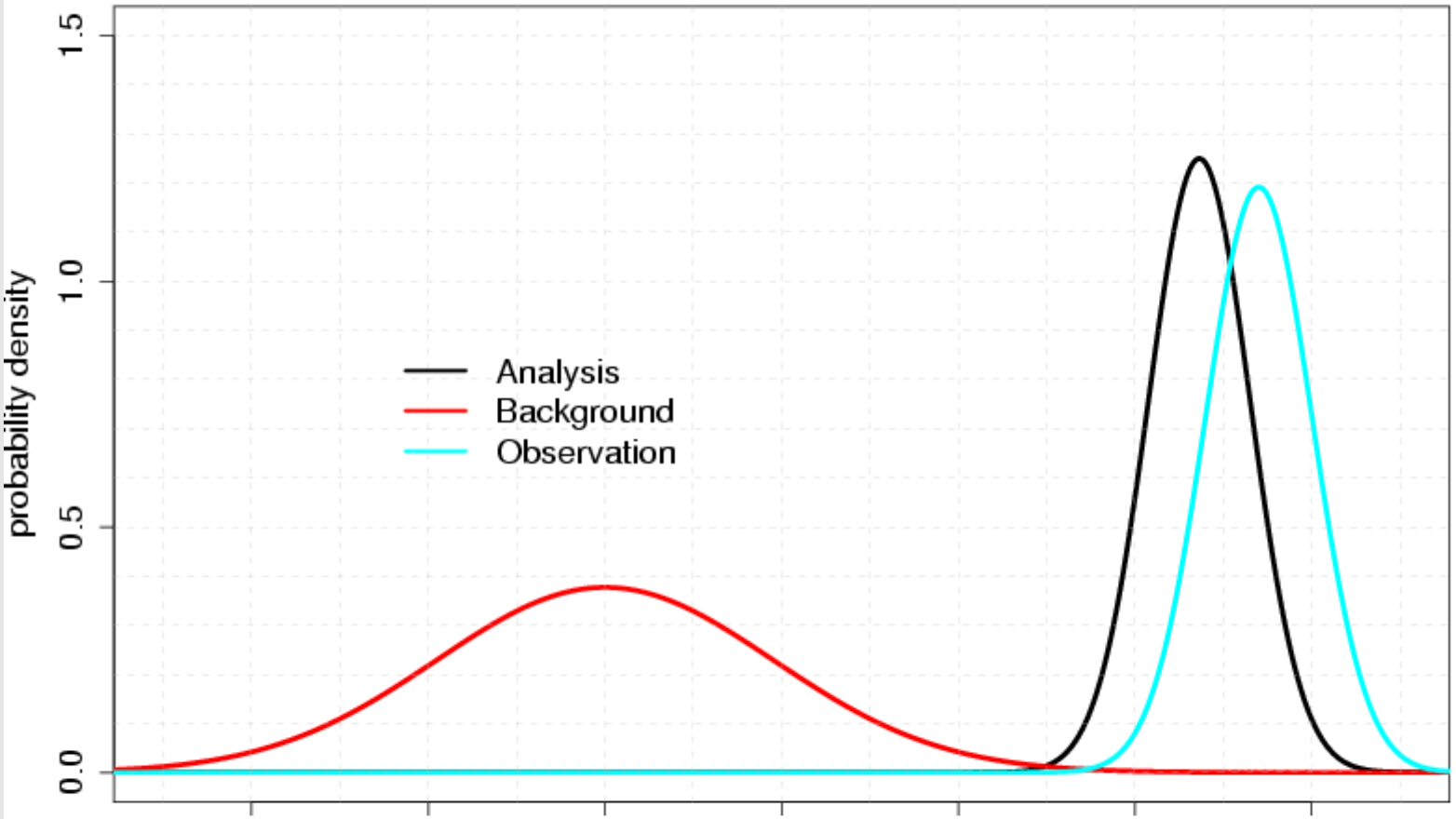


Observation affected by **GE**

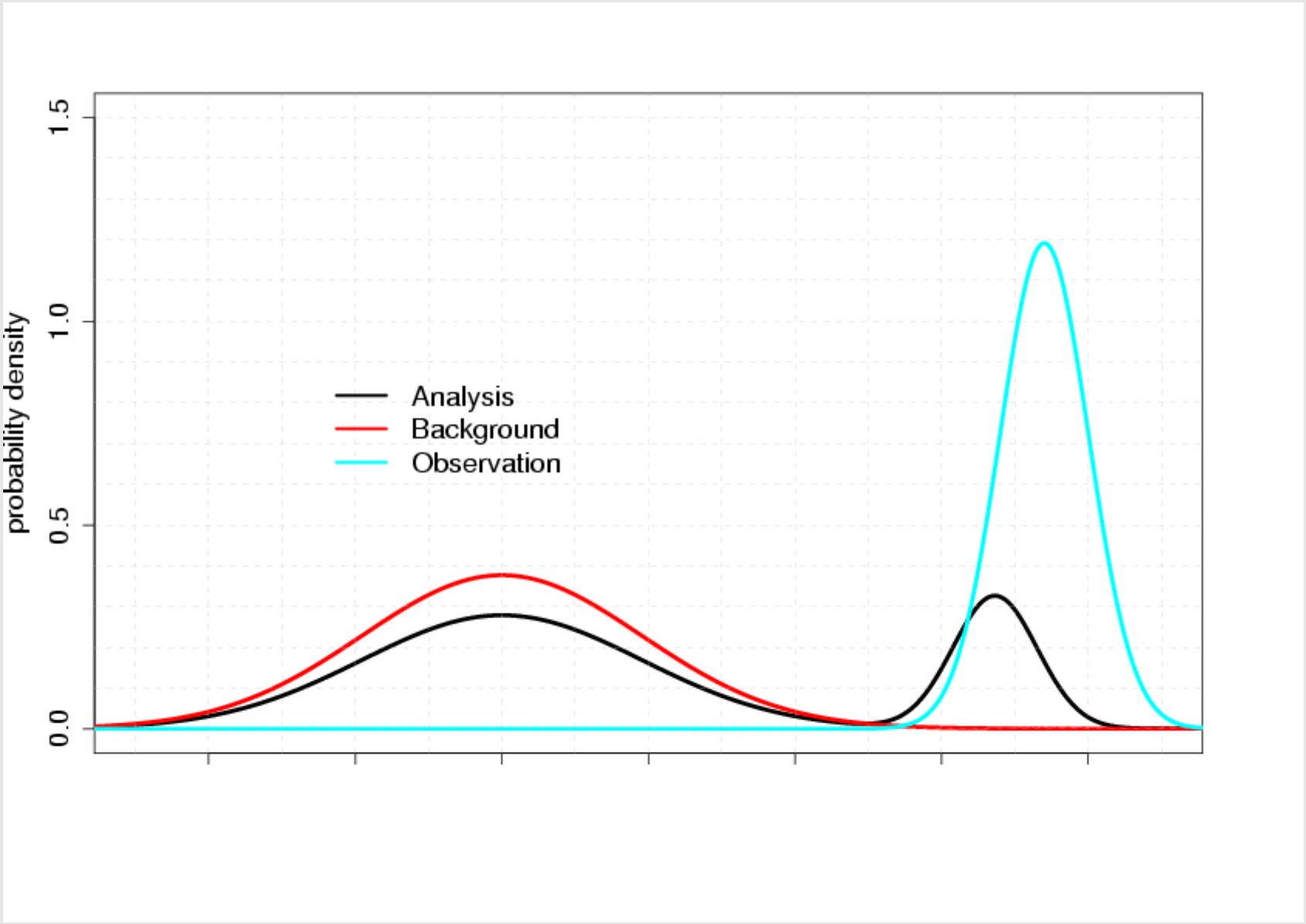
