Norwegian Meteorological Institute

METreport

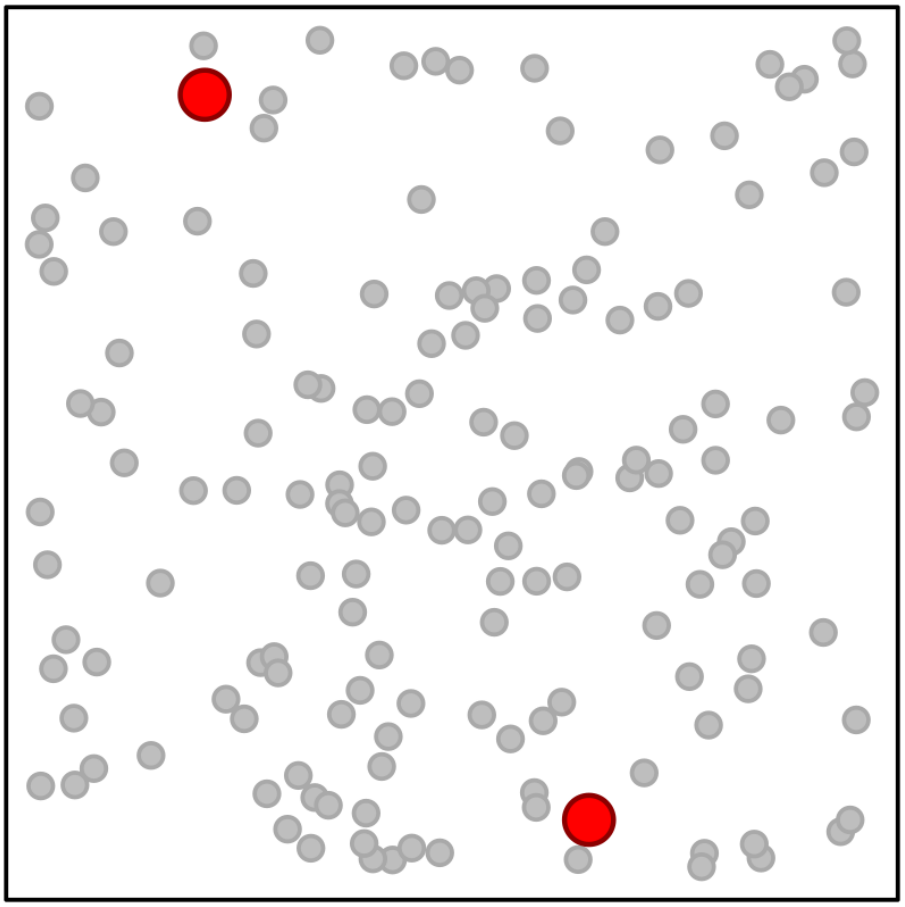
No. 12/2021
ISSN 2387-4201
Free

A spatial consistency test for the quality control of meteorological observations

Part II: Idealized experiments

Cristian Lussana and Line Båserud

The Norwegian Meteorological Institute, Oslo, Norway



Norwegian Meteorological Institute

METreport

Title	Date
A spatial consistency test for the quality control of meteorological observations. Part II: Idealized experiments	October 27, 2021
Section	**Report no.**
Division for Climate Services	12/2021
Author(s)	**Classification**
Cristian Lussana and Line Båserud	● Free ○ Restricted

Abstract

This report is the second of three reports describing the work that has been done at the Norwegian Meteorological institute (MET Norway) to develop and test spatial quality control methods for temperature and precipitation. The first report describes the methodology applied for spatial quality control. This report describes a set of idealized experiments that have been made on temperature and precipitation data. The third report describes the application of spatial quality control procedures to data collected by the observational network available at the Norwegian meteorological institute (MET Norway)

Keywords

data quality control, in-situ observations, near surface temperature, precipitation, spatial statistics

Disciplinary signature	Responsible signature
Hans Olav Hygen	Cecilie Stenersen

Abstract

This report is the second of three reports describing the work that has been done at the Norwegian Meteorological institute (MET Norway) to develop and test spatial quality control methods for temperature and precipitation. The first report describes the methodology applied for spatial quality control. This report describes a set of idealized experiments that have been made on temperature and precipitation data. The third report describes the application of spatial quality control procedures to data collected by the observational network available at the Norwegian meteorological institute (MET Norway).

Norwegian Meteorological Institute
Org.no 971274042
post@met.no
www.met.no / www.yr.no

Oslo
P.O. Box 43, Blindern
0313 Oslo, Norway
T. +47 22 96 30 00

Bergen
Allégaten 70
5007 Bergen, Norway
T. +47 55 23 66 00

Tromsø
P.O. Box 6314, Langnes
9293 Tromsø, Norway
T. +47 77 62 13 00

Contents

1 Introduction **5**

2 Idealized experiments on temperature **6**

2.1 Production of synthetic data in a mountainous region 6

2.2 Results 6

3 Idealized experiments on precipitation **15**

3.1 Stochastic generator for spatially consistent data 15

3.2 Results 16

4 Conclusions **23**

A Categorical statistics used for evaluation **23**

B Supplementary material **25**

1 Introduction

Idealized experiments are used to study the SCT (*Lussana and Båserud*, 2021) behaviour in a totally controlled environment. The atmospheric conditions simulated are simplified versions of complex real-world situations, where just a few parameters are allowed to vary. Besides, the simulations must have a known statistics for the occurrence of GEs, such that our spatial checks can be tested. In practice, we have implemented stochastic weather generators that simulate temperature and precipitation fields for predefined weather situations. The fields are spatially continuous and the values are physically consistent. After the simulation of a spatial field, we need to perturb the simulated data to account for the occurrence of GEs. Because the data are generated through stochastic procedures, the sample size can be as large as needed to achieve statistically robust results.

The statistics of GE occurrence is based on a uniform distribution bounded by lower and upper limits, typically some climatological values. In fact, outside the range of climatological values it is easy to identify GEs without having to compare values against other observations. In practice, suppose there are: i) p observation sites where the variable has been simulated and ii) a prescribed prior probability of GE occurrence P(GE) (0–1). Then, a sample of size $p \cdot \mathrm{P(GE)}$ of random values is drawn from a uniform distribution. The values are then used to replace the simulated values at locations that are randomly chosen.

The idealized experiments on temperature are described in Sec. 2. The simulations focus on the correct identification of small-scale processes (e.g. temperature inversions) occurring in a single valley and not across a wider region. The goal is to recognize that the observations may incorporate significant REs but they are not to be mistaken as GEs.

The idealized experiments on precipitation are described in Sec. 3. In this case, precipitation fields characterized by structures having different length scales are simulated. From the smallest scales, typical of convective precipitation, to the broad systems that usually are associated with stratiform precipitation. The goal is to assess the performances of SCT for the different weather regimes.

2 Idealized experiments on temperature

2.1 Production of synthetic data in a mountainous region

We have considered a region surrounding the Folldal municipality in Innlandet county, Norway. The region is shown in the left panel of Fig. 1, it is a mountainous area with a climate characterized by short, cool summers and long, cold winters. The experiments presented are based on synthetic data for temperature, however we have used the actual spatial distribution of the available observational network (*Nipen et al.*, 2020). In total, there are 395 observation locations (white dots and squares) in the region shown in Fig. 1. The observation locations near Folldal (white squares) are 17. We have often noticed that the observations near Folldal show a different temperature pattern, if compared to the surrounding valleys, and this makes this valley a perfect candidate for our ideal experiments. The elevations at the observation locations can be seen on the y-axis of the right panel of Fig. 1, the range is from 200 m to 1450 m. The elevations of the Folldal observation locations (white squares) range from approximately 600 m to 900 m.

The idea at the basis of our experiments is that the regional near-surface temperature profile decreases with the elevation, having a prescribed lapse rate of $-0.0065°\mathrm{C}/m$. The gray points in the right panel of Fig. 1 show the near-surface temperature at the observation locations marked as dots in the left panel of the same figure. In the region surrounding Folldal, the temperature profile simulated is different. We have considered 41 profiles, nine of them are shown in the right panel of Fig. 1. Half of the cases simulate cold pools (i.e., temperature in Folldal is lower than that observed in the nearby valleys) and at the same time a temperature inversion occurs (i.e. temperature increases with elevation). The difference between those cases is in the strength of the simulated temperature inversion, which is determined by the parameter a (*Frei*, 2014). Positive a values generate temperature inversions, while negative values generate profiles where the valley floor is warmer than those of the surrounding valleys The strongest vertical gradients simulated are shown with dark blue and red points. The light blue and pink points display weaker gradients.

2.2 Results

The simulations have been initialized with the parameter values reported in Tab. 1 (see the definitions in *Lussana and Båserud* (2021)). In addition, the background values in the outer circle are obtained using Theil-Sen linear regression (*Wilks*, 2019), which is also

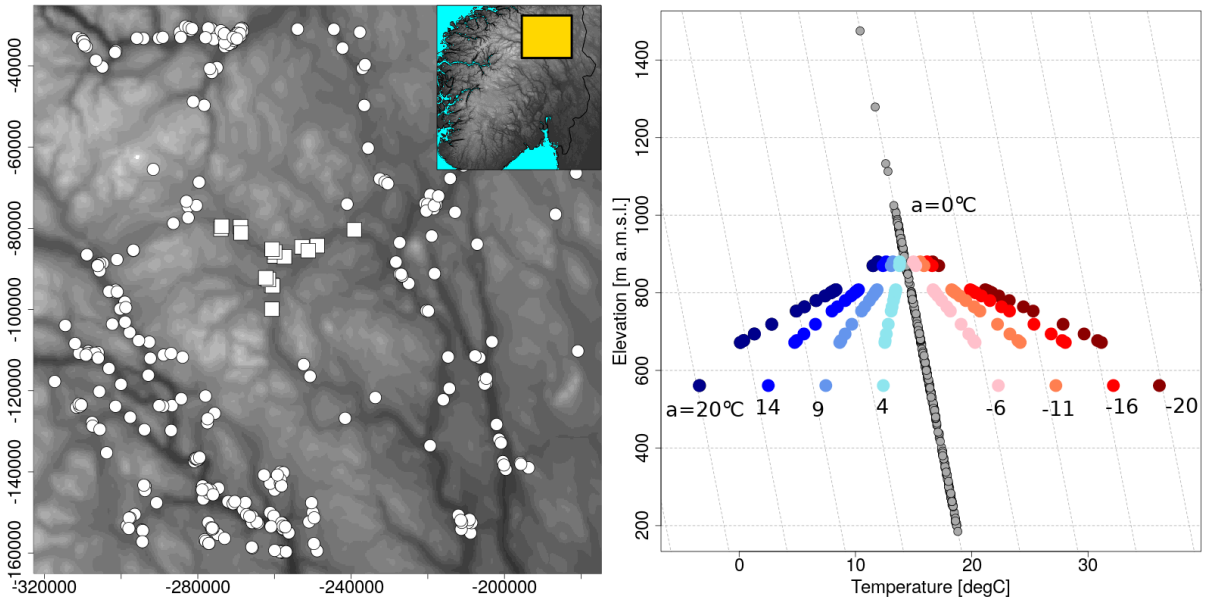


Figure 1: Region around Folldal in Norway and synthetic temperature data. The left panel shows the spatial distribution of observation locations in the region as white dots and squares. The topography is shown in shades of gray. The top-right inset shows South Norway and the region around Folldal is the one in the yellow box. The 378 white dots are the locations where the vertical profile of near-surface temperature follows a constant linear profile, which corresponds to the one shown by the gray dots in the right panel. The 17 white squares are the locations in the Folldal valley where the temperature profile differs from the regional profile. The blue and red dots show 9 of the 41 simulated local profiles. The temperature inversion strength a (*Frei*, 2014) is indicated. In the left panel, the units of the easting and northing coordinates are in metres.

called "median of pairwise slopes regression" (*Lanzante*, 1996). The SCT thresholds used do not distinguish between positive and negative cv-increments, we have used a unique value T (i.e. $T = T_+ = T_-$).

parameter	value
ε^2	0.5
r_{in}	20 km
r_{out}	40 km
k	3
$p_{\text{out,mn}}$	10
$p_{\text{out,mx}}$	50
D_{z}	400 m
D_{mn}	$r_{\text{in}}/10$
D_{mx}	r_{in}
$\boldsymbol{v}^{\text{mn}}$	$\boldsymbol{y}^{\text{o}} - 1°\text{C}$
$\boldsymbol{v}^{\text{mx}}$	$\boldsymbol{y}^{\text{o}} + 1°\text{C}$
$\boldsymbol{a}^{\text{mn}}$	$\boldsymbol{y}^{\text{o}} - 20°\text{C}$
$\boldsymbol{a}^{\text{mx}}$	$\boldsymbol{y}^{\text{o}} + 20°\text{C}$
$\boldsymbol{y}^{\text{b}}$	Theil-Sen linear regression

Table 1: SCT parameters used for temperature.

The SCT performances when all observations are set to good are shown in Fig. 2 for different values of a. In all the graphs, we can distinguish three different regimes. First, if $T < 1.5$, POFD reaches its highest values, between 0.3–0.75, depending on a. Examples of two simulations that are part of the dataset used are shown in Figs. 12– 13, for $T = 0.5$ and $T = 1.5$, respectively. Second, if $1.5 \leq T < 2.5$, POFD is equal to approximately 0.05 for all panels. Finally, when $T \geq 3$, POFD=0 for all values of a. The results support the use of T higher than 1.5 (when $\text{P(GE)} = 0\%$), otherwise the risk of having a good observation mistaken as a bad one (false alarm) is quite high.

The POD, POFD and ETS are shown in Figs. 3– 4- 5, respectively, for different values of P(GE). In all graphs, there are two lines. The gray lines show the aggregated skill-score when all observations are considered. The pink lines show the results for Folldal. When all observations are considered, the results show better performances because of the more spatially coherent temperature field and the higher -on average- spatial density

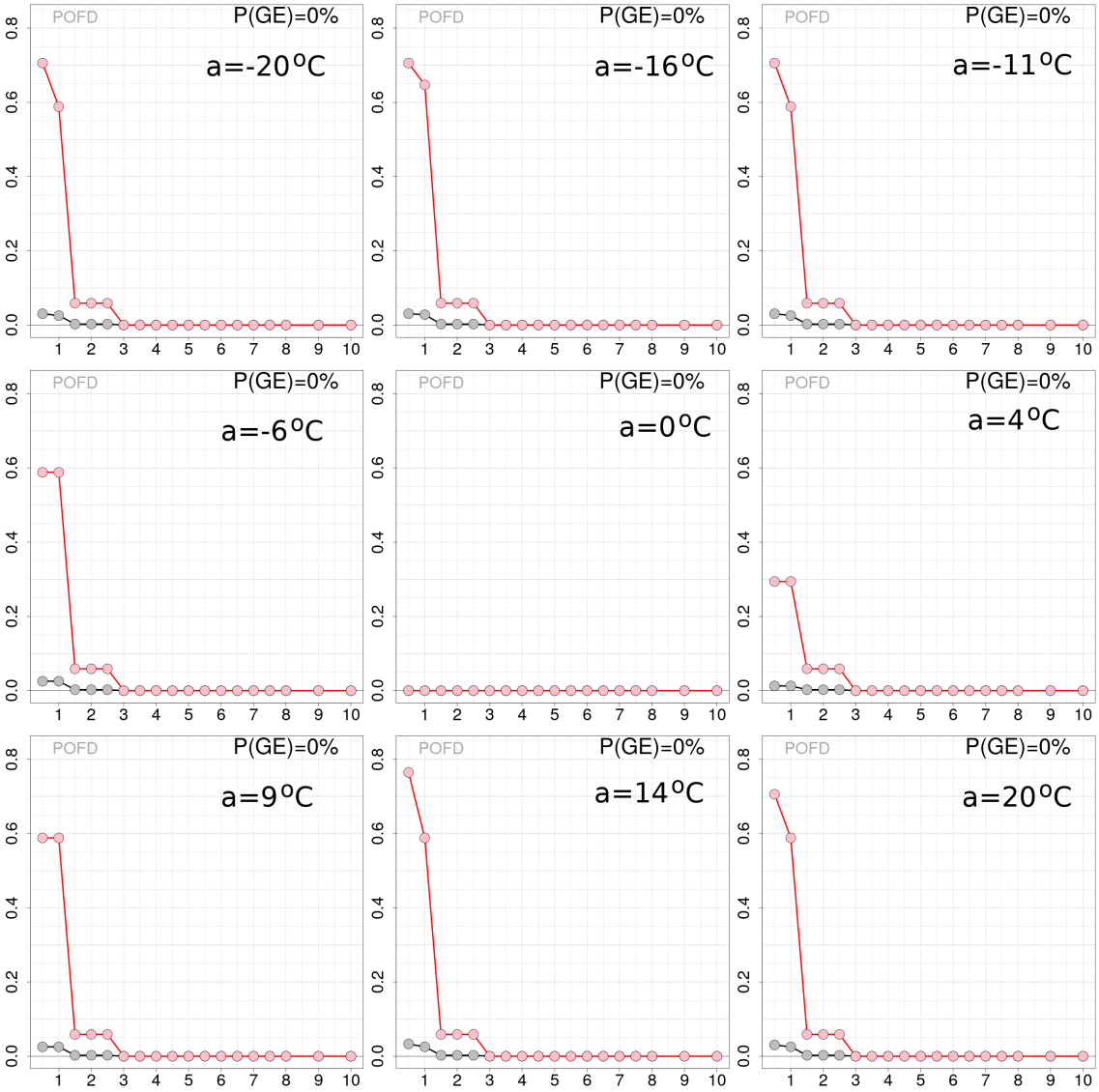


Figure 2: Experiments on synthetic data of temperature generated with $\mathrm{P(GE)} = 0\%$ (see Fig. 1). Probability of false detection (POFD) as a function of the SCT threshold for different values of the inversion strength a, as indicated on the panels. The dataset is based on 500 simulations. The pink lines show the results when considering observations in Folldal only (i.e. white squares in Fig. 1). The gray lines show the results when considering all observations. See also Figs. 12- 13

.

of observations considered. This is especially true for the POFD in Fig. 4. However, in Fig. 3, the PODs for P(GE) $= 40\%$ and P(GE) $= 50\%$ are not that different when comparing Folldal to the whole region. As the value of P(GE) increases, the two curves become more and more similar, even though the gray curve is always displaying the best performance. When P(GE) is higher or equal to 30%, the spatial coherence of the samples is significantly reduced and disturbed by the presence of GEs. For high values of P(GE), the range of optimal values of T is drastically reduced, regardless of the observational density. On the other hand, when P(GE) is small, it is the synergy between a dense network and spatially coherent samples that determines stable performances over a wide range of T values. If compared to POD and POFD, the ETS in Fig. 4 constitutes a more balanced assessment of the SCT performances, in the sense that all components of the contingency table are taken into account. Given a P(GE) value, the best SCT threshold is the one that guarantees the highest ETS. The best thresholds are shown in Fig. 6. In the supplementary material, Figs. 14– 22 show results from individual simulations used to obtain Fig. 5. In the figures, the best thresholds T are used.

The best values of T shown in Fig. 6 are between 0.5 and 4: lower values when P(GE) is small, higher values when P(GE) increases. The magnitude of the inversion strength a does not influence significantly the form of this relationship, instead it has an influence on the SCT performance. For instance, the best values of the ETS for $a = 4°$C and P(GE) $= 1\%$ are between 0.7–0.8, while for $a = 14°$C and P(GE) $= 1\%$ the ETS is at most equal to 0.4.

In the supplementary material, the same figures that are shown for $a = 20$ are shown for the other values of a reported in Fig 1. The POD is shown in Figs. 23– 30. The POFD is shown in Figs. 31– 38. The ETS is shown in Figs. 39– 46.

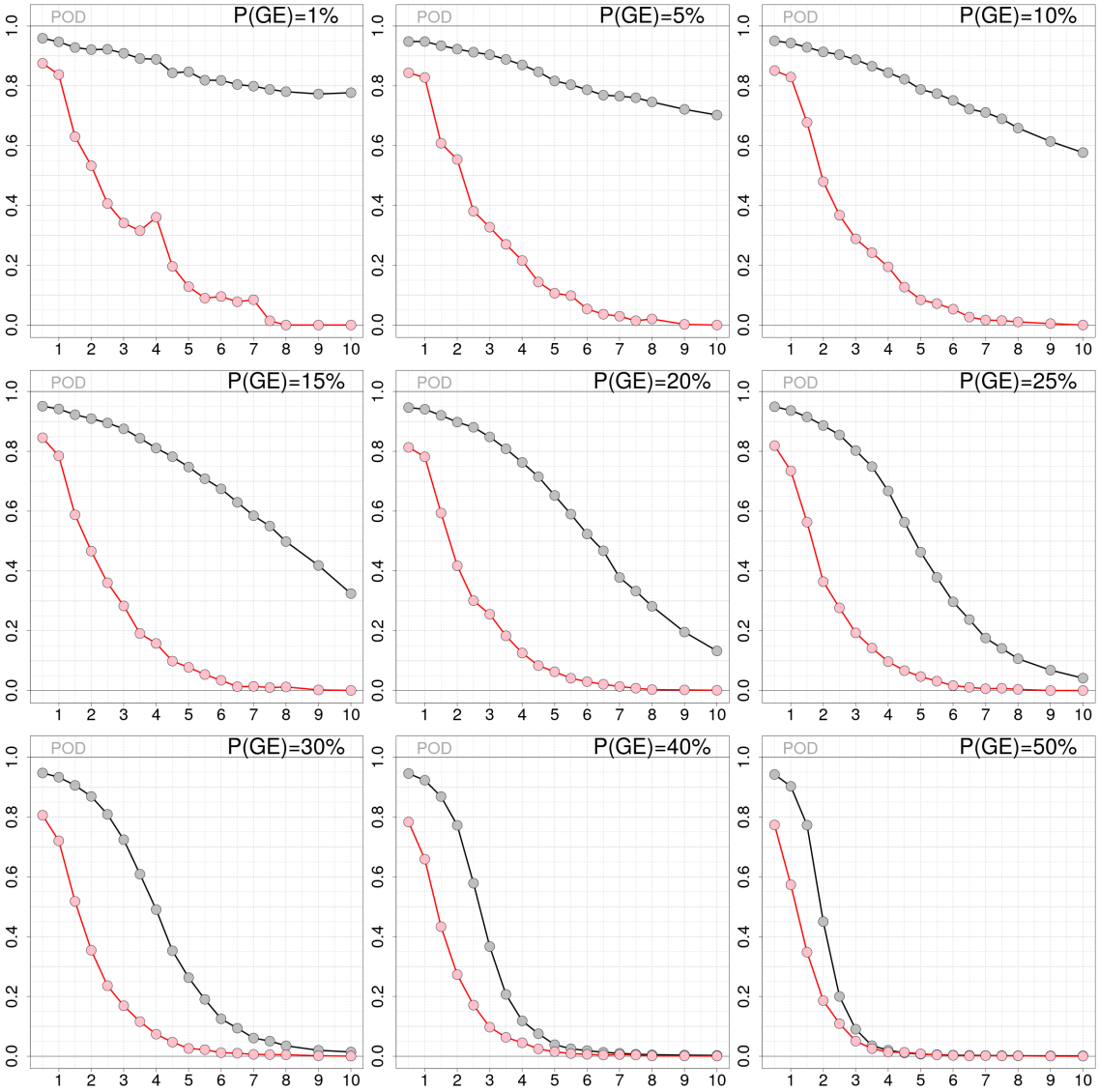


Figure 3: Experiments on synthetic data of temperature generated with $a = 20°C$ (see Fig. 1). Probability of detection (POD) as a function of the SCT threshold (x-axis) for different values of the prior probability of GE occurrence (P(GE)), as indicated on the panels. The dataset is based on 500 simulations. The pink lines show the results when considering observations in Folldal only (i.e. white squares in Fig. 1). The gray lines show the results when considering all observations.

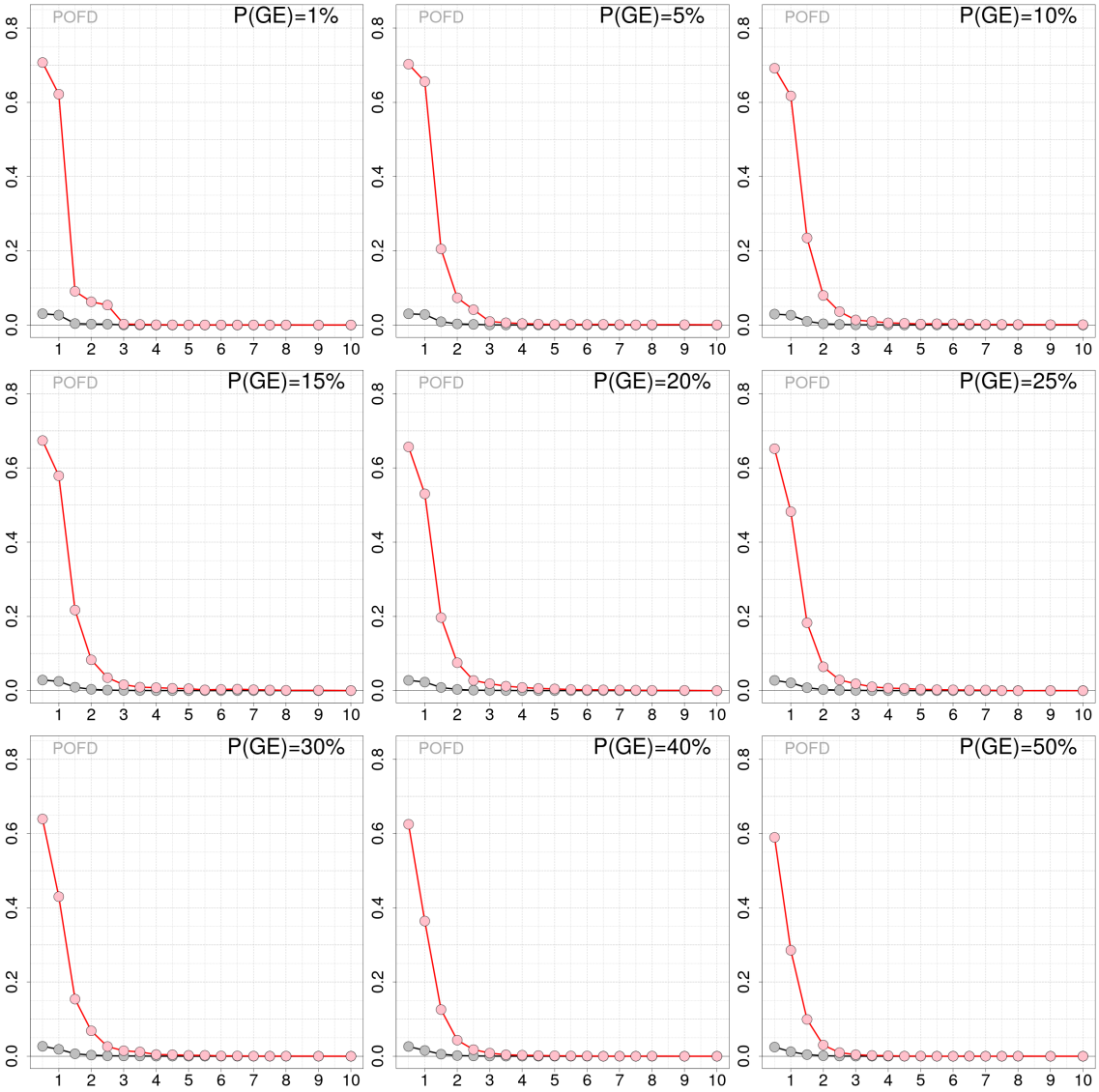


Figure 4: Experiments on synthetic data of temperature generated with $a = 20°C$ (see Fig. 1). Probability of false detection (POFD) as a function of the SCT threshold for different values of the prior probability of GE occurrence (P(GE)), as indicated on the panels. The dataset is based on 500 simulations. The pink lines show the results when considering observations in Folldal only (i.e. white squares in Fig. 1). The gray lines show the results when considering all observations.

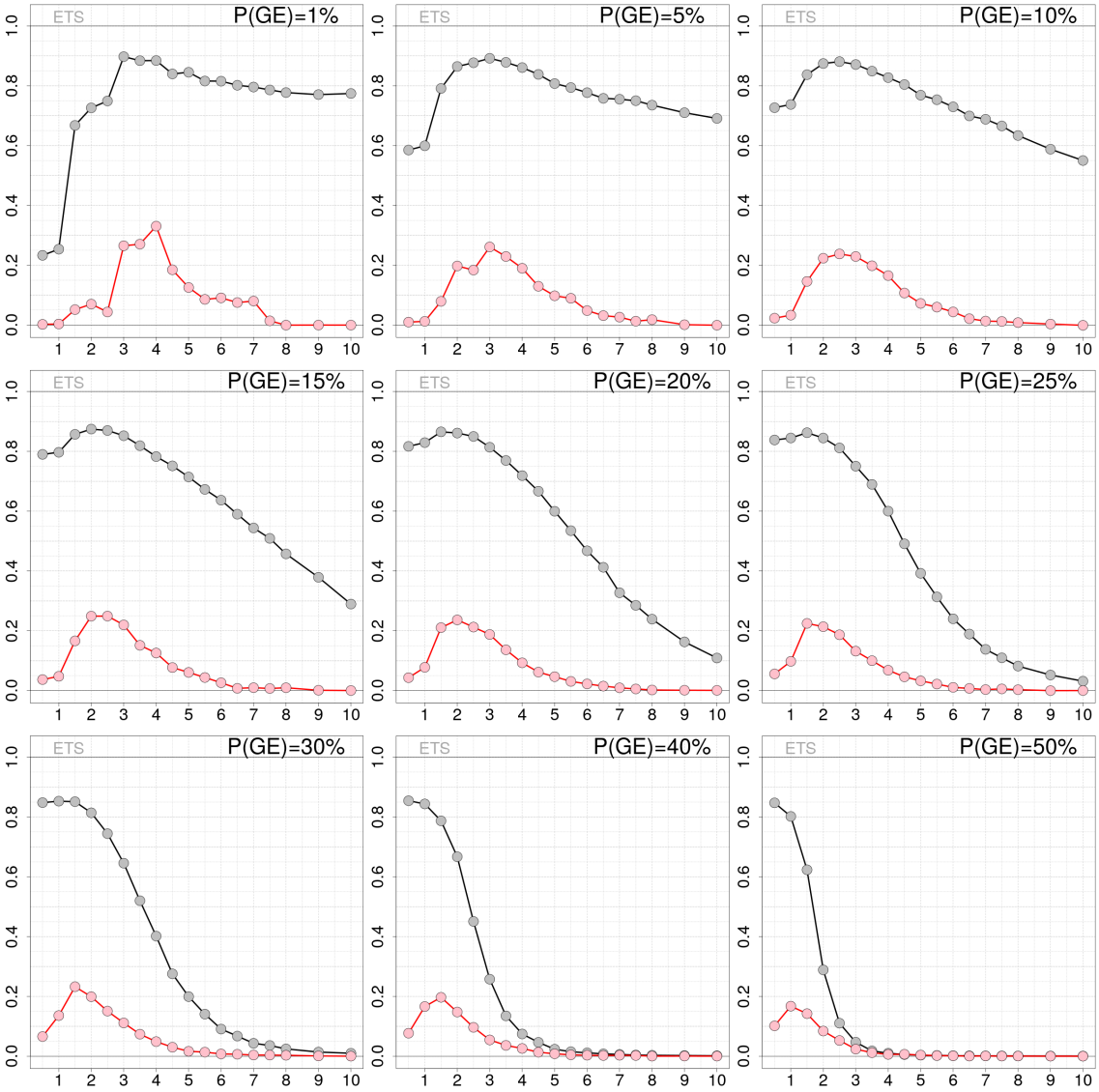


Figure 5: Experiments on synthetic data of temperature generated with $a = 20°C$ (see Fig. 1). Equitable threat score (ETS) as a function of the SCT threshold for different values of the prior probability of GE occurrence (P(GE)), as indicated on the panels. The dataset is based on 500 simulations. The pink lines show the results when considering observations in Folldal only (i.e. white squares in Fig. 1). The gray lines show the results when considering all observations.

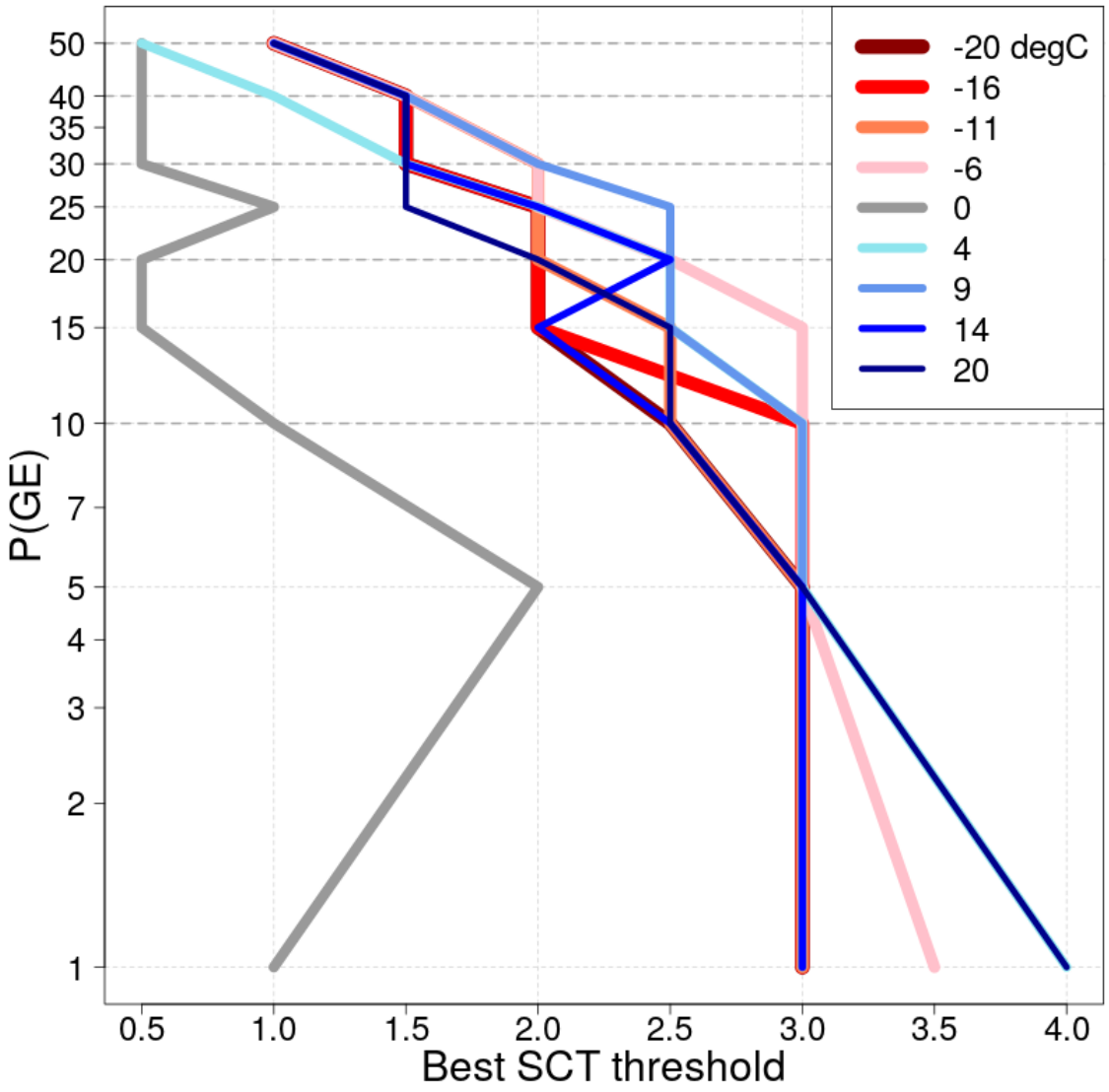


Figure 6: Experiments on synthetic data of temperature (see Fig. 1). The relationship between the best SCT threshold T and the prior probability of gross error P(GE) is shown for different values of a (see Fig. 1). Given P(GE), the best SCT threshold is the one maximizing the ETS. The y-axis has a logarithmic scale.

3 Idealized experiments on precipitation

Precipitation is highly variable over short distances. Distribution of values is positively skewed. The observational network is sometimes dense enough to represent this variability but often it is too sparse. Precipitation errors (uncertainties) are heteroskedastic, that is the variability is different across the range of values. In particular, the variability is directly proportional to the value. Often, the observation error standard deviation is considered to be approximately from 10% to 20% of the observed values, depending on the atmospheric conditions (*Wolff et al.*, 2015).

3.1 Stochastic generator for spatially consistent data

The objective is to generate as many random realizations of a precipitation field as one may wish. The implementation consists of two steps. First, random realizations of a 2D Gaussian random field are generated. Second, a Gaussian anamorphism is used to transform the normally distributed values into a Gamma distribution, which better represents the skewed distribution of precipitation values. A similar data transformation has been used by *Lussana et al.* (2021).

The larger the number of grid points, the more computationally expensive the simulation is. We opted for a grid covering South Norway, with 30 points in the easting dimension and 40 points in the northing dimension. The grid spacing is 12.5 km in both dimensions.

The stochastic generator samples a conditional Gaussian random field with prescribed mean and covariance matrix. The samples from the multivariate normal distribution, with a prescribed continuous spatial structure, are obtained as described by *Wilks* (2019) in chap. 12.4. The mean (30 x 40 elements) vector is set to 0 everywhere and the covariance matrix (1200 rows x 1200 columns) is defined by means of a second-order autoregressive function (*Gaspari and Cohn*, 1999) of the form:

$$\mathrm{C}(\boldsymbol{r}) = \left(1 + \frac{|\boldsymbol{r}|}{L}\right) \exp\left(-\frac{|\boldsymbol{r}|}{L}\right) \tag{1}$$

where $|\boldsymbol{r}|$ represents distance between two grid points in R^3 and L is the fixed length scale. Similar formulations of the error covariance matrix for spatial analysis of precipitation have been used in the past (*Mahfouf et al.*, 2007; *Soci et al.*, 2016). The spatial length scale L used may vary, such that precipitation fields with different characteristics can be simulated.

We have considered the following values of L (km): 10, 50, 100, 200, 400. For each of the L values, 100 random realizations are generated.

The Gaussian random fields are adjusted such that the final distribution of values is a Gamma distribution, with shape and rate parameters derived from the typical weather of the region. We have used daily precipitation fields derived from numerical weather predictions considering the time period from June 2019 to June 2020. The values of shape and rate are the maximum likelihood estimators calculated iteratively by means of the Newton–Raphson method as described by *Wilks* (2019), Sect. 4.6.2. The Gamma shape is set to 0.1328493, while the rate is set 0.0285855. Examples of the resulting fields are shown in Fig. 7. The smaller spatial length scale of 10 km is used to simulate convective situations, while the largest spatial scales resemble cases of stratiform precipitation over large parts of the region. We are interested in precipitation values at station locations. In Fig. 7, panel *a* shows the spatial distribution of the MET Norway observational network (July 2020). The precipitation value assigned to each station location is extracted from the precipitation fields by means of a nearest neighbour interpolation.

The other important points to know are:

- A Box-Cox transformation with transformation parameter set to 0.3 has been applied before running the SCT
- The first guess in the outer circle is the mean of all values
- As for temperature, also for precipitation we have included a known fraction of GEs to study the behaviour of the SCT

3.2 Results

The parameters used for the SCT are reported in Tab. 2.

The POD, POFD and ETS for $L = 50$ km are shown in Figs. 8– 10. The POD, shown in Fig. 8, is close to 1 for $T = 1$, then it decays exponentially towards 0. The decay rate is faster for higher values of P(GE). Note that for P(GE) less than 5%, the POD is quite high even for $T = 12$. The POFD, shown in Fig. 9, does not change significantly for P(GE) between 1% and 30%. In this range of P(GE) values, the POFD is always less than 5% and less than 1% when $T \geq 4$. The situation changes for P(GE) equal to 40% and 50%, with POFD that doubles its values. The ETS is shown in Fig. 10 and the best SCT thresholds (i.e. those that maximise the ETS) are shown in Fig. 11. For values of L in the

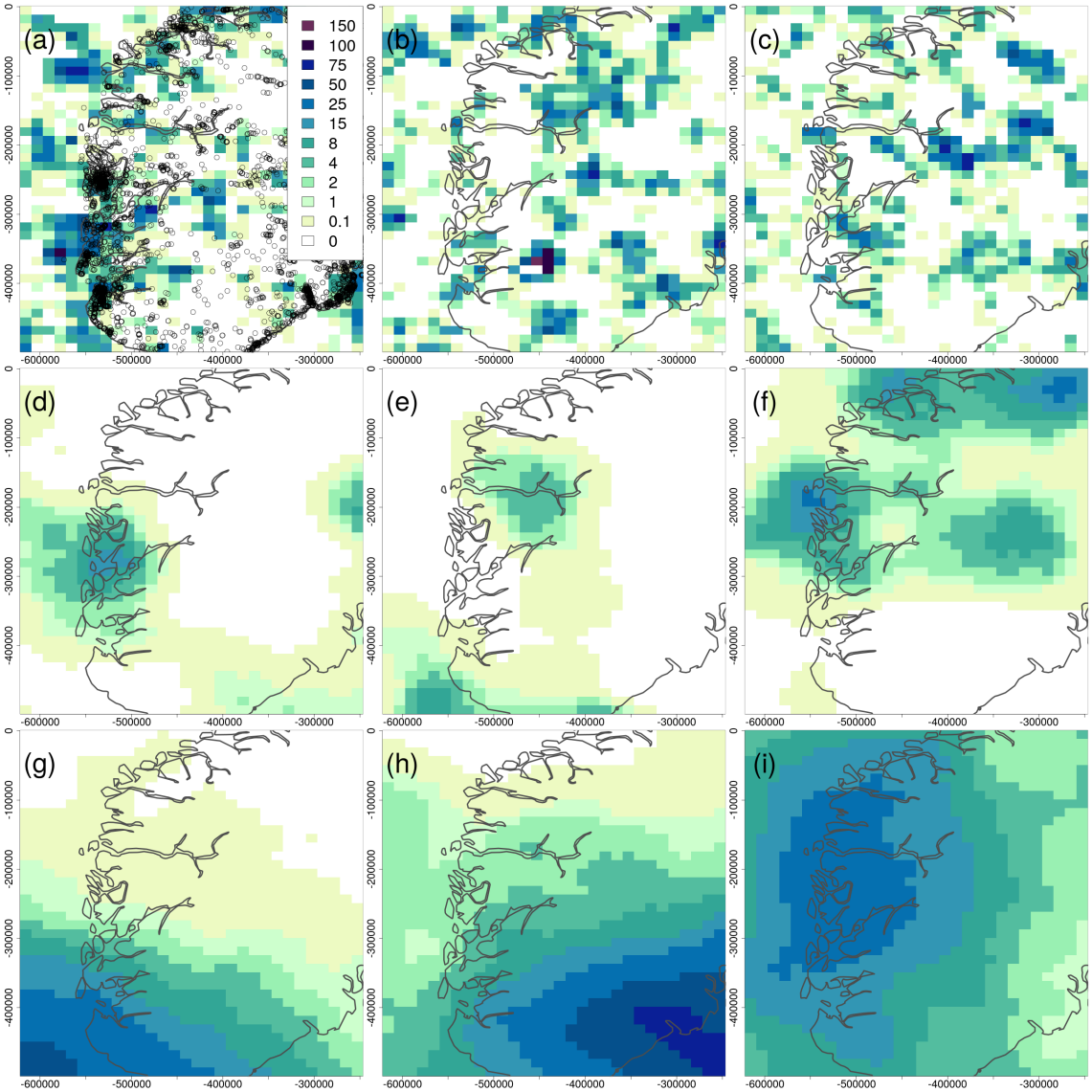


Figure 7: Precipitation synthetic gridded fields created with a stochastic weather generator. All values are in mm. The panels a, b and c in the top row are generated with a spatial length scale of $L = 10$ km (Eq.(1)). The panels d, e and f in the middle row have $L = 100$ km. The panels g, h and i in the bottom row have $L = 200$ km. In panel a, the actual observational network of in-situ weather stations available at MET Norway in July 2020 is shown (black dots). The Gaussian anamorphosis parameters are Gamma shape=0.1328493, rate=0.0285855. The units of easting and northing coordinates are metres.

parameter	value
ε^2	0.1
r_{in}	50 km
r_{out}	100 km
k	3
$p_{\text{out,mn}}$	10
$p_{\text{out,mx}}$	60
D_{z}	-
D_{mn}	$r_{\text{in}}/10$
D_{mx}	r_{in}
$\boldsymbol{v}^{\text{mn}}$	min($\boldsymbol{y}^{\text{o}} - 1\,\text{mm}$ OR $\boldsymbol{y}^{\text{o}} - 10\%$)
$\boldsymbol{v}^{\text{mx}}$	max($\boldsymbol{y}^{\text{o}} + 1\,\text{mm}$ OR $\boldsymbol{y}^{\text{o}} + 10\%$)
$\boldsymbol{a}^{\text{mn}}$	min($\boldsymbol{y}^{\text{o}} - 10\,\text{mm}$ OR $\boldsymbol{y}^{\text{o}} - 50\%$)
$\boldsymbol{a}^{\text{mx}}$	max($\boldsymbol{y}^{\text{o}} + 10\,\text{mm}$ OR $\boldsymbol{y}^{\text{o}} + 50\%$)
$\boldsymbol{y}^{\text{b}}$	mean outer circle

Table 2: SCT parameters used for precipitation.

range from 10 km to 200 km, the best T values are remarkably similar. In the case of L equal to 400 km, the best T value is systematically smaller than for the other values. In both figures, Fig. 6 and Fig. 11, the relationship between the best T and P(GE) could be approximated by the analytical function:

$$\log(\text{P(GE)}) = aT^2 + b^T + c \tag{2}$$

with negative values of the coefficient a. The range of values for T could depend on the interval used in the definition of the uniform distribution considered as the statistical model for the GE occurrence.