$$P(O|GE)$$

uniform *pdf*
over the climatological interval

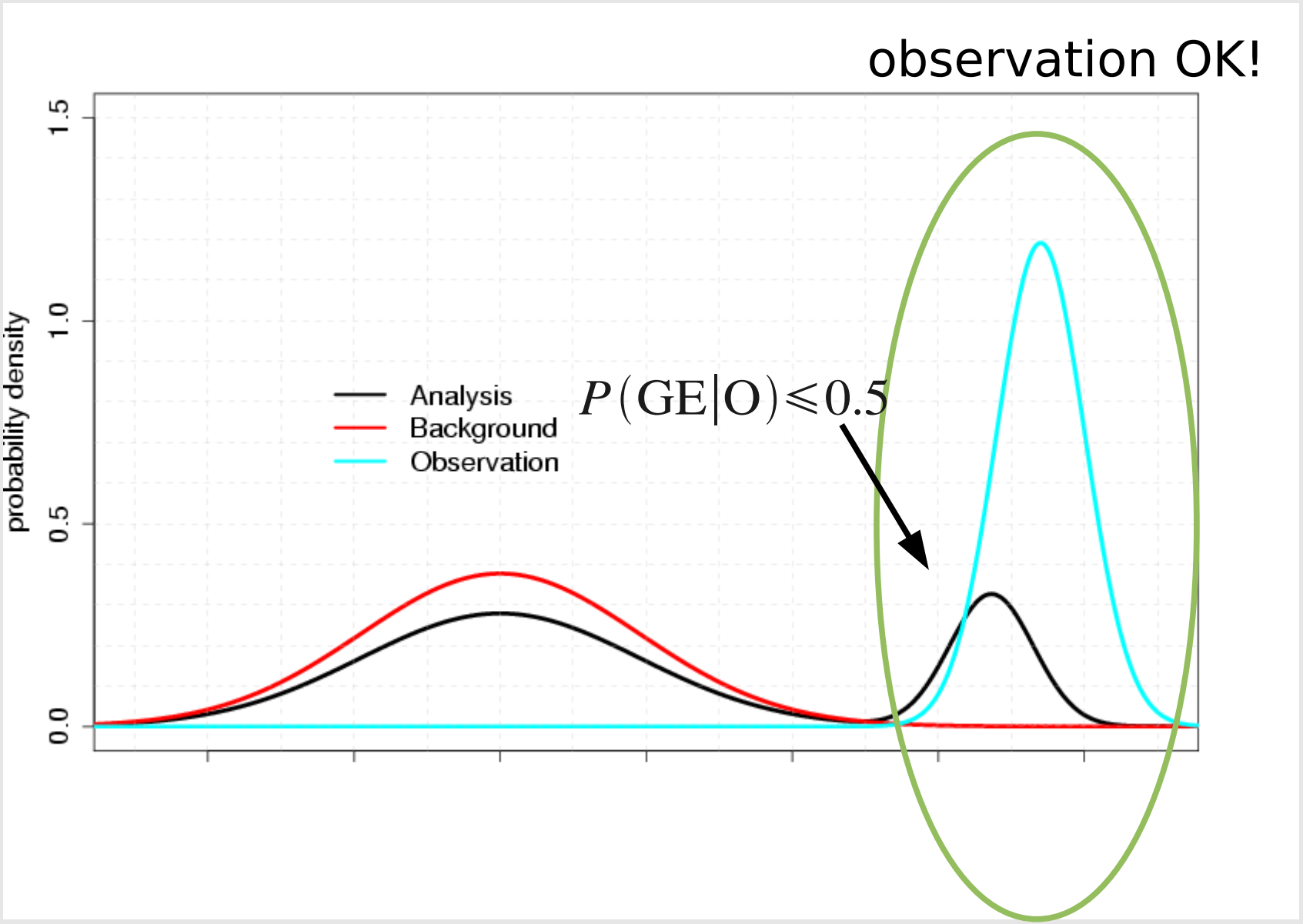
Observational Error Model: Gaussian only
GE present but NOT accounted for



Observational Error Model: accounting for a GE

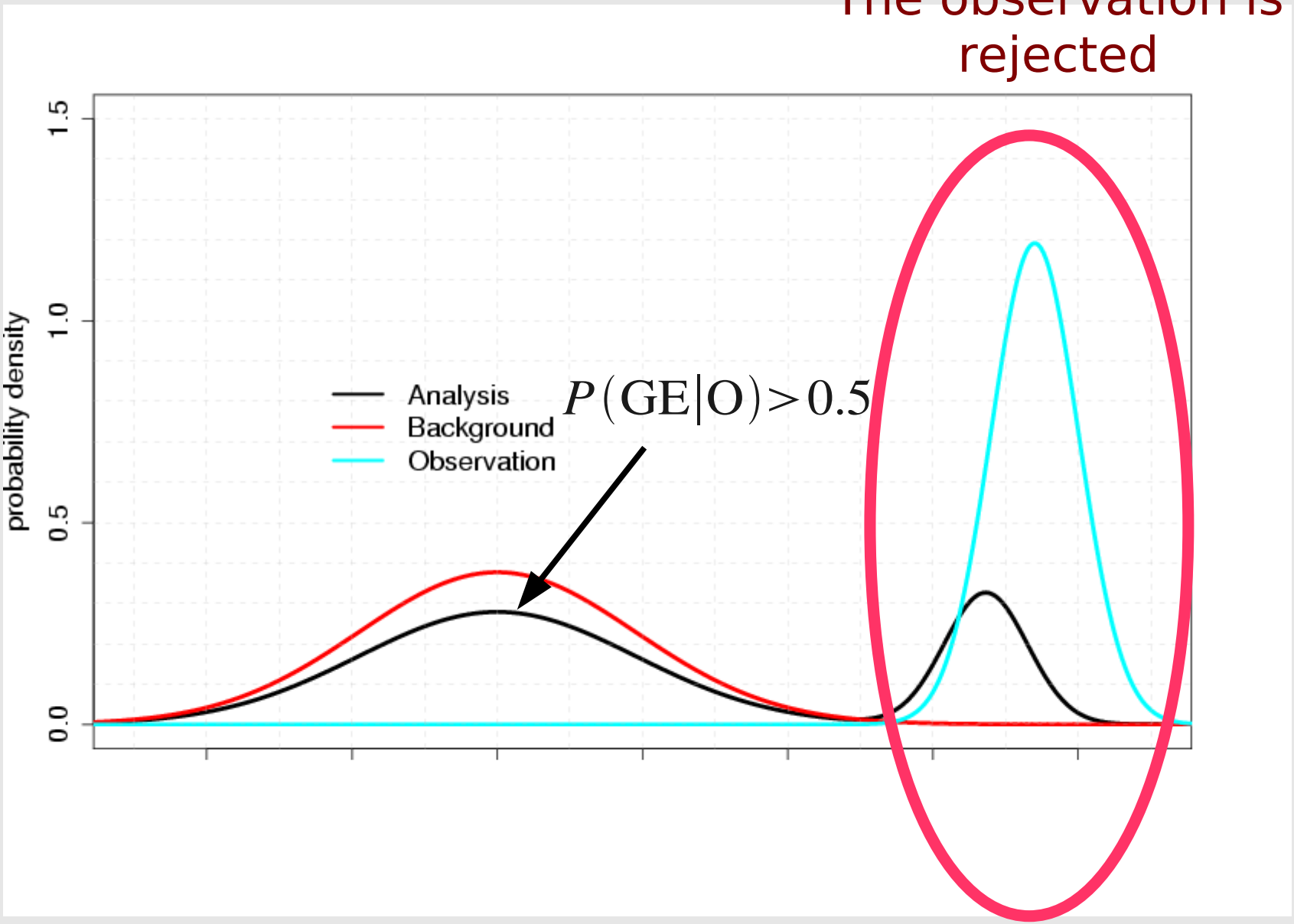


Observational Error Model



Observational Error Model

The observation is rejected



Cross Validation (CV) Analysis

In our case (no independent background field) the *a priori* estimate is the CV analysis:

$$y^{CVa}$$

The CV analysis is produced using ALL observations EXCEPT the observation undergoing the SCT.

As a consequence, observation error and CV analysis error are uncorrelated.

Spatial Consistency Test

$$(y^o - y^{CVa})^2 > T^2(\sigma_o^2 + \sigma_{CVa}^2)$$

Only T^2 , σ_o^2 are needed: objectively estimated from the statistical hypothesis and 3-year statistics

The SCT automatically accounts for local data density:

- completely isolated stations (CV analysis = background): *permissive*

$$(y^o - y^b)^2 > T^2(\sigma_o^2 + \sigma_b^2)$$

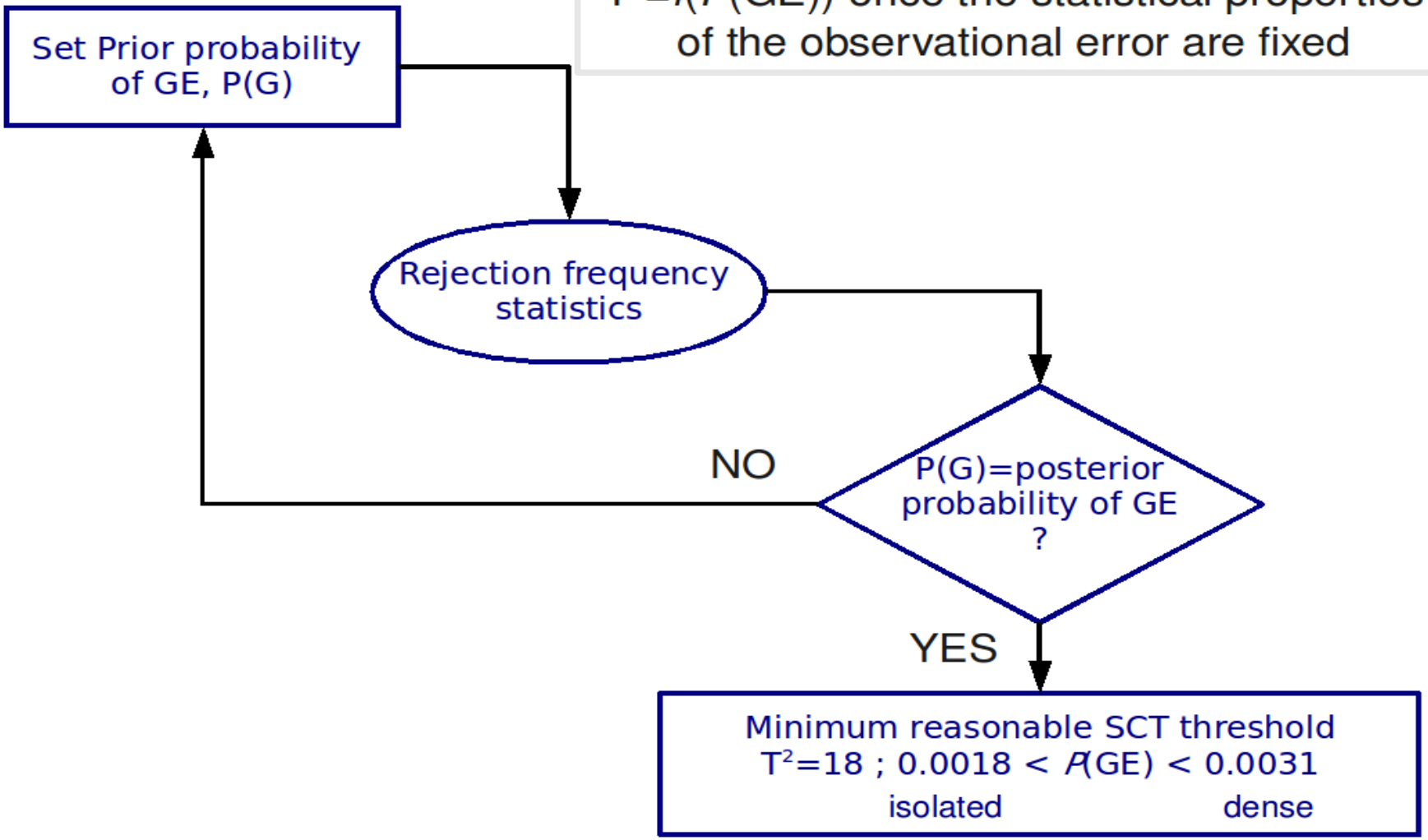
- totally redundant stations (CV analysis = analysis): *restrictive*

$$(y^o - y^{CVa})^2 = (y^o - y^a)^2 > T^2\sigma_o^2$$

meteo ARPA

T^2 estimation

$T^2 = f(P(GE))$ once the statistical properties of the observational error are fixed



SCT Performance monitoring

			SCT	
Persistence	Step	Pass	Warning	Fail
Pass	Pass	1151771	2568	1252
Pass	Warning	426	177	1285
Fail	Pass	4815	1716	4323
Fail	Warning	0	0	0

Total observation tested = 1165339

January – June 2009 temperature observations

meteo ARPA

SCT Performance monitoring

SCT rejection frequency = 0.0045 (expected: from 0.0018 to 0.0031)

Estimate *f*(pass|GE) = 0.29 (expected: from 0.13 to 0.23)

- Network management differences
- Temporal GE correlations
- Systematic errors

meteo ARPA

Conclusions

SCT:

- particularly efficient for a high density network
- based on clear statistical hypothesis (error model + OI)
- objective scheme for estimating thresholds and parameters
- *a priori* estimate of false and missed rejections
- local station density automatically taken into account

- the threshold estimation procedure also provides a measure of network reliability;
- preliminary comparison of test performance in the first 6 months of operational use with the 3 years statistics used to estimate test parameters show agreement in the order of magnitude but larger values of P(GE) in the network.
- large amplitude representativity errors

contact: c.lussana@arpalombardia.it