The figures reported in this section have been obtained with $L = 50$ km. For the other values of L considered, which are 10 km, 100 km, 200 km and 400 km, the corresponding figures are shown in the Supplementary material Figs. 48– 59. The SCT performance worsens when we consider the smaller scales.

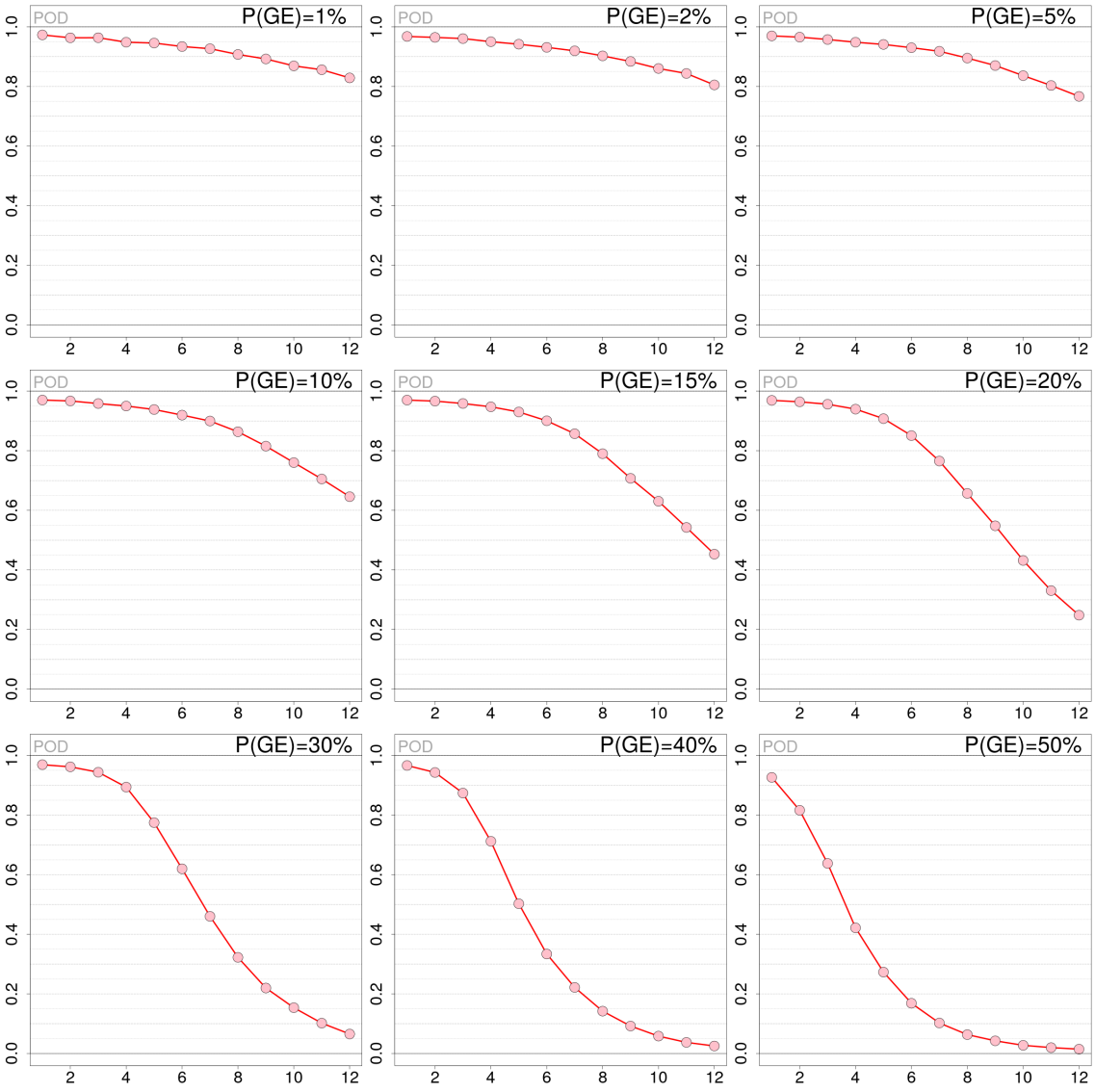


Figure 8: Experiments on synthetic precipitation data produced with $L = 50$ km. Probability of detection (POD) as a function of the SCT threshold T for different values of the predefined probability of gross error occurrence, which are indicated on the top right corner of each panel. The dataset analyzed is based on 100 simulated fields.

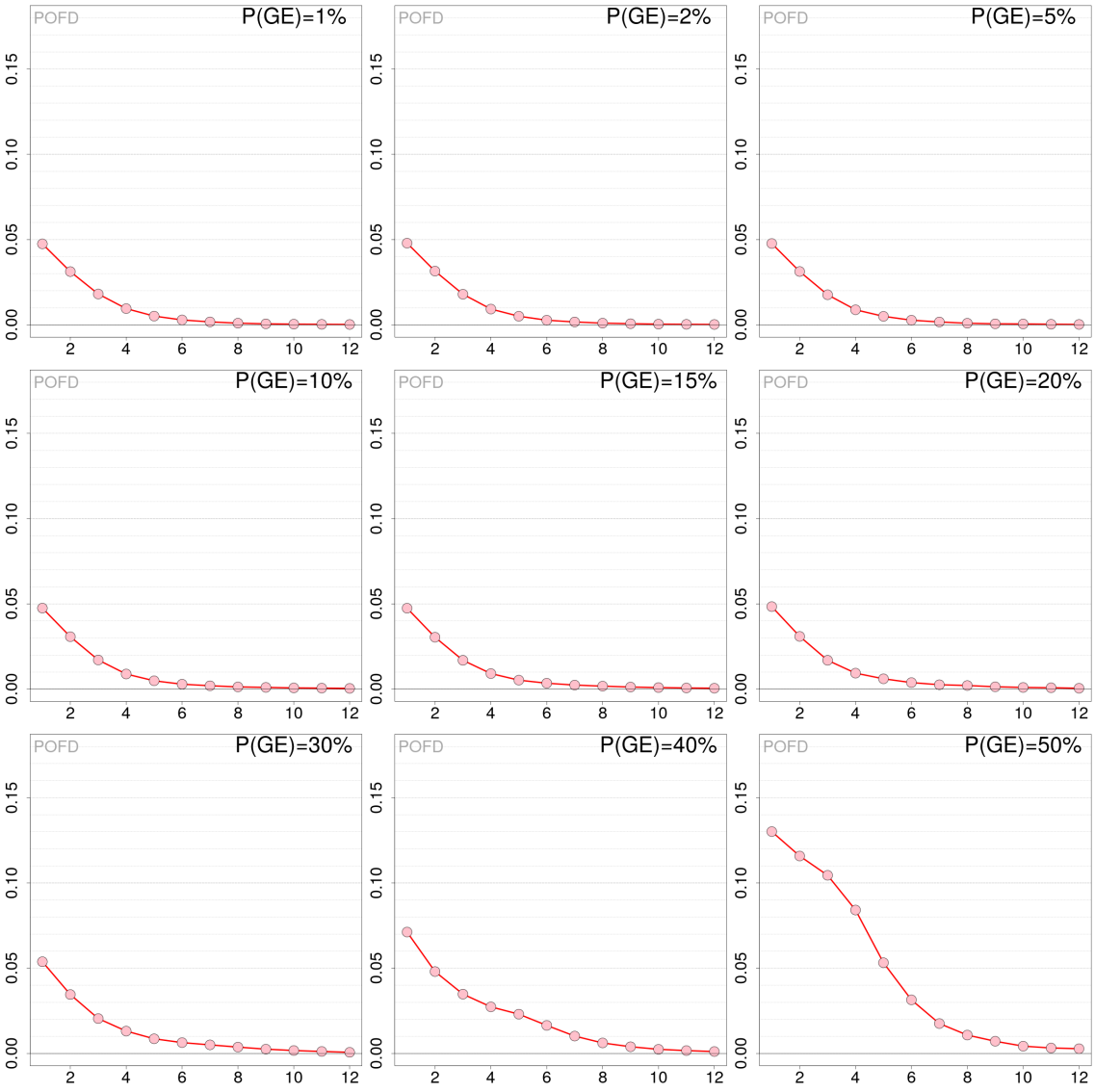


Figure 9: Experiments on synthetic precipitation data produced with $L = 50$ km. Probability of false detection (POFD) as a function of the SCT threshold T for different values of the predefined probability of gross error occurrence, which are indicated on the top right corner of each panel. The dataset analyzed is based on 100 simulated fields.

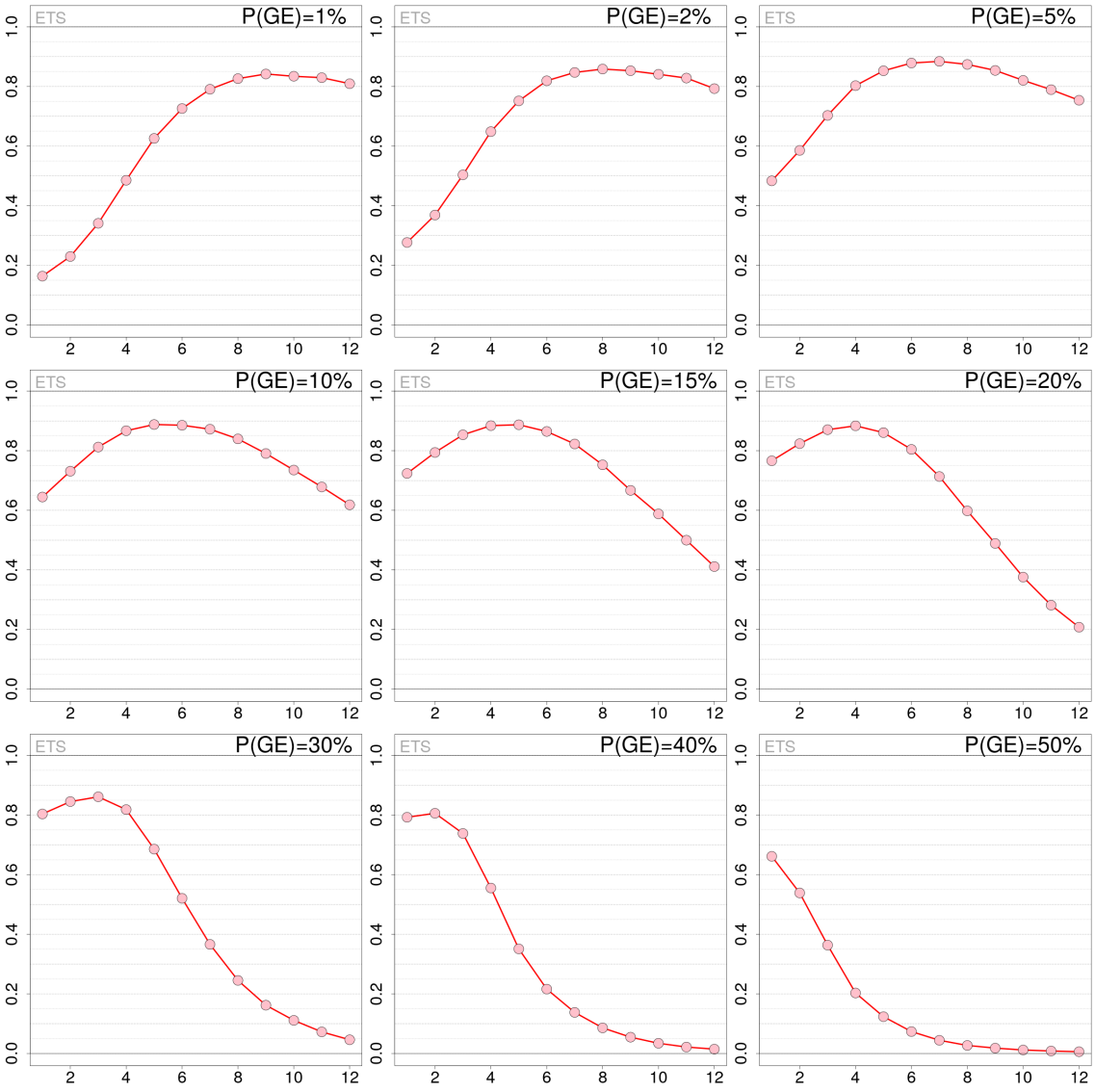


Figure 10: Experiments on synthetic precipitation data produced with $L = 50$ km. Equitable threat score (ETS) as a function of the SCT threshold T for different values of the predefined probability of gross error occurrence, which are indicated on the top right corner of each panel. The dataset analyzed is based on 100 simulated fields.

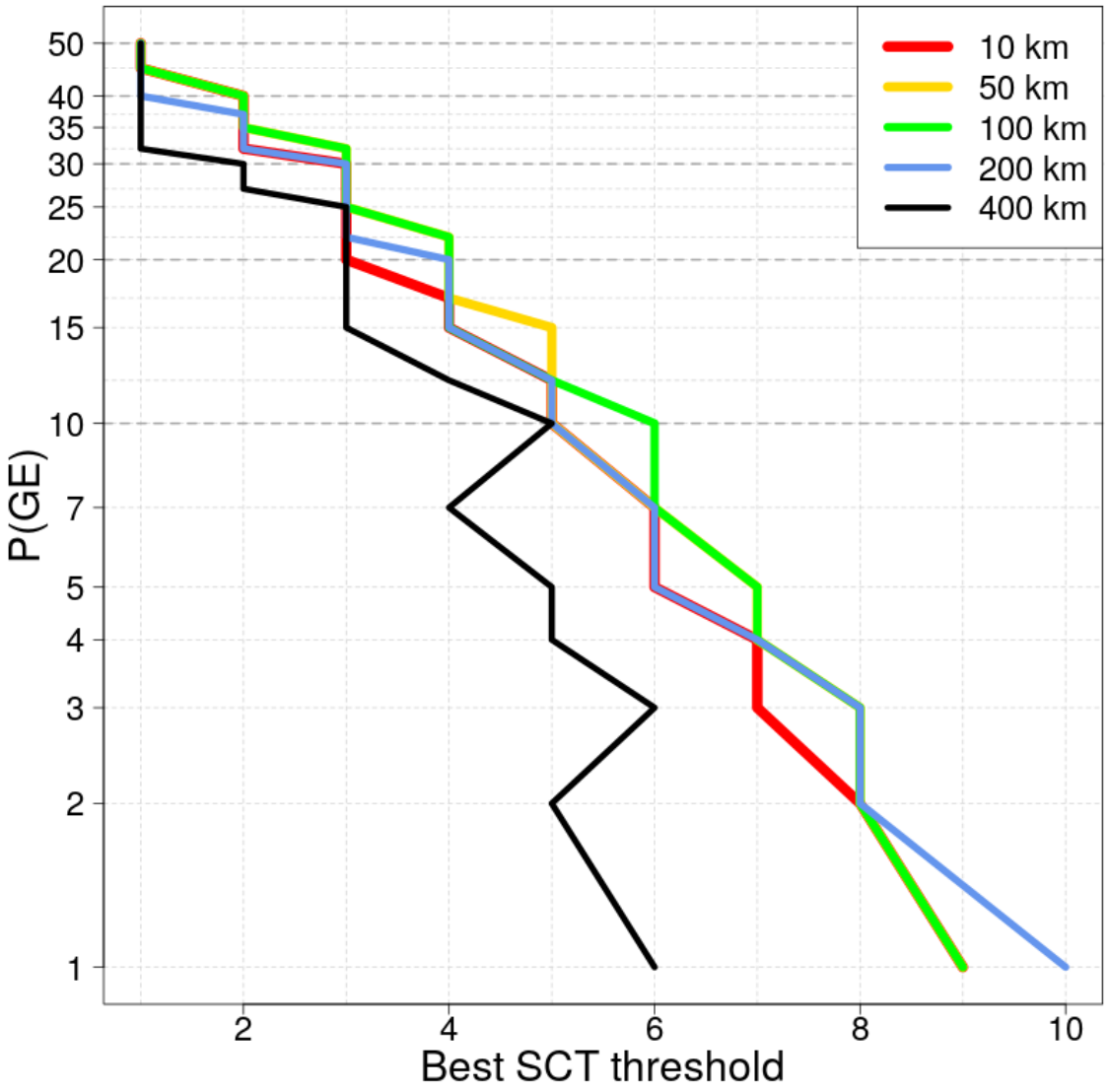


Figure 11: Experiments on synthetic precipitation data. Equitable threat score (ETS) as a function of the SCT threshold T for different values of the predefined probability of gross error occurrence, which are indicated on the top right corner of each panel. The dataset analyzed is based on 100 simulated fields.

4 Conclusions

The stochastic generator implemented for temperature and precipitation fields allowed us to test the SCT under well-defined conditions.

The main conclusions are valid both for temperature and precipitation. When P(GE) is high (above 20%), the range of values determining the optimal SCT thresholds is narrow and centered over small values (less than 2). On the other hand, when P(GE) is smaller (less than 20%), the range of optimal SCT threshold values is wider (between 2 and 4, or more).

The best value of the SCT threshold is mostly determined by P(GE) and not by the weather conditions. The occurrence of small-scale weather phenomena in a region worsen the SCT performances, even though for the observational network under study, useful results are still obtained even when 50% of the observations are affected by gross-measurements errors.

A Categorical statistics used for evaluation

A useful reference is the website `https://www.cawcr.gov.au/projects/verification/` and references reported there. In our study, the observed event is ”observation is bad" and the predicted event is ”SCT flags observation as bad". The four combinations of SCT flags and GEs occurrence are:

- hit, SCT flags a bad observation as bad
- miss, SCT flags a bad observation as good
- false alarm, SCT flags a good observation as bad
- correct negative, SCT flags a good observation as good

The contingency table used:
The statistics we use are reported in the following.
Probability of detection (POD):

$$\mathrm{POD} = \frac{\mathrm{hits}}{\mathrm{hits} + \mathrm{misses}} \tag{3}$$

Probability of false detection (POFD):

$$\mathrm{POFD} = \frac{\mathrm{false\ alarms}}{\mathrm{correct\ negatives} + \mathrm{false\ alarms}} \tag{4}$$

Event predicted	Event observed	
	observation is bad	observation is good
SCT flags observation as bad	hits	false alarms
SCT flags observation as good	misses	correct negatives

Table 3: Contingency table used to evaluate the SCT.

Equitable threat score (ETS):

$$\text{ETS} = \frac{\text{hits} - \text{hits}_{\text{random}}}{\text{hits} + \text{misses} + \text{false alarms} - \text{hits}_{\text{random}}} \tag{5}$$

where:

$$\text{hits}_{\text{random}} = \frac{(\text{hits} + \text{misses})(\text{hits} + \text{false alarms})}{\text{total}} \tag{6}$$

B Supplementary material

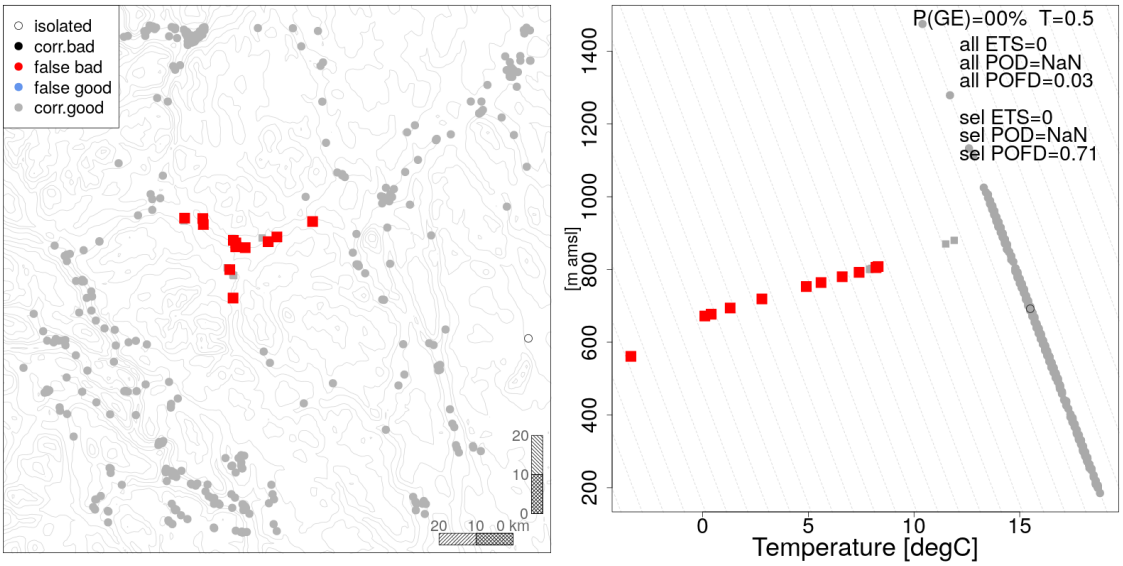


Figure 12: Experiments on synthetic data of temperature generated with $a = 20°C$. The figure refers to a simulation (randomly chosen) among the 500 generated. All observations are set to good (P(GE) $= 0\%$), the threshold is $T = 0.5$. The left panel shows the spatial distribution of the observational network, where the dots and squares mark the observation locations. The meanings associated with the colors are indicated in the legend. The right panel shows the vertical profile of temperature, the colors are the same as in the left panel.

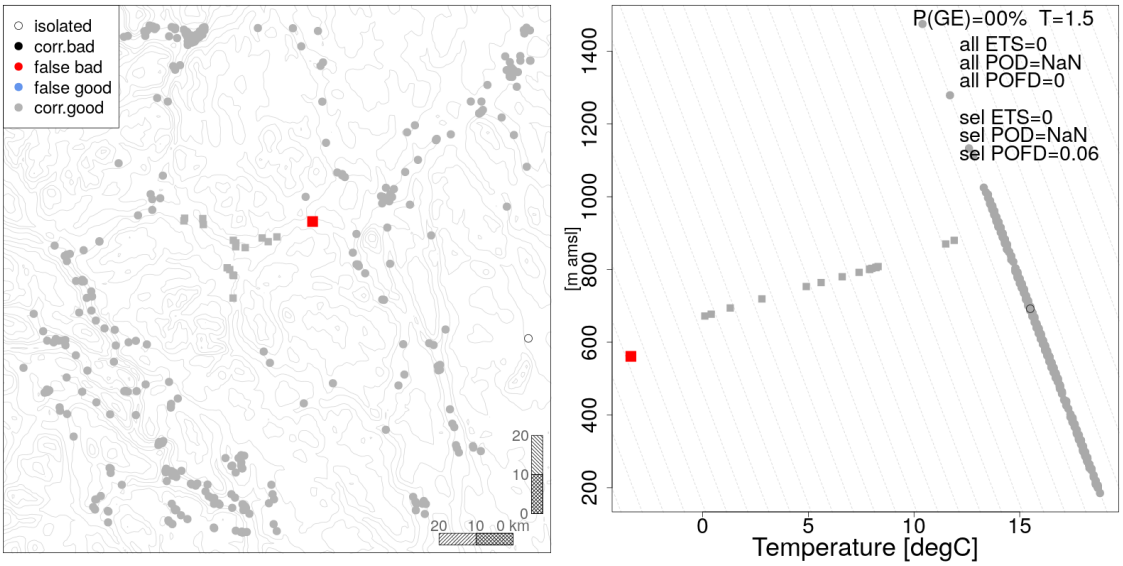


Figure 13: Same as Fig. 12, but with $T = 1.5$.

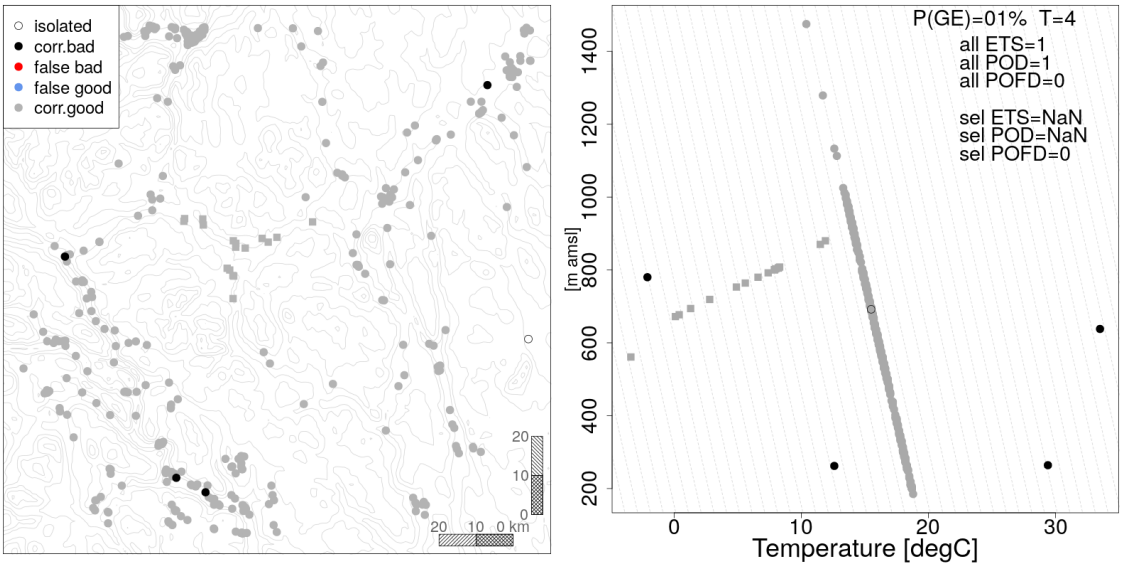


Figure 14: Experiments on synthetic data of temperature generated with $a = 20°\text{C}$, $\text{P(GE)} = 1\%$ and $T = 4$. The layout is similar to Fig. 12.

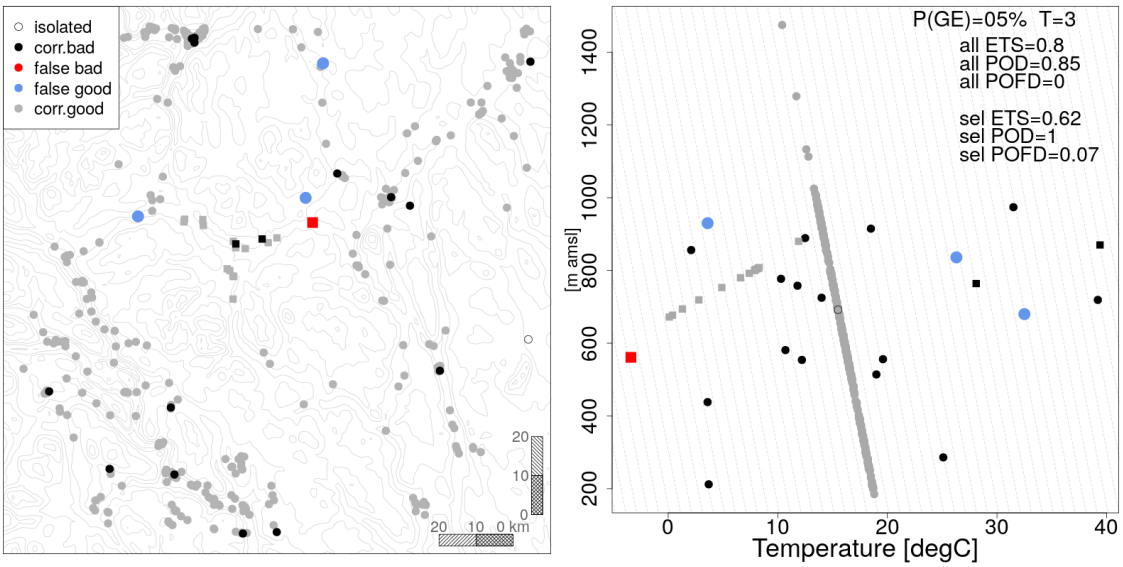


Figure 15: Experiments on synthetic data of temperature generated with $a = 20°\text{C}$, $\text{P(GE)} = 5\%$ and $T = 3$. The layout is similar to Fig. 12.

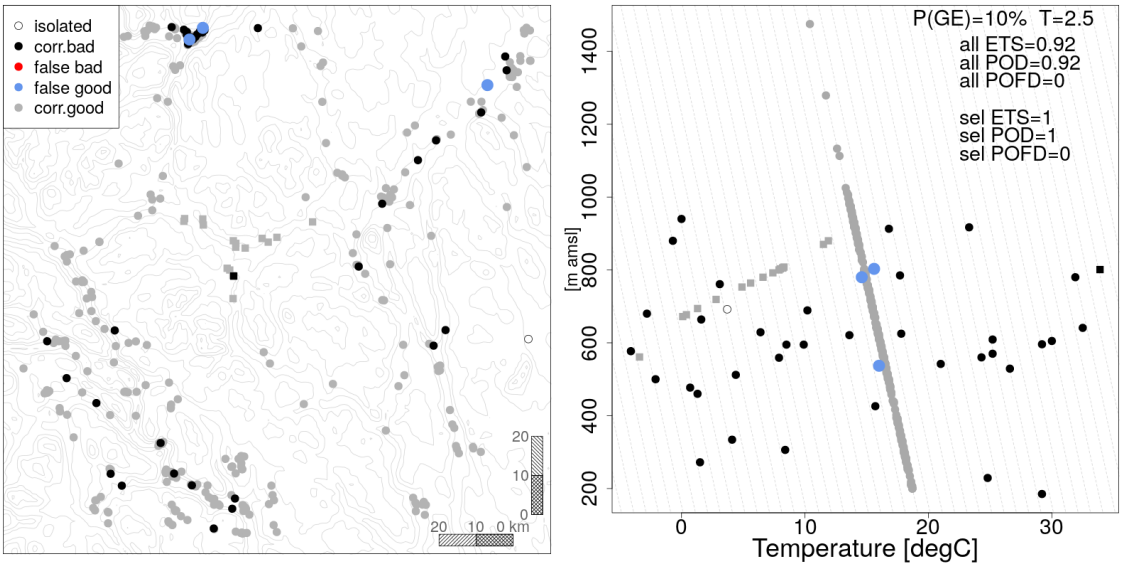


Figure 16: Experiments on synthetic data of temperature generated with $a = 20°\text{C}$, $\text{P(GE)} = 10\%$ and $T = 2.5$. The layout is similar to Fig. 12.

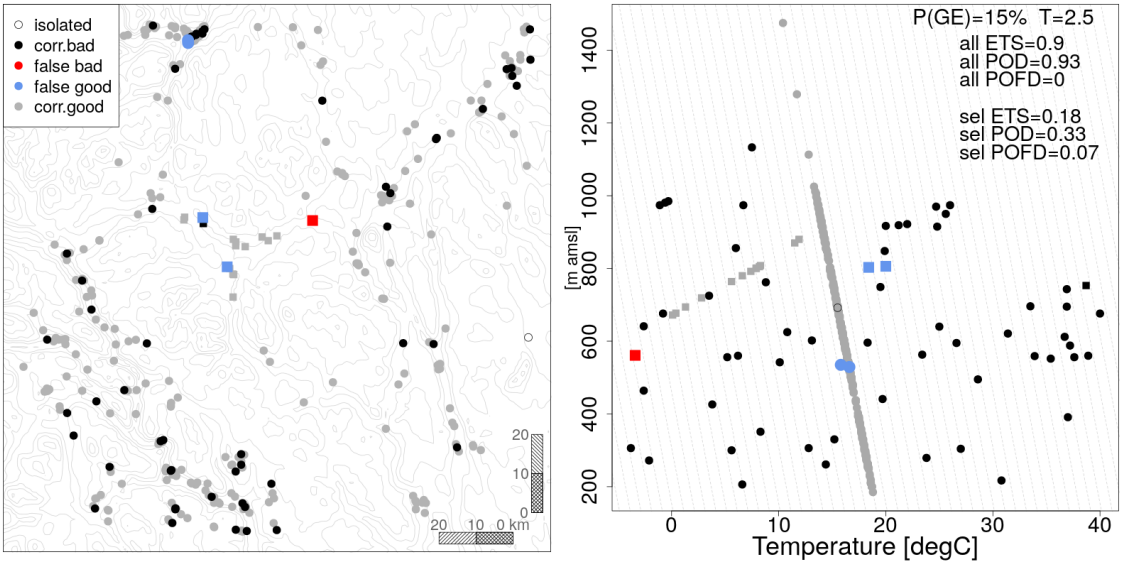


Figure 17: Experiments on synthetic data of temperature generated with $a = 20°\text{C}$, $\text{P(GE)} = 15\%$ and $T = 2.5$. The layout is similar to Fig. 12.

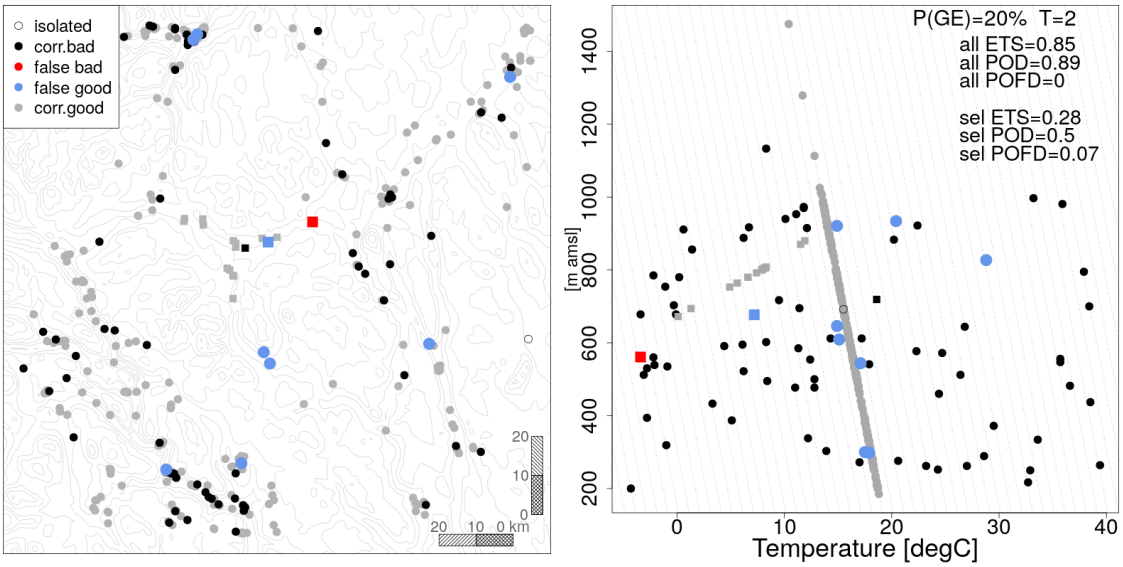


Figure 18: Experiments on synthetic data of temperature generated with $a = 20°\text{C}$, $\text{P(GE)} = 20\%$ and $T = 2$. The layout is similar to Fig. 12.

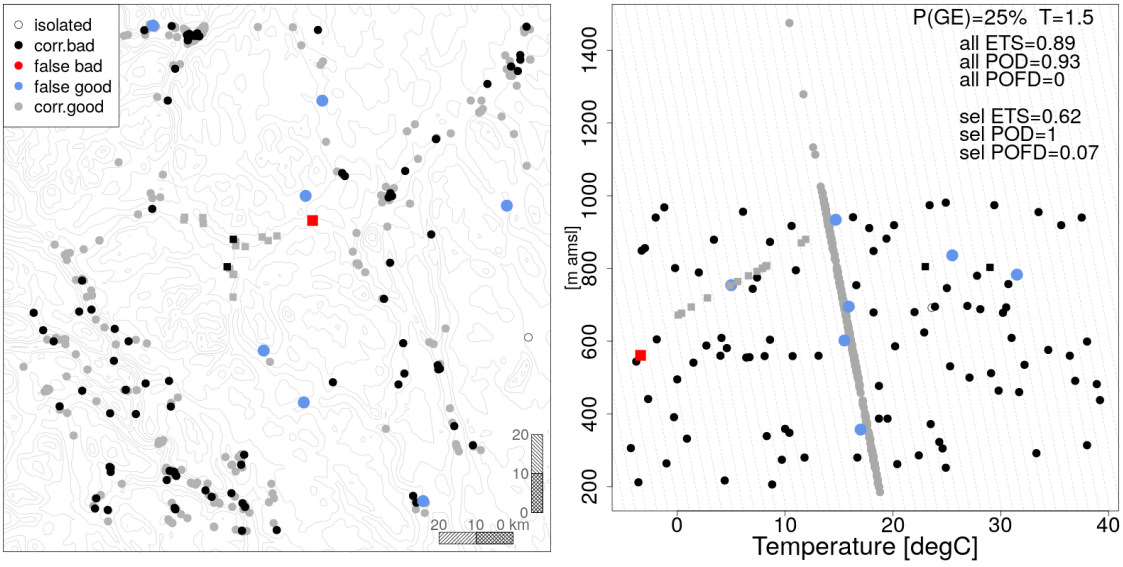


Figure 19: Experiments on synthetic data of temperature generated with $a = 20°\text{C}$, $\text{P(GE)} = 25\%$ and $T = 1.5$. The layout is similar to Fig. 12.

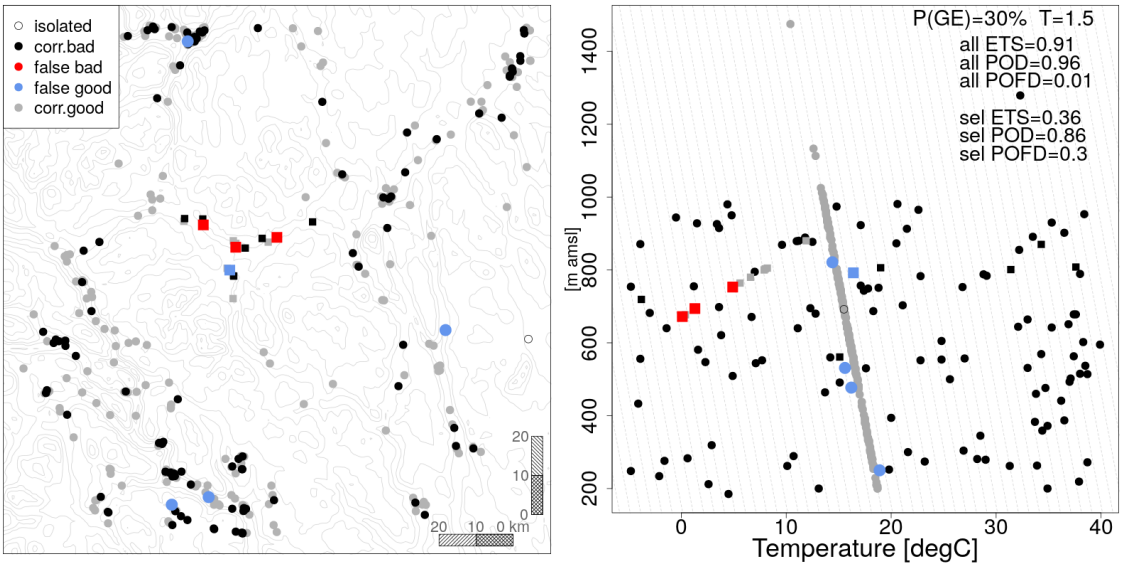


Figure 20: Experiments on synthetic data of temperature generated with $a = 20°\text{C}$, $\text{P(GE)} = 30\%$ and $T = 1.5$. The layout is similar to Fig. 12.

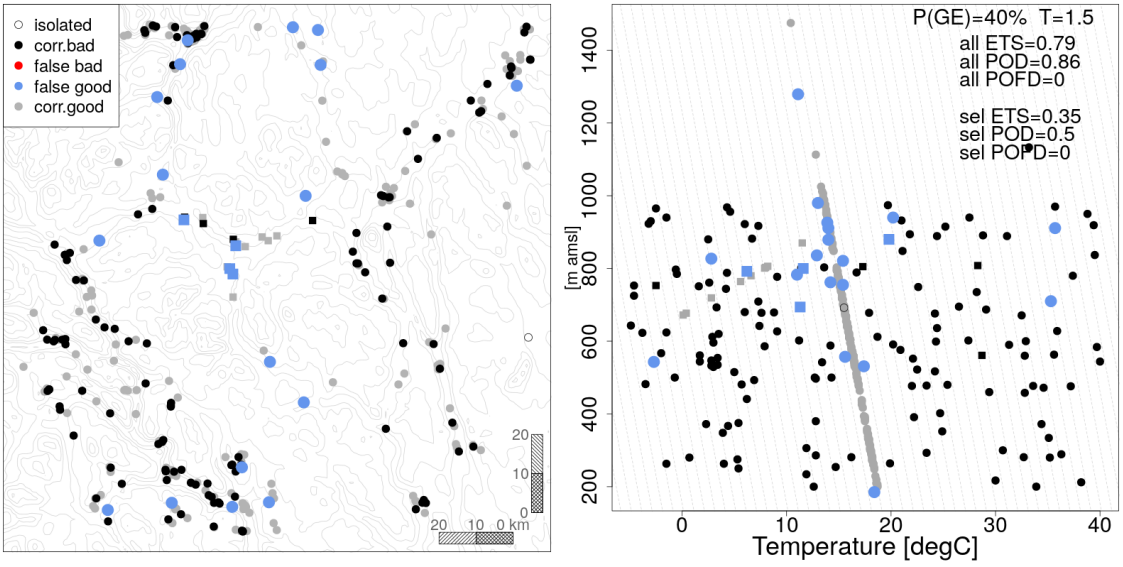


Figure 21: Experiments on synthetic data of temperature generated with $a = 20°\text{C}$, $\text{P(GE)} = 40\%$ and $T = 1.5$. The layout is similar to Fig. 12.

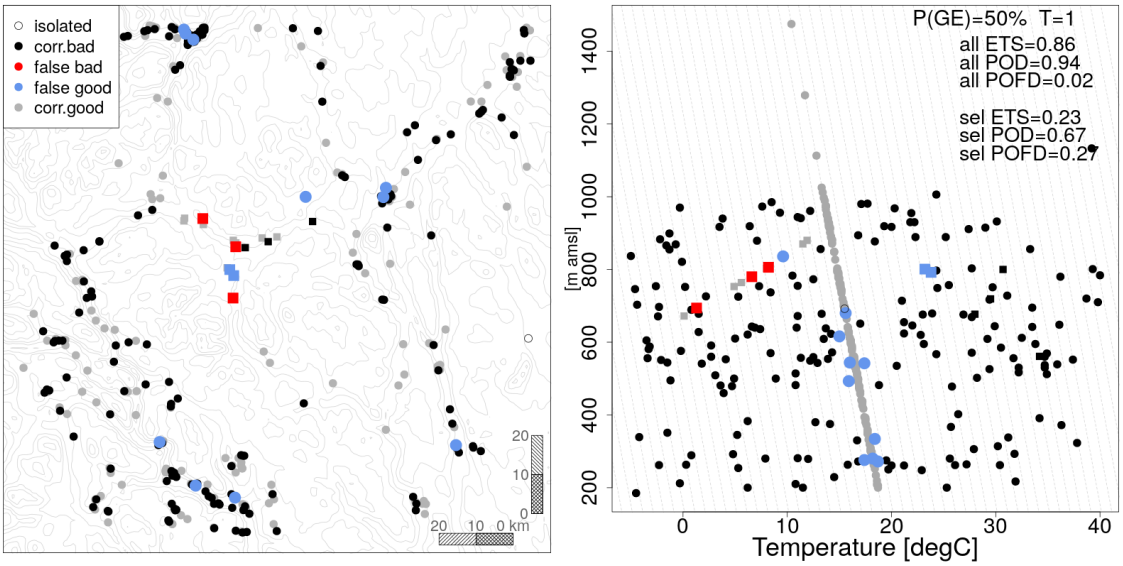


Figure 22: Experiments on synthetic data of temperature generated with $a = 20°\text{C}$, $\text{P(GE)} = 50\%$ and $T = 1$. The layout is similar to Fig. 12.

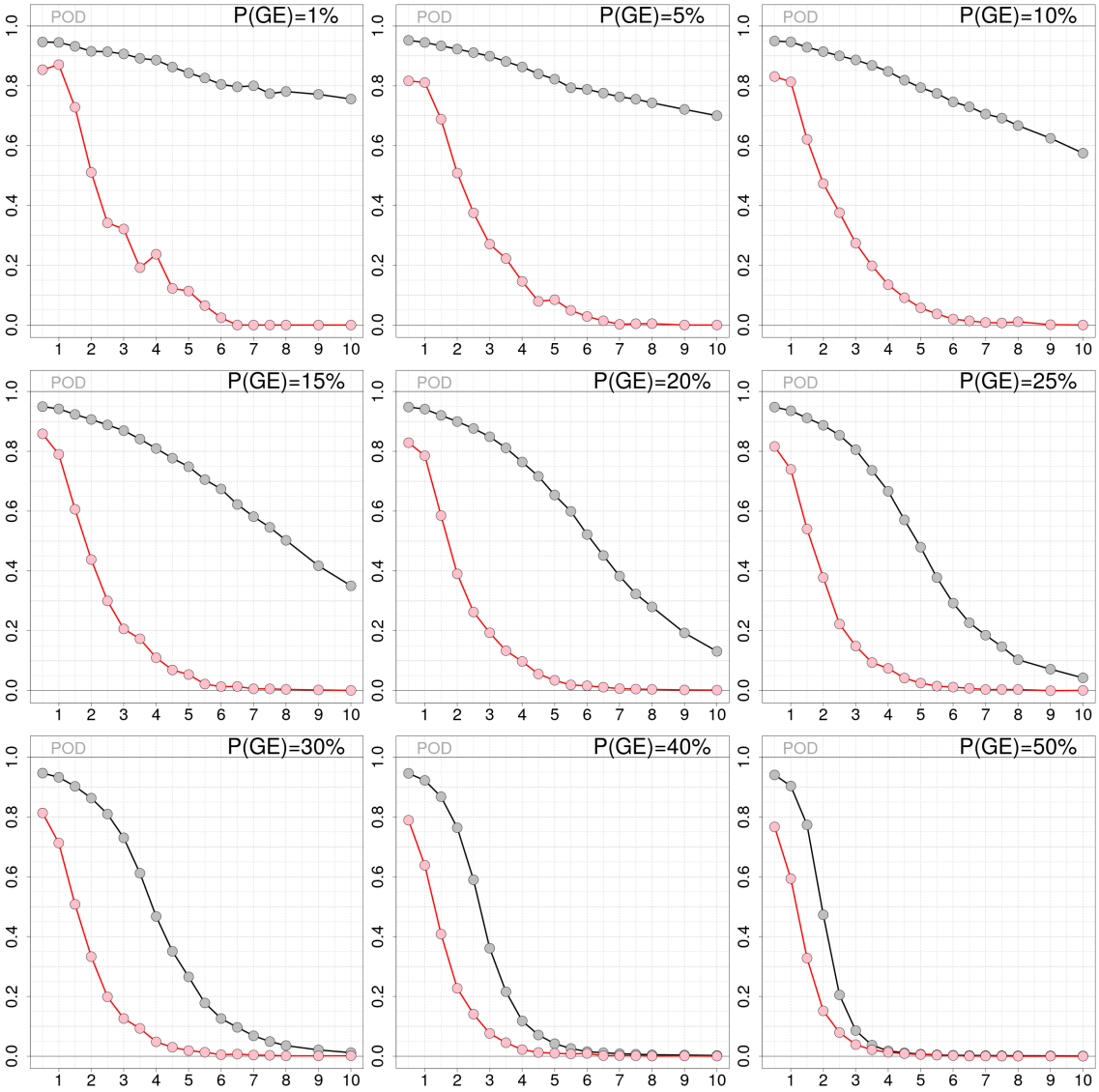


Figure 23: Experiments on synthetic data of temperature generated with $a = -20°C$ (see Fig. 1). Probability of detection (POD) as a function of the SCT threshold for different values of the prior probability of GE occurrence (P(GE)), as indicated on the panels. The dataset is based on 500 simulations. The pink lines show the results when considering observations in Folldal only (i.e. white squares in Fig. 1). The gray lines show the results when considering all observations.

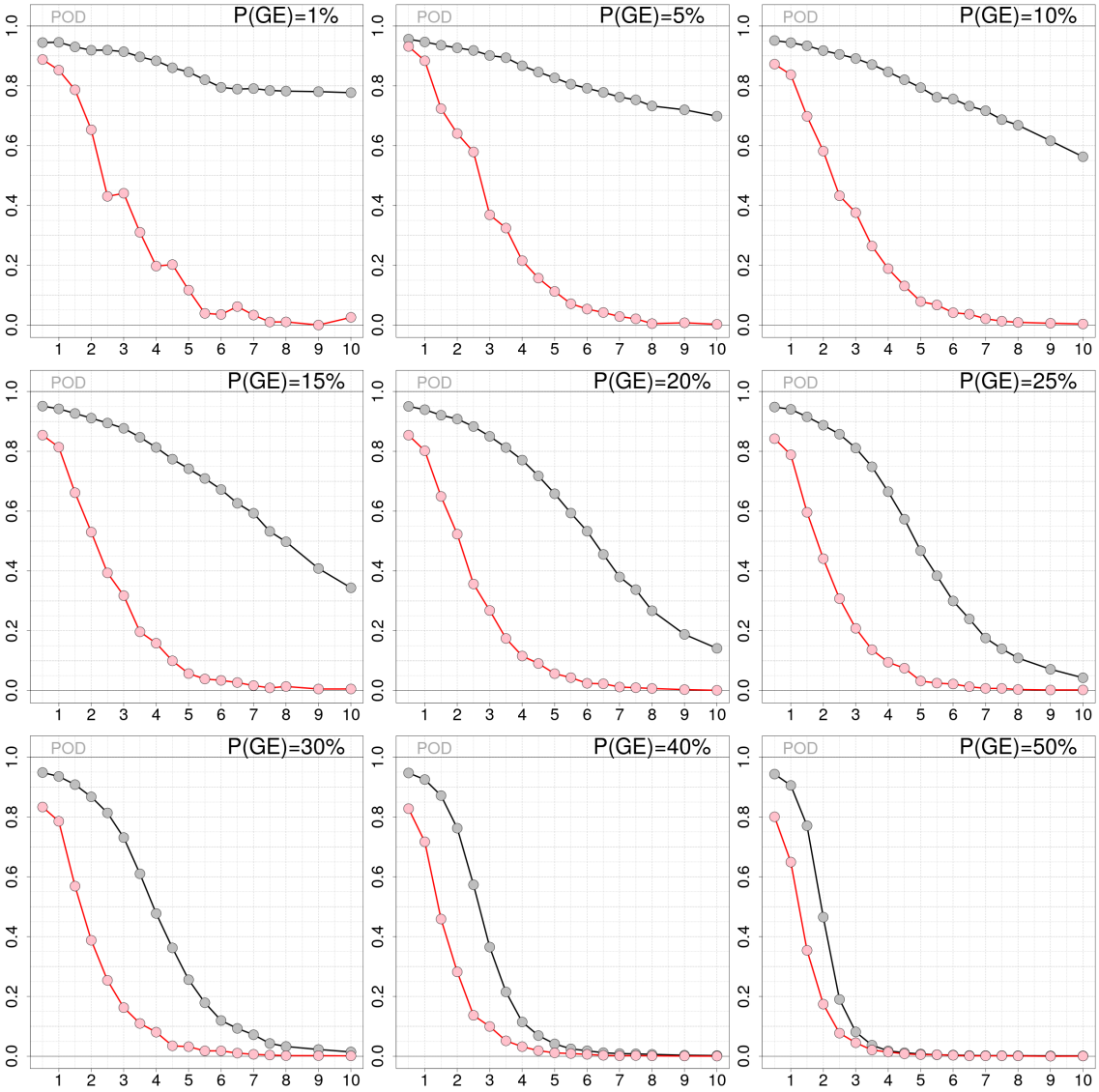


Figure 24: Experiments on synthetic data of temperature generated with $a = -16°C$ (see Fig. 1). Probability of detection (POD) as a function of the SCT threshold for different values of the prior probability of GE occurrence (P(GE)), as indicated on the panels. The dataset is based on 500 simulations. The pink lines show the results when considering observations in Folldal only (i.e. white squares in Fig. 1). The gray lines show the results when considering all observations.

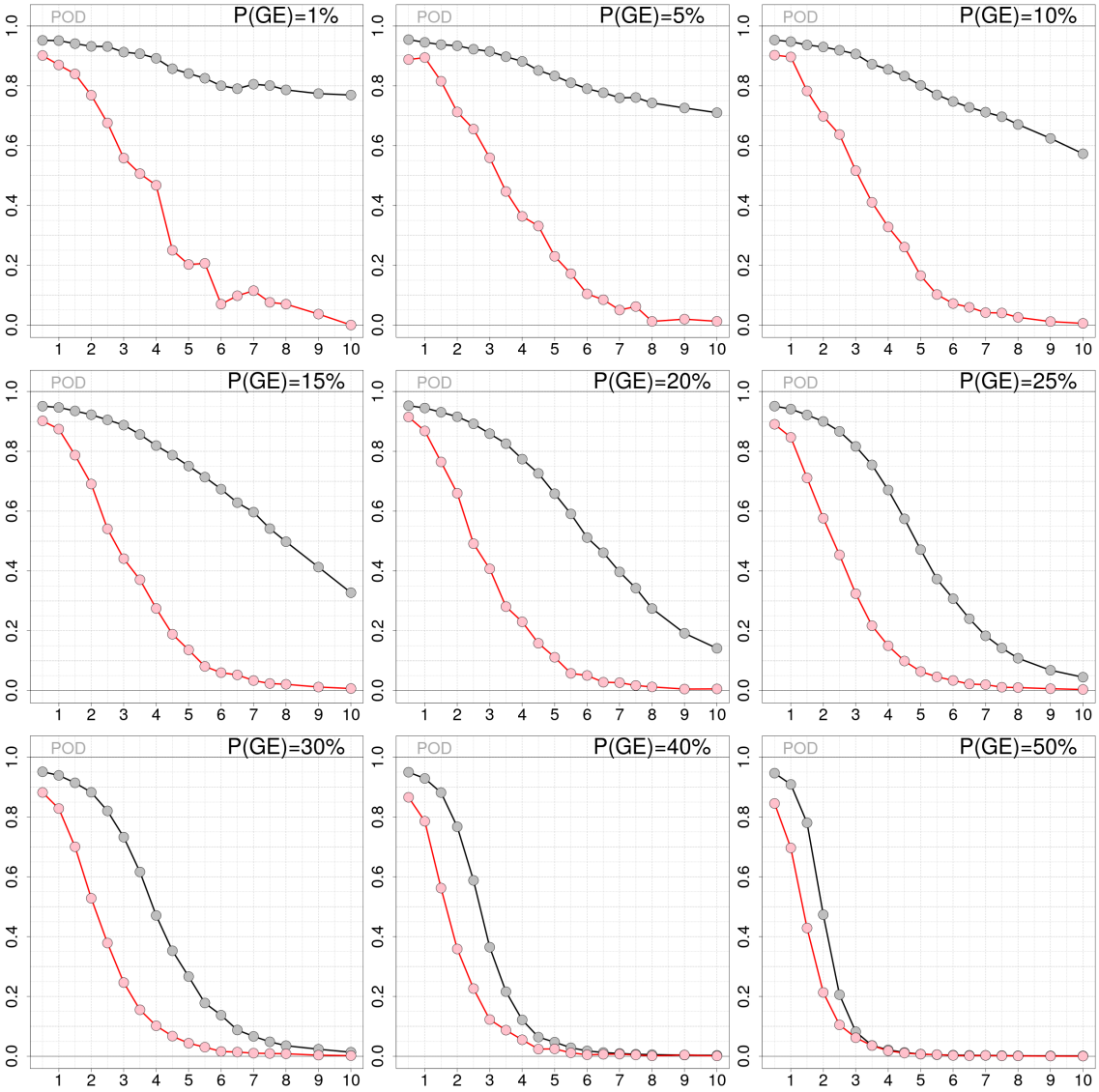


Figure 25: Experiments on synthetic data of temperature generated with $a = -11°C$ (see Fig. 1). Probability of detection (POD) as a function of the SCT threshold for different values of the prior probability of GE occurrence (P(GE)), as indicated on the panels. The dataset is based on 500 simulations. The pink lines show the results when considering observations in Folldal only (i.e. white squares in Fig. 1). The gray lines show the results when considering all observations.

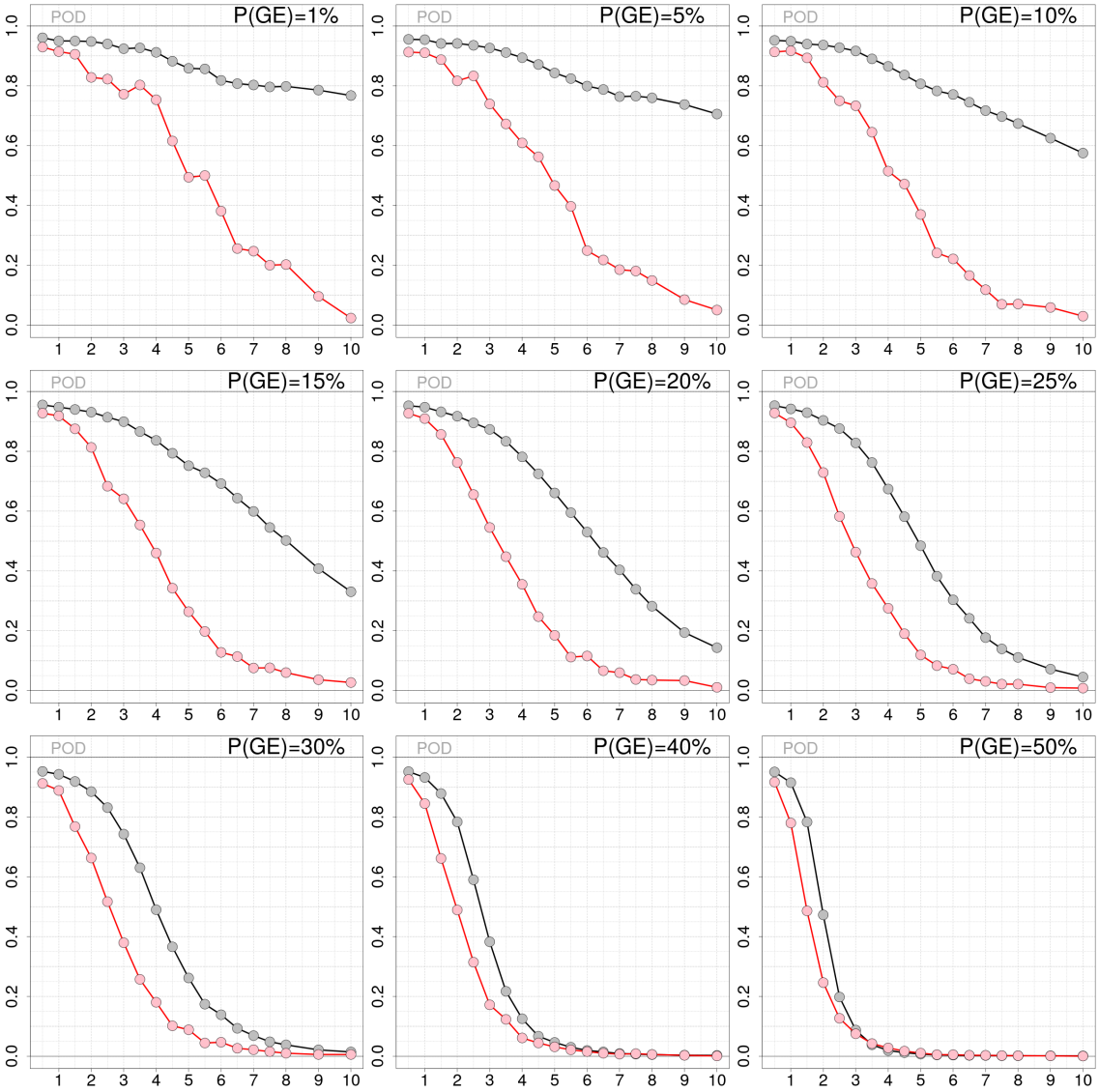


Figure 26: Experiments on synthetic data of temperature generated with $a = -6°C$ (see Fig. 1). Probability of detection (POD) as a function of the SCT threshold for different values of the prior probability of GE occurrence (P(GE)), as indicated on the panels. The dataset is based on 500 simulations. The pink lines show the results when considering observations in Folldal only (i.e. white squares in Fig. 1). The gray lines show the results when considering all observations.

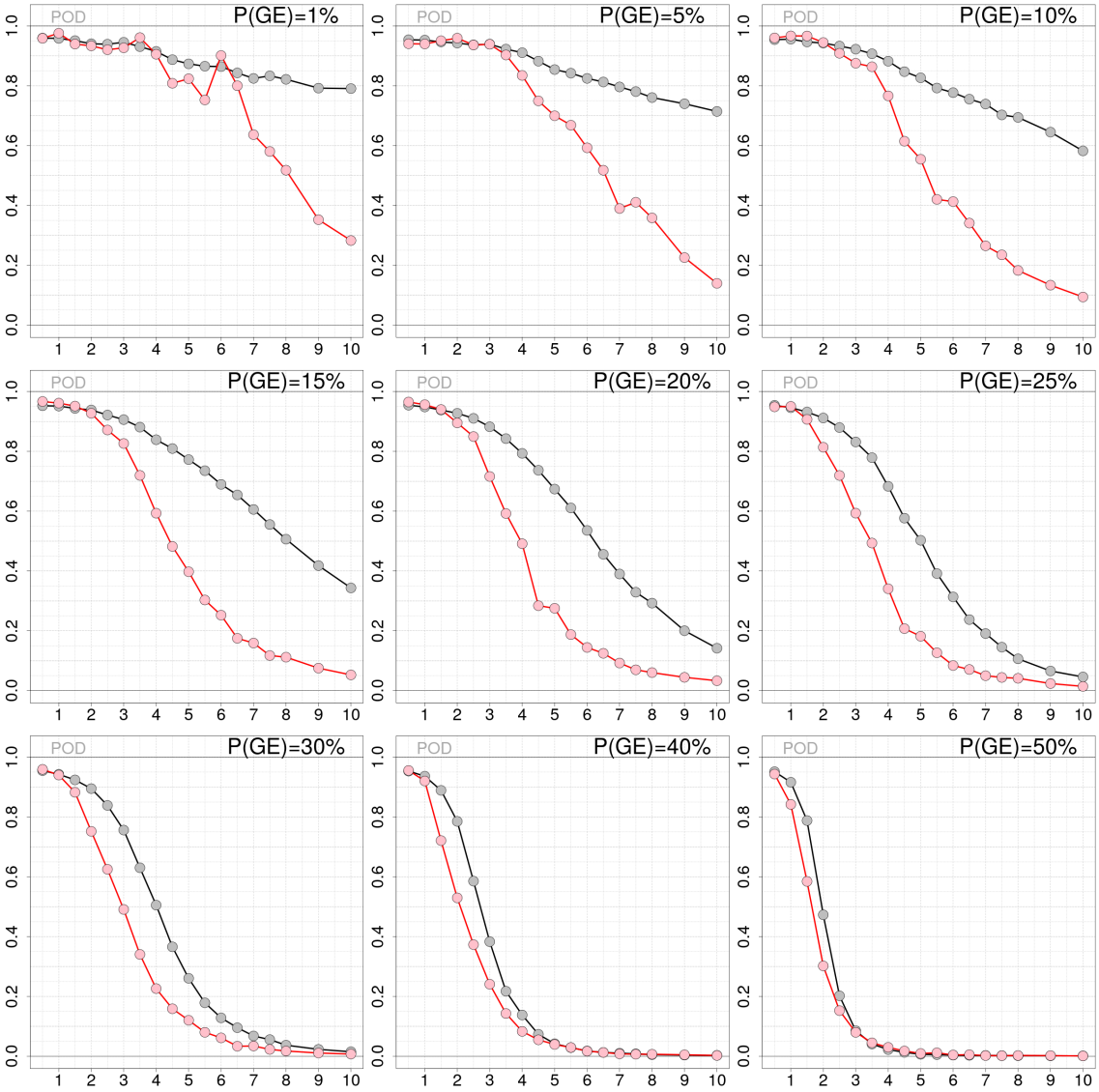


Figure 27: Experiments on synthetic data of temperature generated with $a = 0°C$ (see Fig. 1). Probability of detection (POD) as a function of the SCT threshold for different values of the prior probability of GE occurrence (P(GE)), as indicated on the panels. The dataset is based on 500 simulations. The pink lines show the results when considering observations in Folldal only (i.e. white squares in Fig. 1). The gray lines show the results when considering all observations.

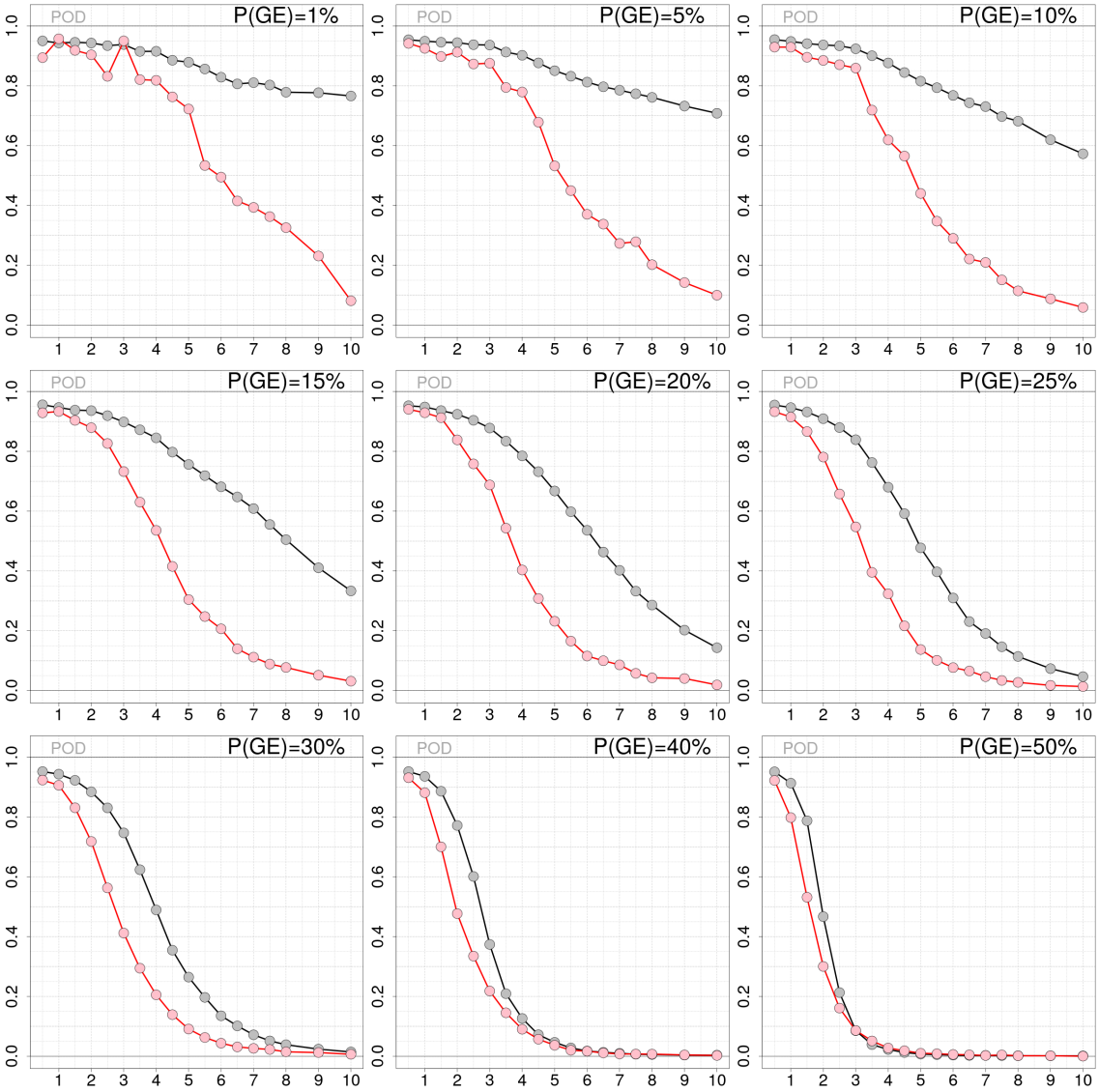


Figure 28: Experiments on synthetic data of temperature generated with $a = 4°C$ (see Fig. 1). Probability of detection (POD) as a function of the SCT threshold for different values of the prior probability of GE occurrence (P(GE)), as indicated on the panels. The dataset is based on 500 simulations. The pink lines show the results when considering observations in Folldal only (i.e. white squares in Fig. 1). The gray lines show the results when considering all observations.

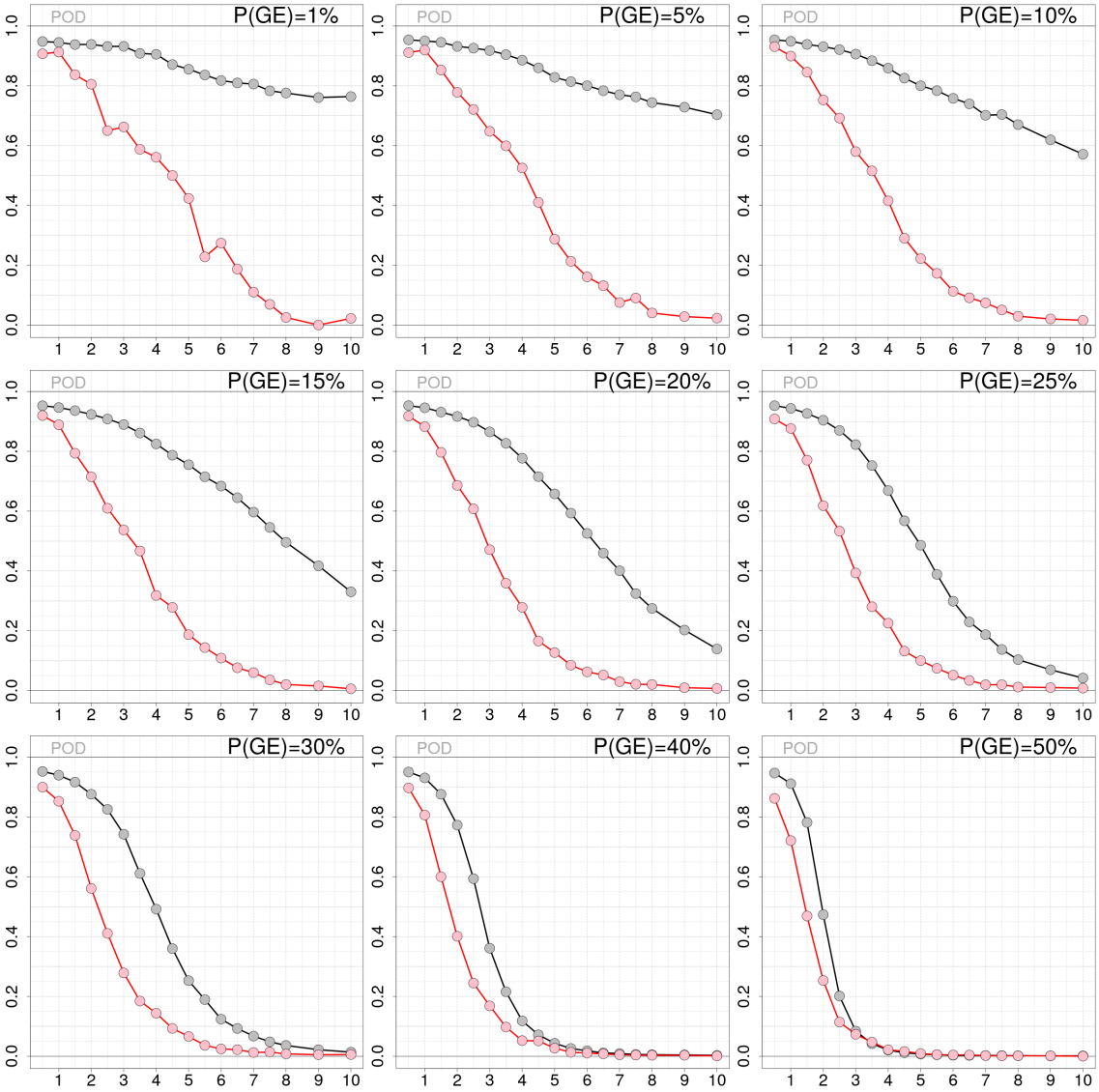


Figure 29: Experiments on synthetic data of temperature generated with $a = 9°C$ (see Fig. 1). Probability of detection (POD) as a function of the SCT threshold for different values of the prior probability of GE occurrence (P(GE)), as indicated on the panels. The dataset is based on 500 simulations. The pink lines show the results when considering observations in Folldal only (i.e. white squares in Fig. 1). The gray lines show the results when considering all observations.

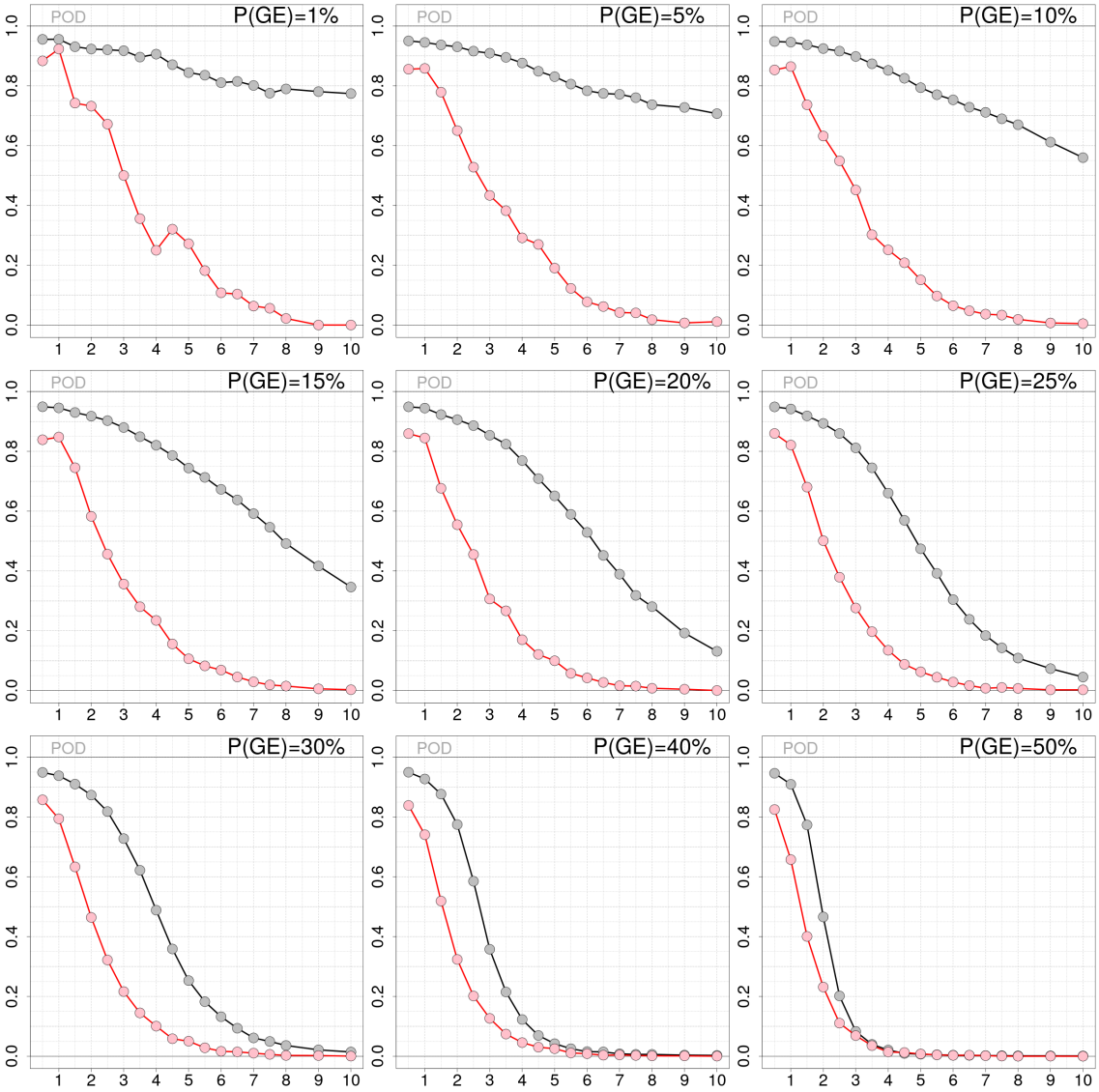


Figure 30: Experiments on synthetic data of temperature generated with $a = 14°C$ (see Fig. 1). Probability of detection (POD) as a function of the SCT threshold for different values of the prior probability of GE occurrence (P(GE)), as indicated on the panels. The dataset is based on 500 simulations. The pink lines show the results when considering observations in Folldal only (i.e. white squares in Fig. 1). The gray lines show the results when considering all observations.

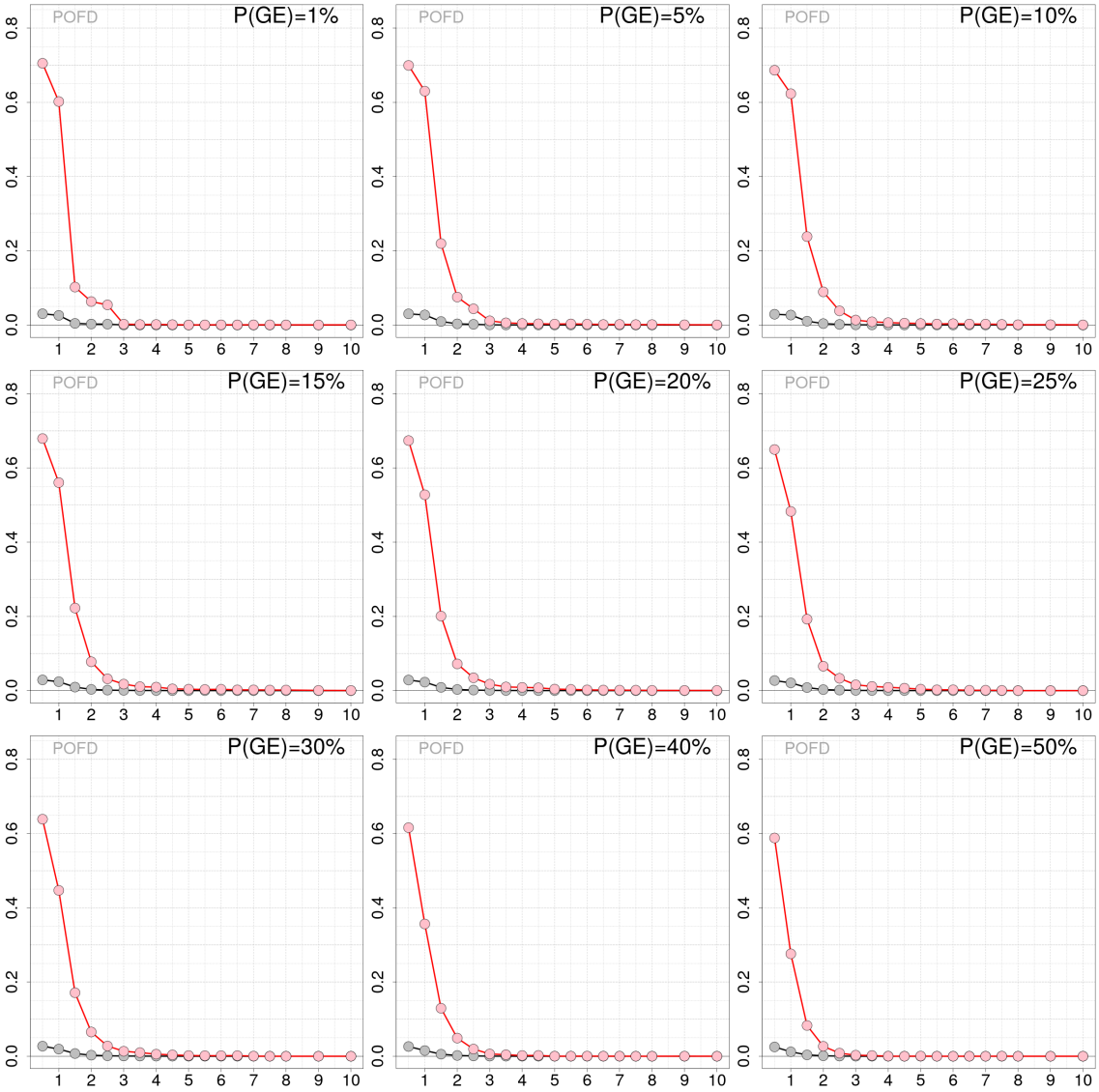


Figure 31: Experiments on synthetic data of temperature generated with $a = -20°C$ (see Fig. 1). Probability of false detection (POFD) as a function of the SCT threshold for different values of the prior probability of GE occurrence (P(GE)), as indicated on the panels. The dataset is based on 500 simulations. The pink lines show the results when considering observations in Folldal only (i.e. white squares in Fig. 1). The gray lines show the results when considering all observations.

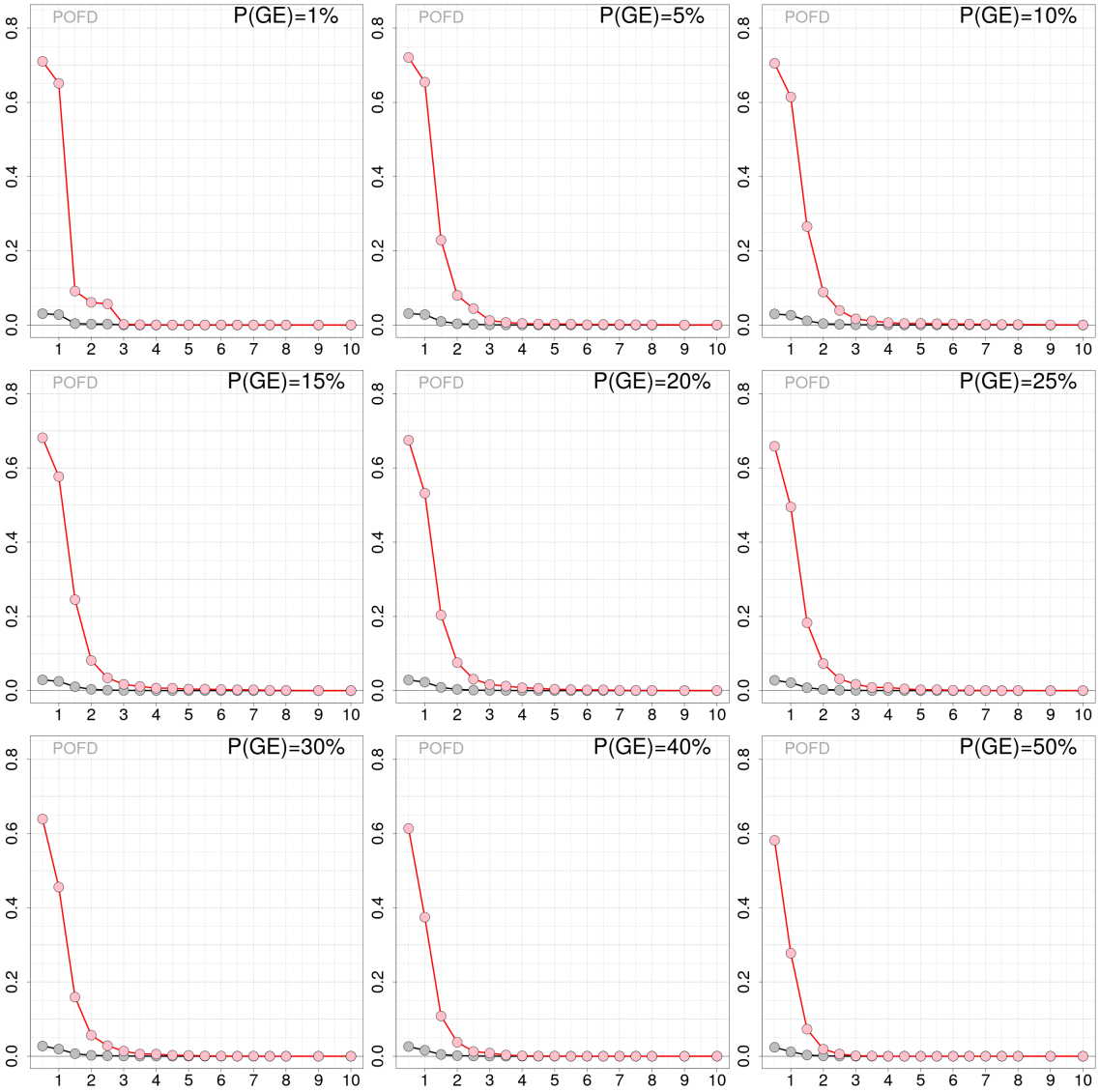


Figure 32: Experiments on synthetic data of temperature generated with $a = -16°C$ (see Fig. 1). Probability of false detection (POFD) as a function of the SCT threshold for different values of the prior probability of GE occurrence (P(GE)), as indicated on the panels. The dataset is based on 500 simulations. The pink lines show the results when considering observations in Folldal only (i.e. white squares in Fig. 1). The gray lines show the results when considering all observations.

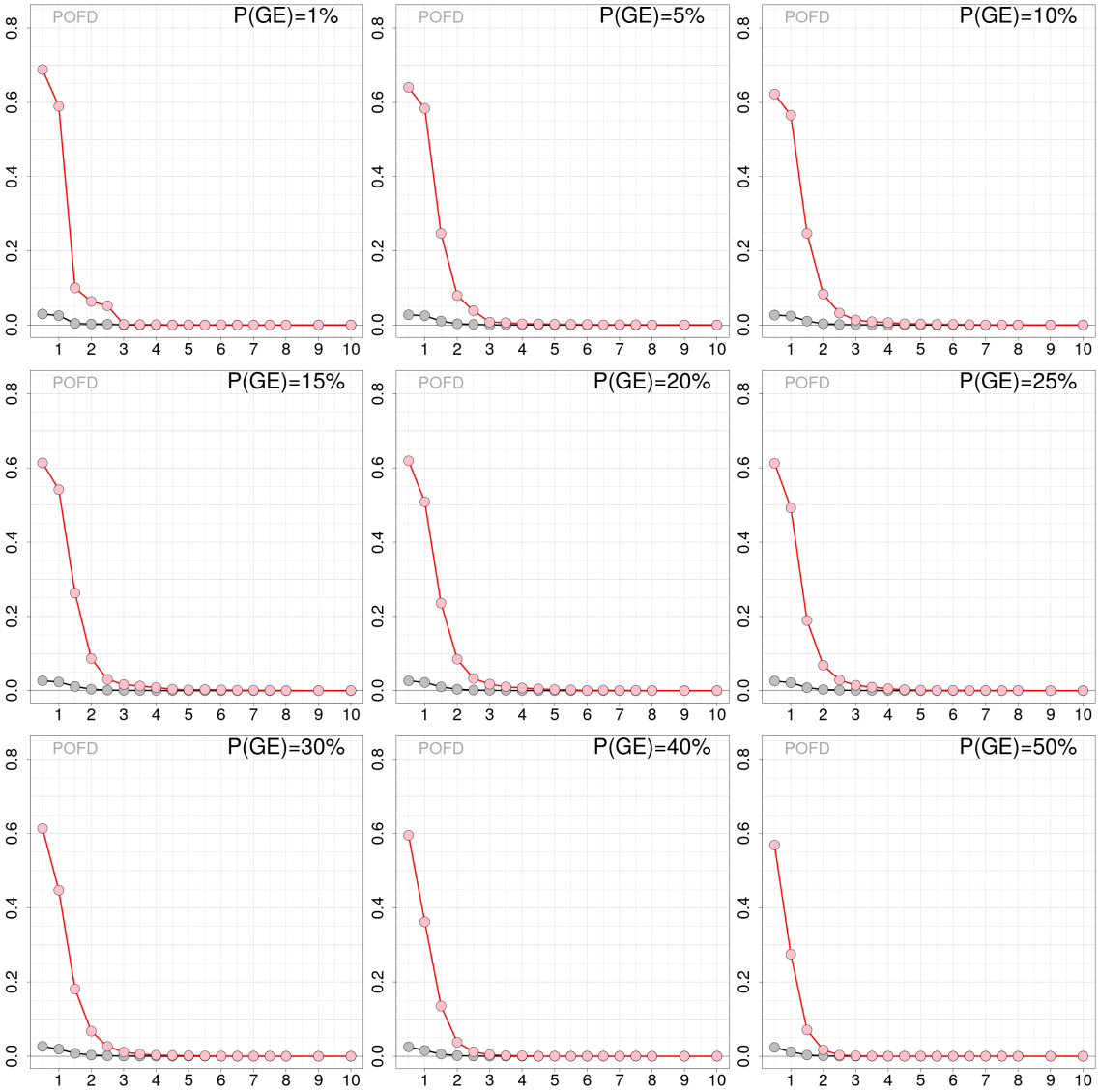


Figure 33: Experiments on synthetic data of temperature generated with $a = -11°C$ (see Fig. 1). Probability of false detection (POFD) as a function of the SCT threshold for different values of the prior probability of GE occurrence (P(GE)), as indicated on the panels. The dataset is based on 500 simulations. The pink lines show the results when considering observations in Folldal only (i.e. white squares in Fig. 1). The gray lines show the results when considering all observations.

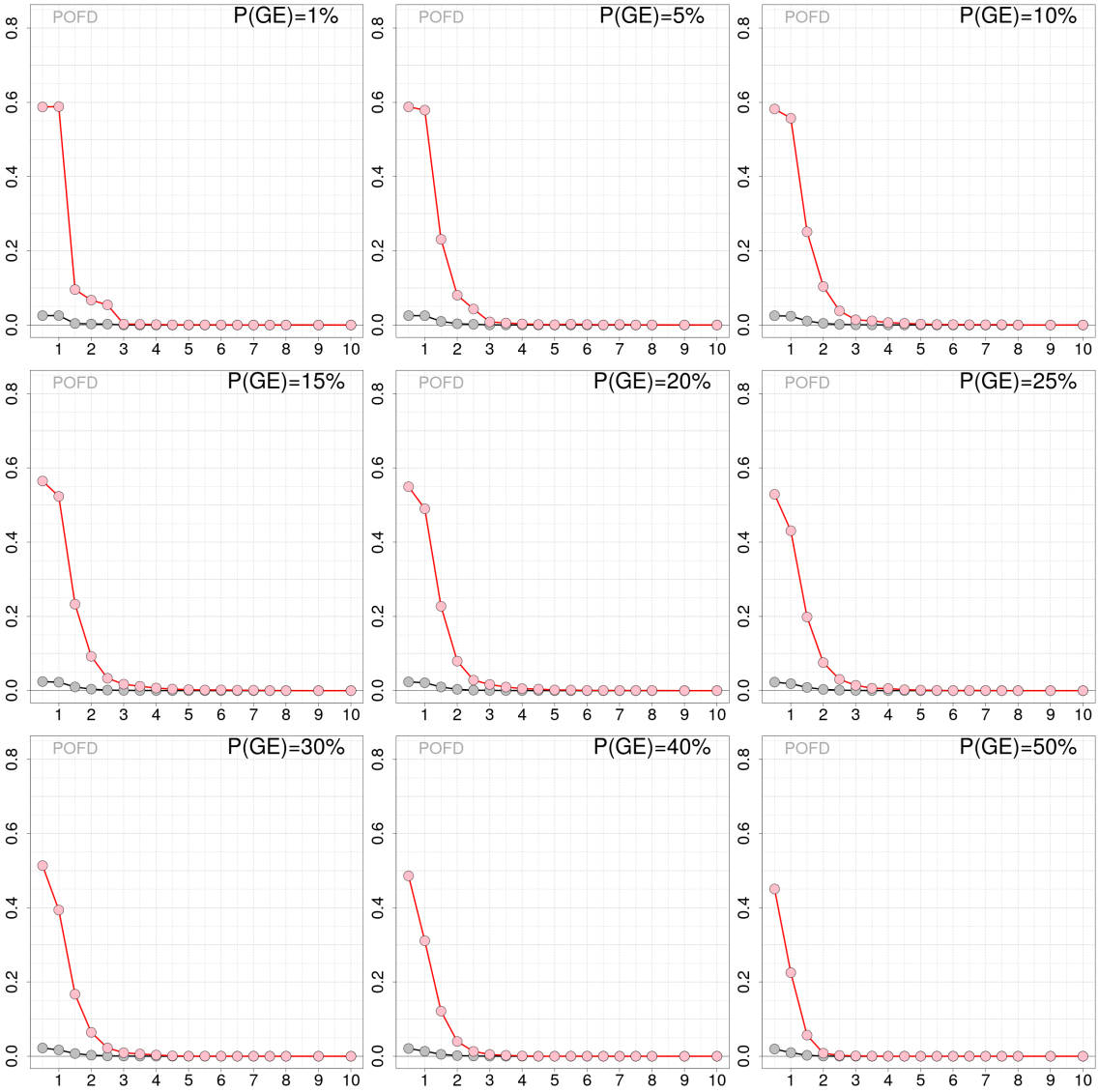


Figure 34: Experiments on synthetic data of temperature generated with $a = -6°C$ (see Fig. 1). Probability of false detection (POFD) as a function of the SCT threshold for different values of the prior probability of GE occurrence (P(GE)), as indicated on the panels. The dataset is based on 500 simulations. The pink lines show the results when considering observations in Folldal only (i.e. white squares in Fig. 1). The gray lines show the results when considering all observations.

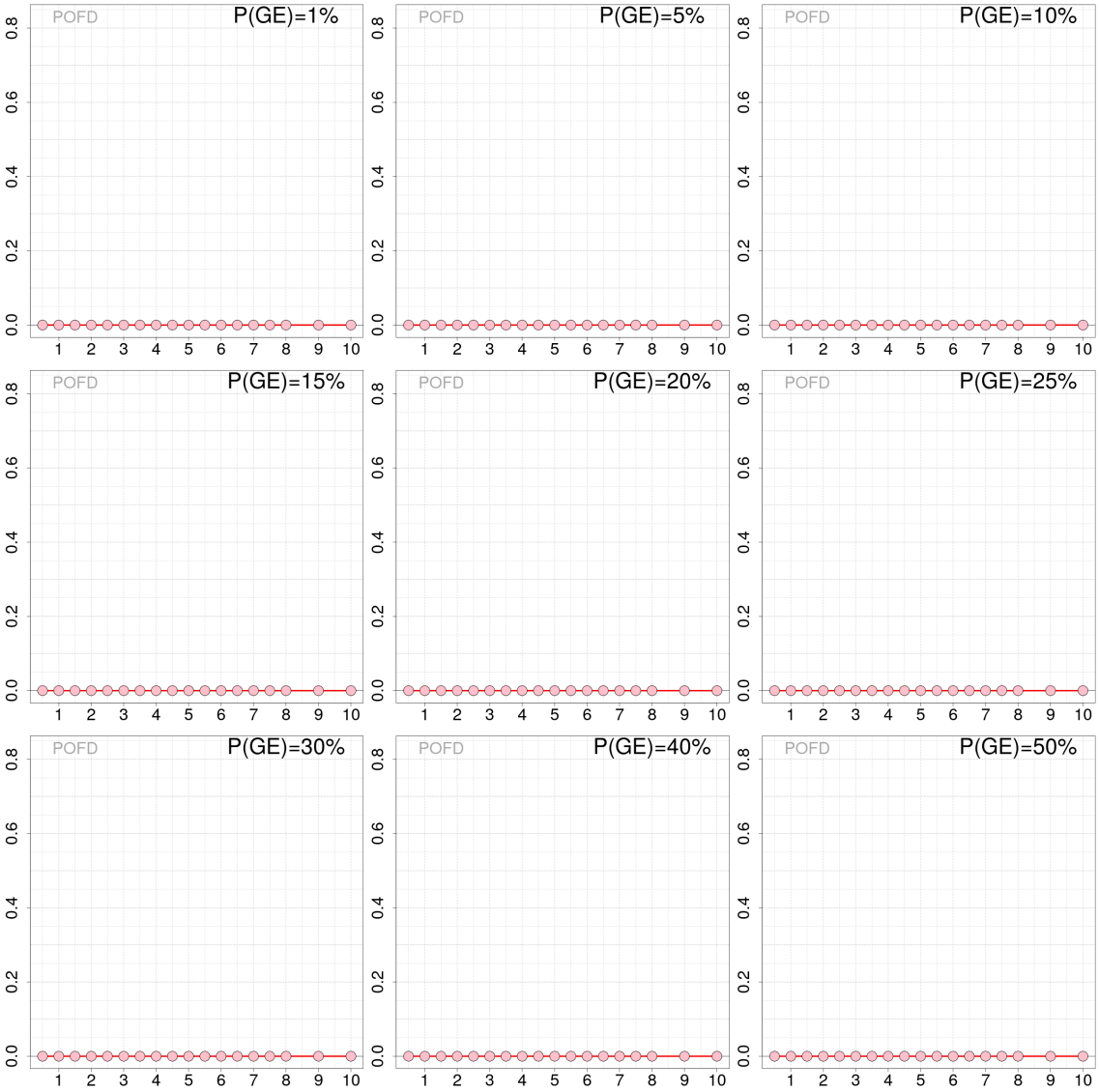


Figure 35: Experiments on synthetic data of temperature generated with $a = 0°C$ (see Fig. 1). Probability of false detection (POFD) as a function of the SCT threshold for different values of the prior probability of GE occurrence (P(GE)), as indicated on the panels. The dataset is based on 500 simulations. The pink lines show the results when considering observations in Folldal only (i.e. white squares in Fig. 1). The gray lines show the results when considering all observations.

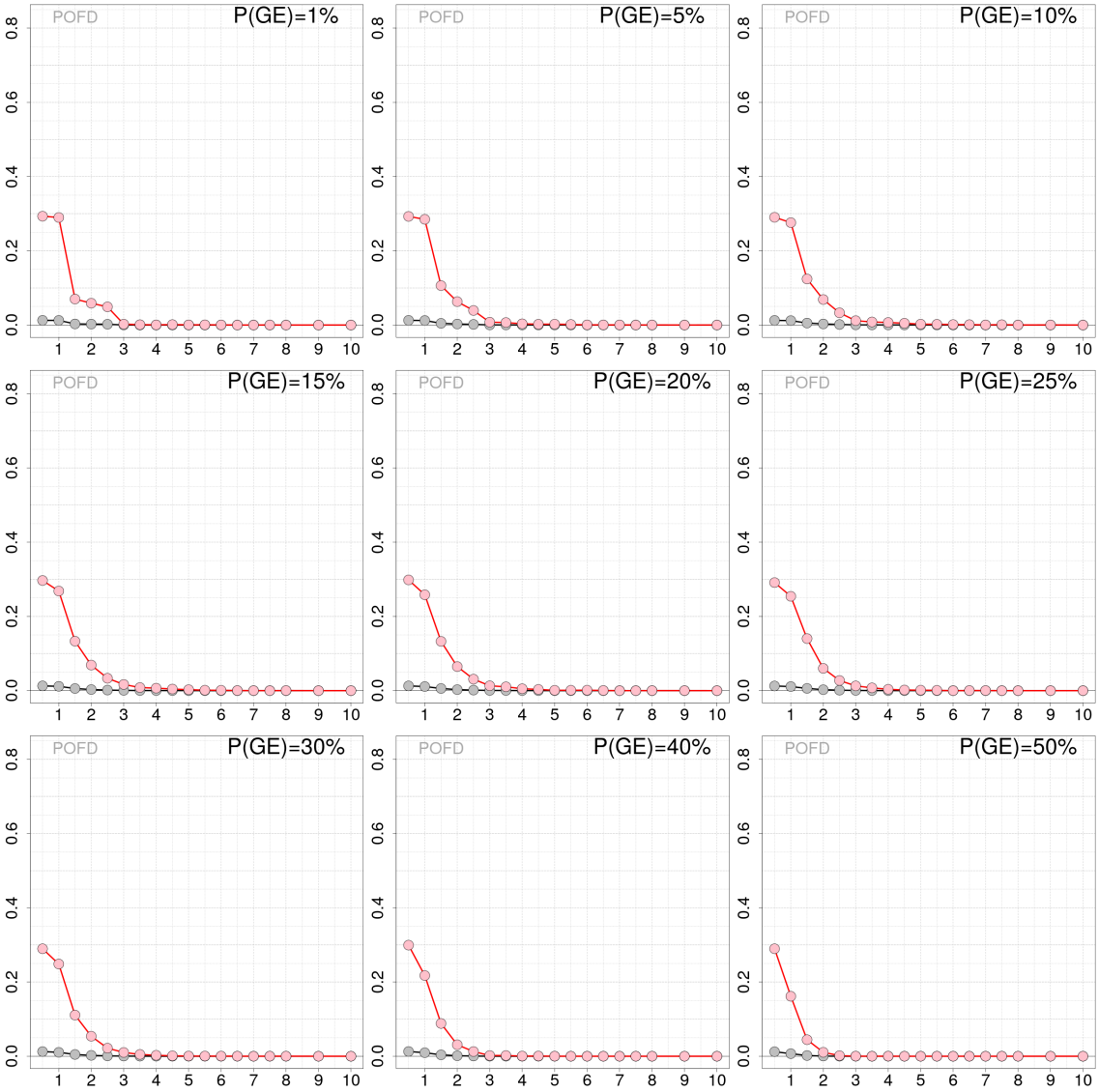


Figure 36: Experiments on synthetic data of temperature generated with $a = 4°C$ (see Fig. 1). Probability of false detection (POFD) as a function of the SCT threshold for different values of the prior probability of GE occurrence (P(GE)), as indicated on the panels. The dataset is based on 500 simulations. The pink lines show the results when considering observations in Folldal only (i.e. white squares in Fig. 1). The gray lines show the results when considering all observations.

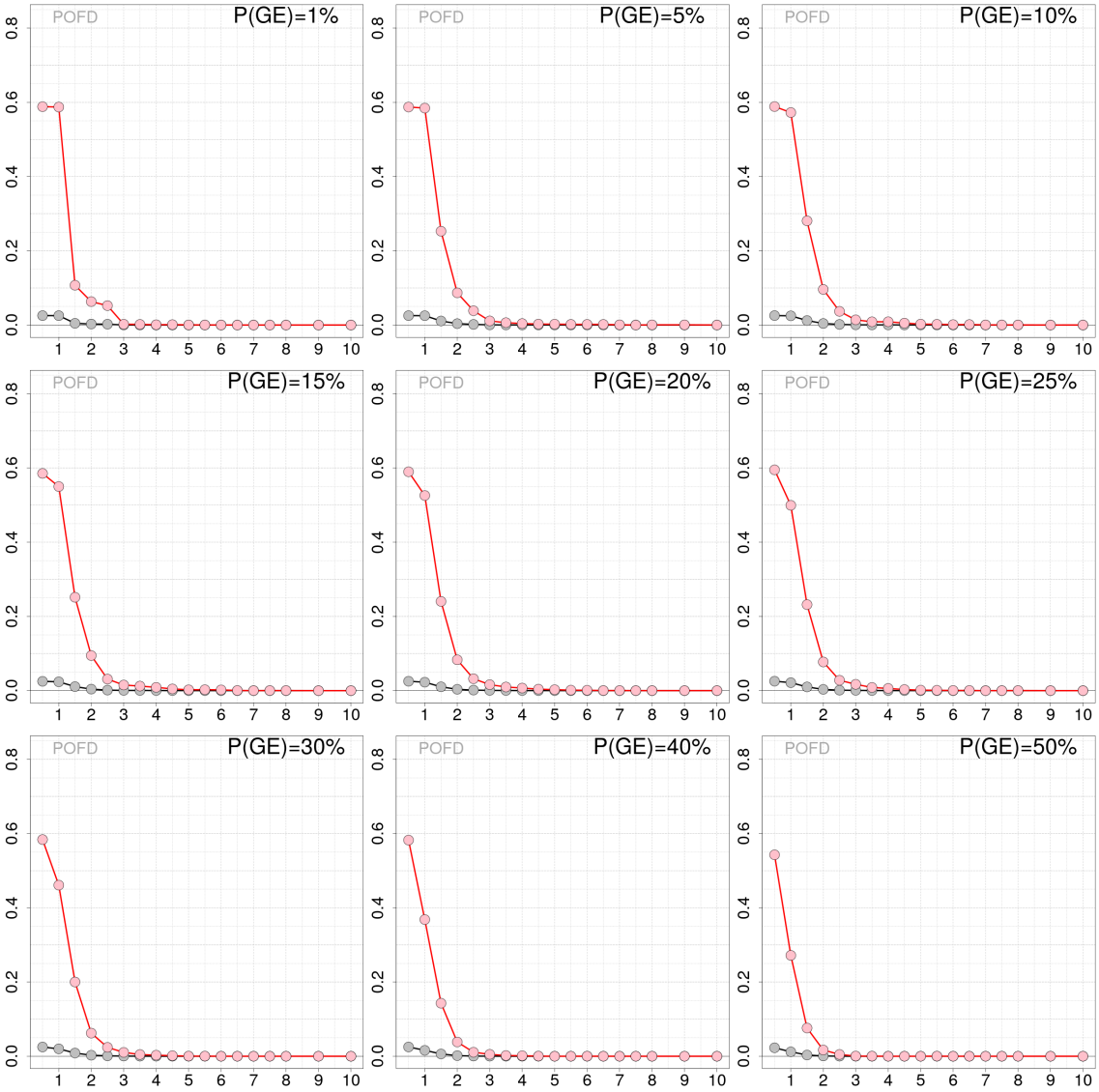


Figure 37: Experiments on synthetic data of temperature generated with $a = 9°C$ (see Fig. 1). Probability of false detection (POFD) as a function of the SCT threshold for different values of the prior probability of GE occurrence (P(GE)), as indicated on the panels. The dataset is based on 500 simulations. The pink lines show the results when considering observations in Folldal only (i.e. white squares in Fig. 1). The gray lines show the results when considering all observations.

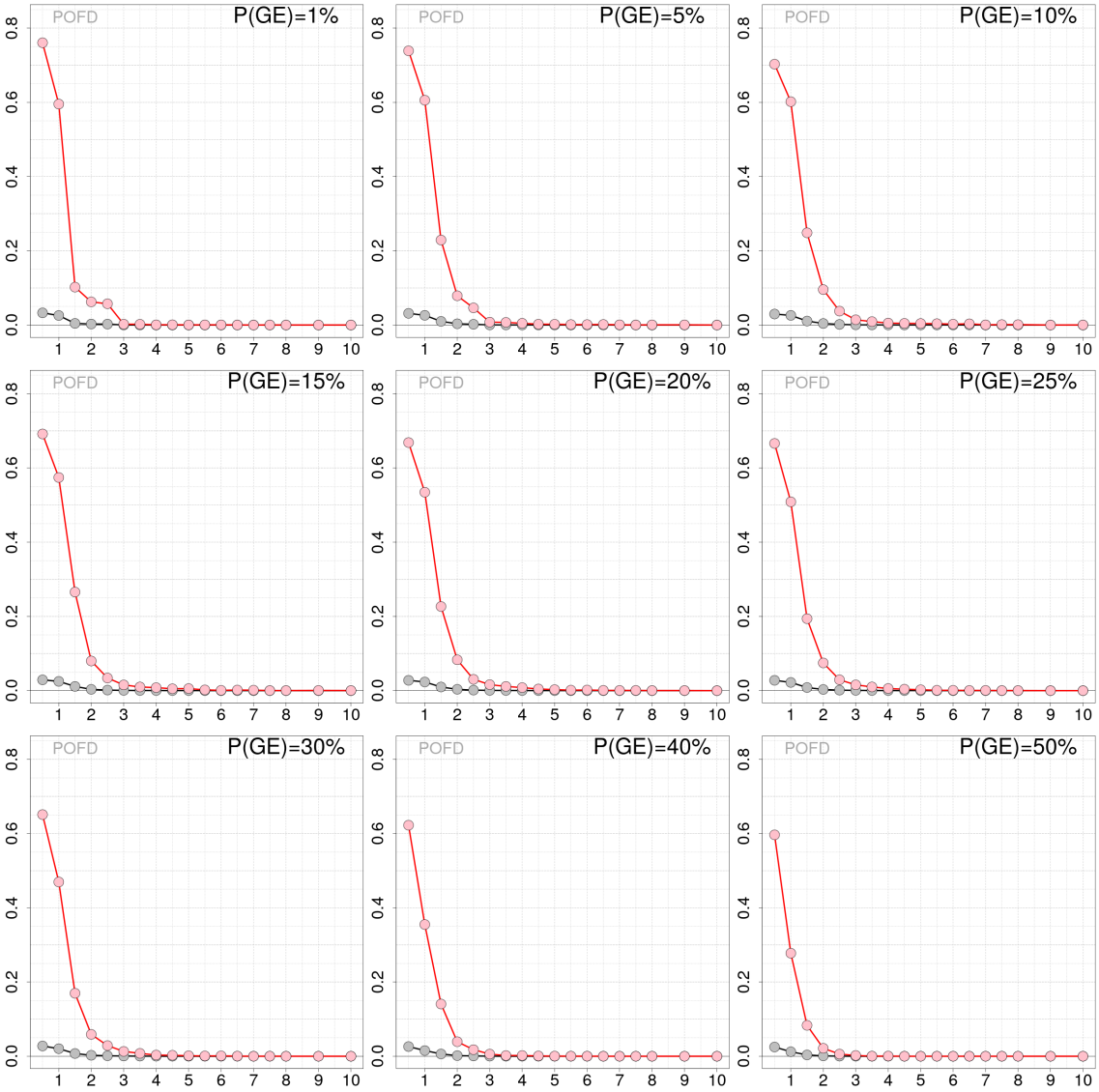


Figure 38: Experiments on synthetic data of temperature generated with $a = 14°C$ (see Fig. 1). Probability of false detection (POFD) as a function of the SCT threshold for different values of the prior probability of GE occurrence (P(GE)), as indicated on the panels. The dataset is based on 500 simulations. The pink lines show the results when considering observations in Folldal only (i.e. white squares in Fig. 1). The gray lines show the results when considering all observations.

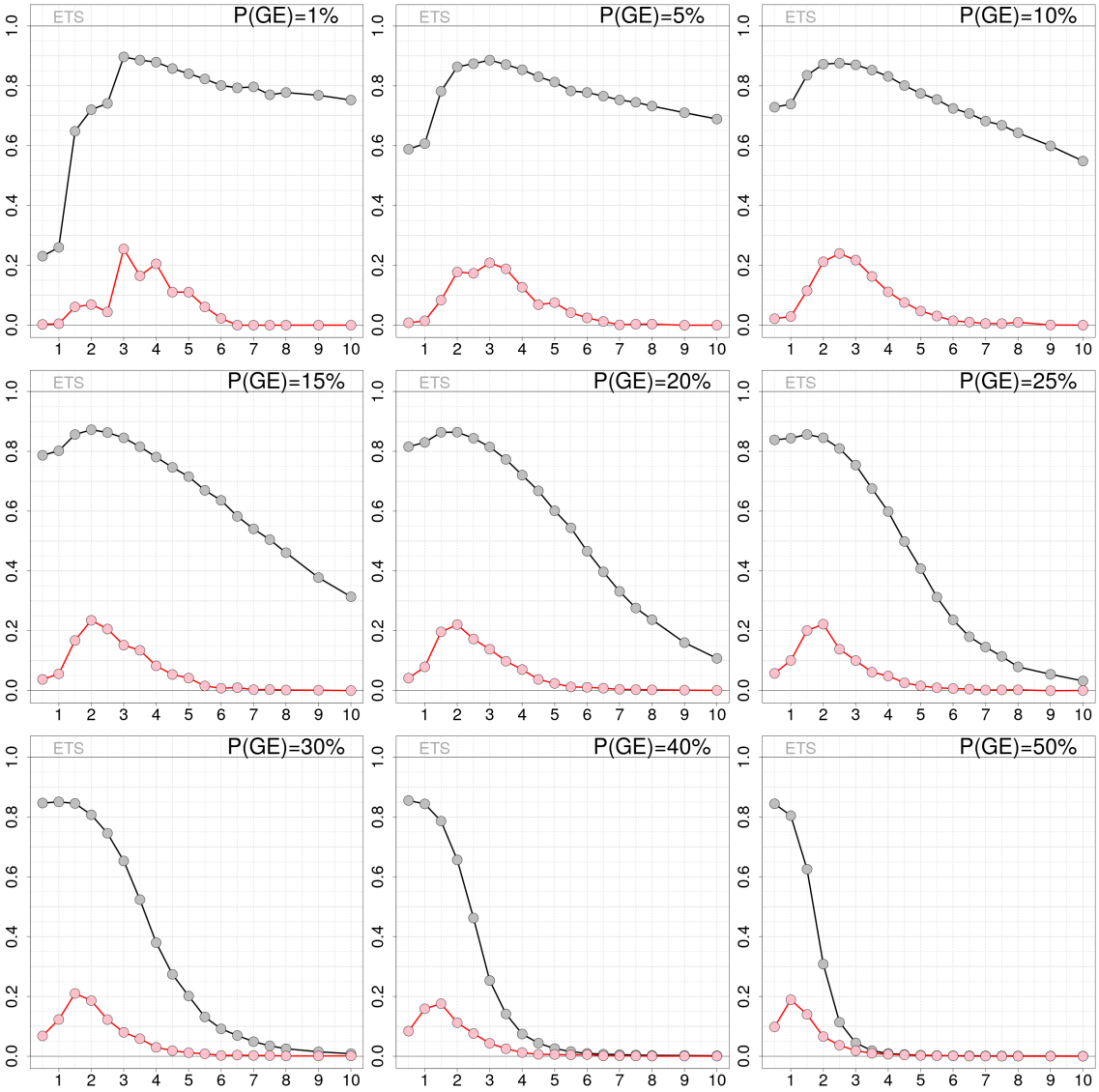


Figure 39: Experiments on synthetic data of temperature generated with $a = -20°C$ (see Fig. 1). Equitable threat score (ETS) as a function of the SCT threshold for different values of the prior probability of GE occurrence (P(GE)), as indicated on the panels. The dataset is based on 500 simulations. The pink lines show the results when considering observations in Folldal only (i.e. white squares in Fig. 1). The gray lines show the results when considering all observations.

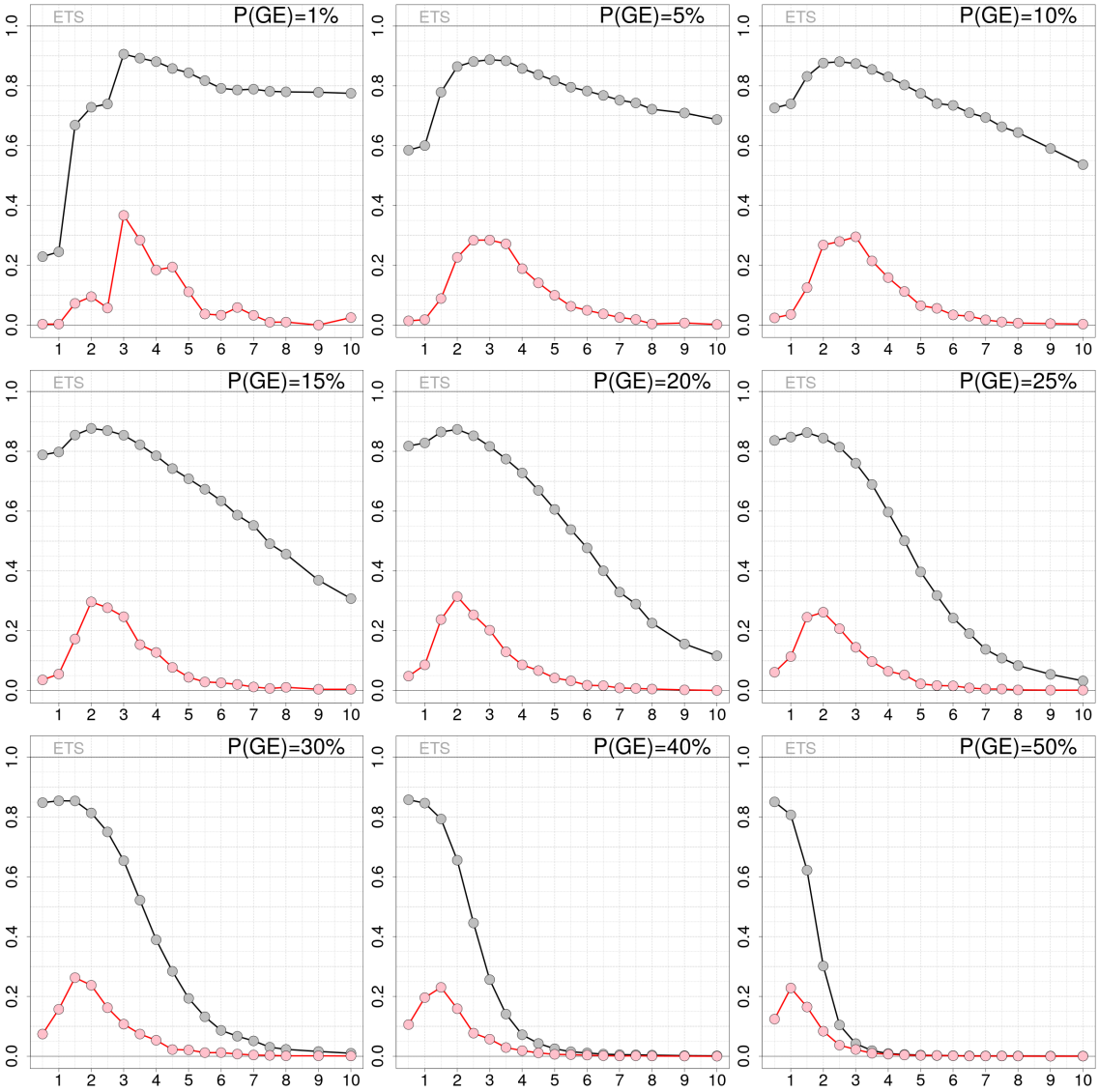


Figure 40: Experiments on synthetic data of temperature generated with $a = -16°C$ (see Fig. 1). Equitable threat score (ETS) as a function of the SCT threshold for different values of the prior probability of GE occurrence (P(GE)), as indicated on the panels. The dataset is based on 500 simulations. The pink lines show the results when considering observations in Folldal only (i.e. white squares in Fig. 1). The gray lines show the results when considering all observations.

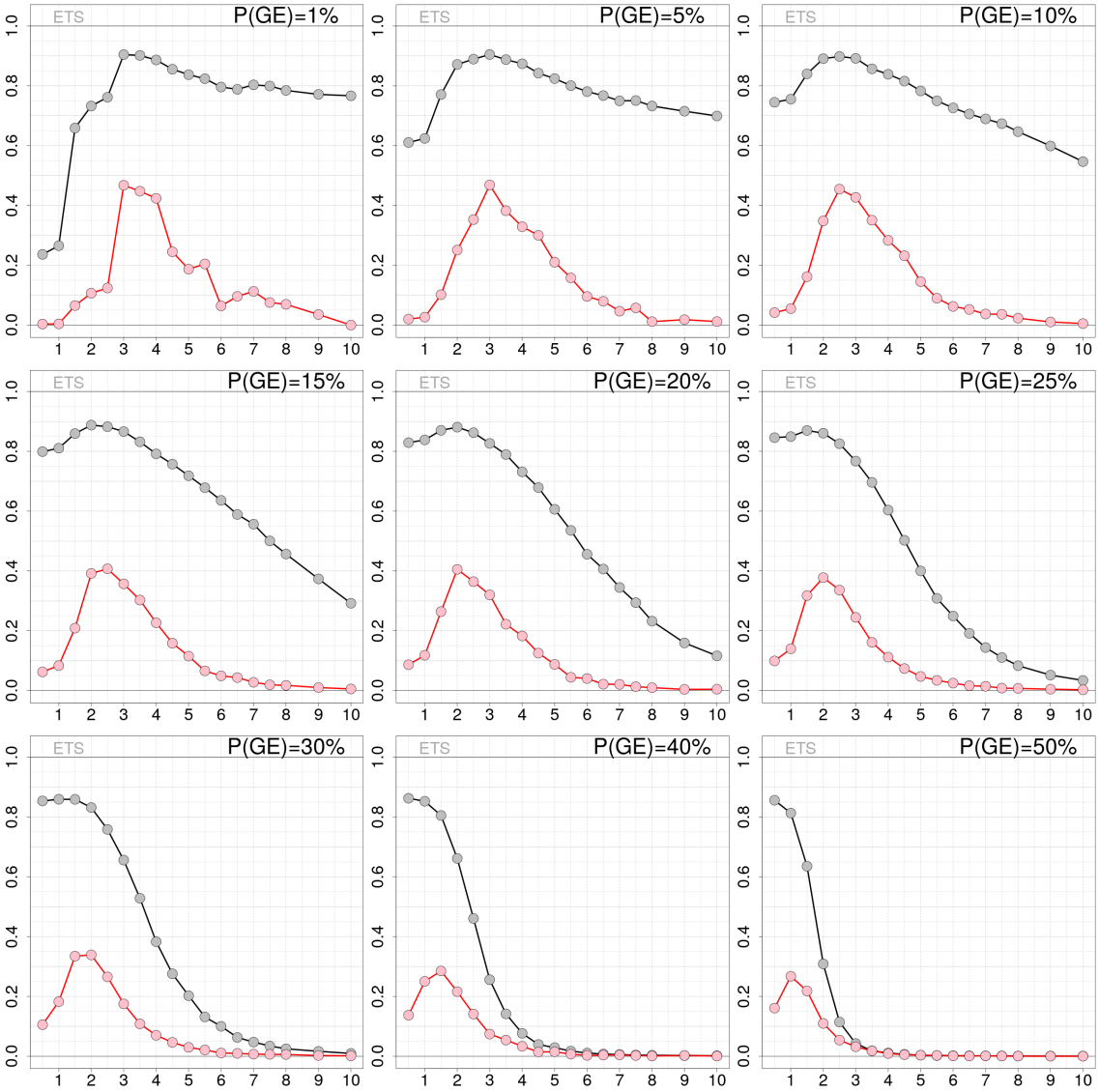


Figure 41: Experiments on synthetic data of temperature generated with $a = -11°C$ (see Fig. 1). Equitable threat score (ETS) as a function of the SCT threshold for different values of the prior probability of GE occurrence (P(GE)), as indicated on the panels. The dataset is based on 500 simulations. The pink lines show the results when considering observations in Folldal only (i.e. white squares in Fig. 1). The gray lines show the results when considering all observations.

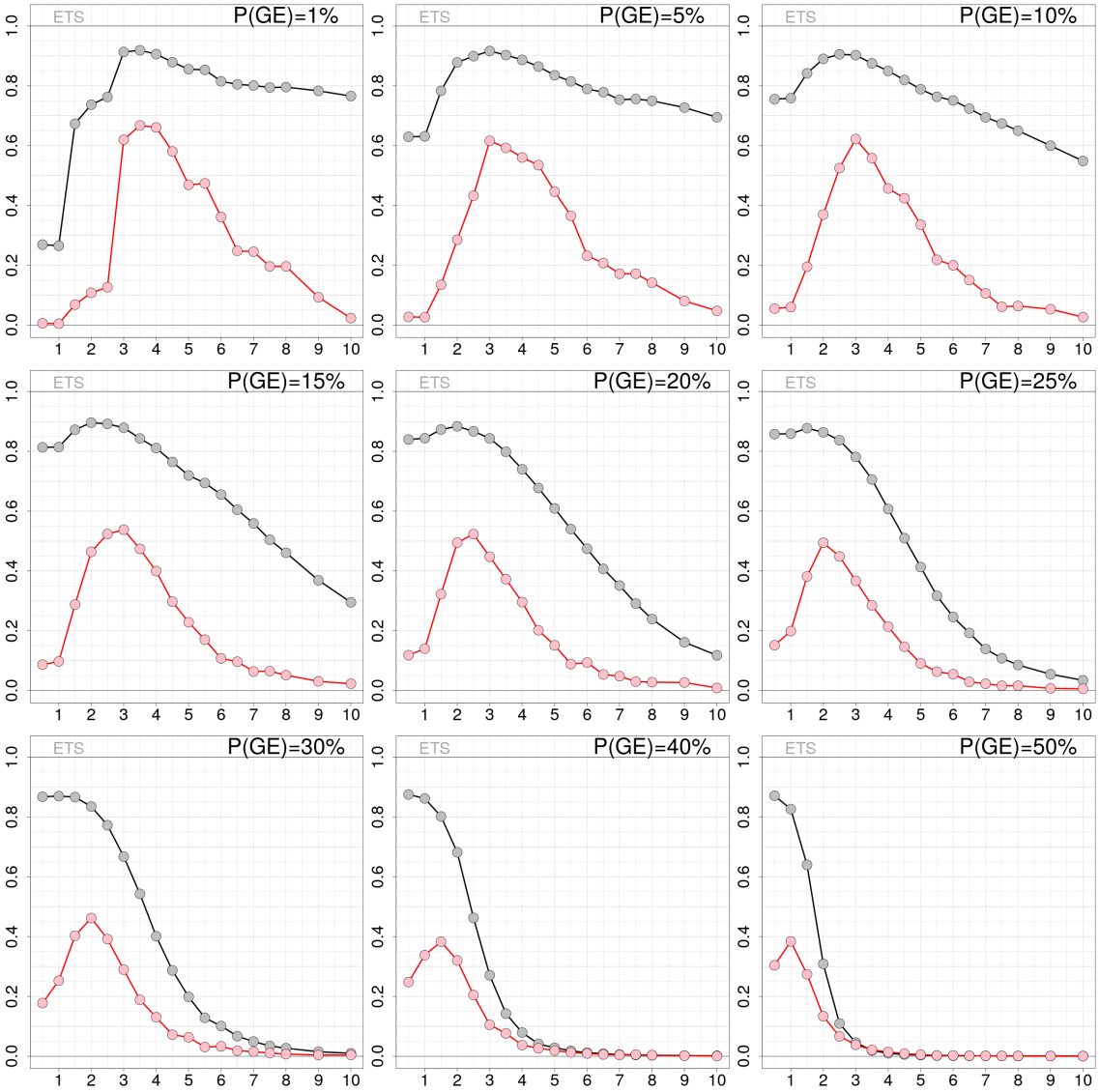


Figure 42: Experiments on synthetic data of temperature generated with $a = -6°C$ (see Fig. 1). Equitable threat score (ETS) as a function of the SCT threshold for different values of the prior probability of GE occurrence (P(GE)), as indicated on the panels. The dataset is based on 500 simulations. The pink lines show the results when considering observations in Folldal only (i.e. white squares in Fig. 1). The gray lines show the results when considering all observations.

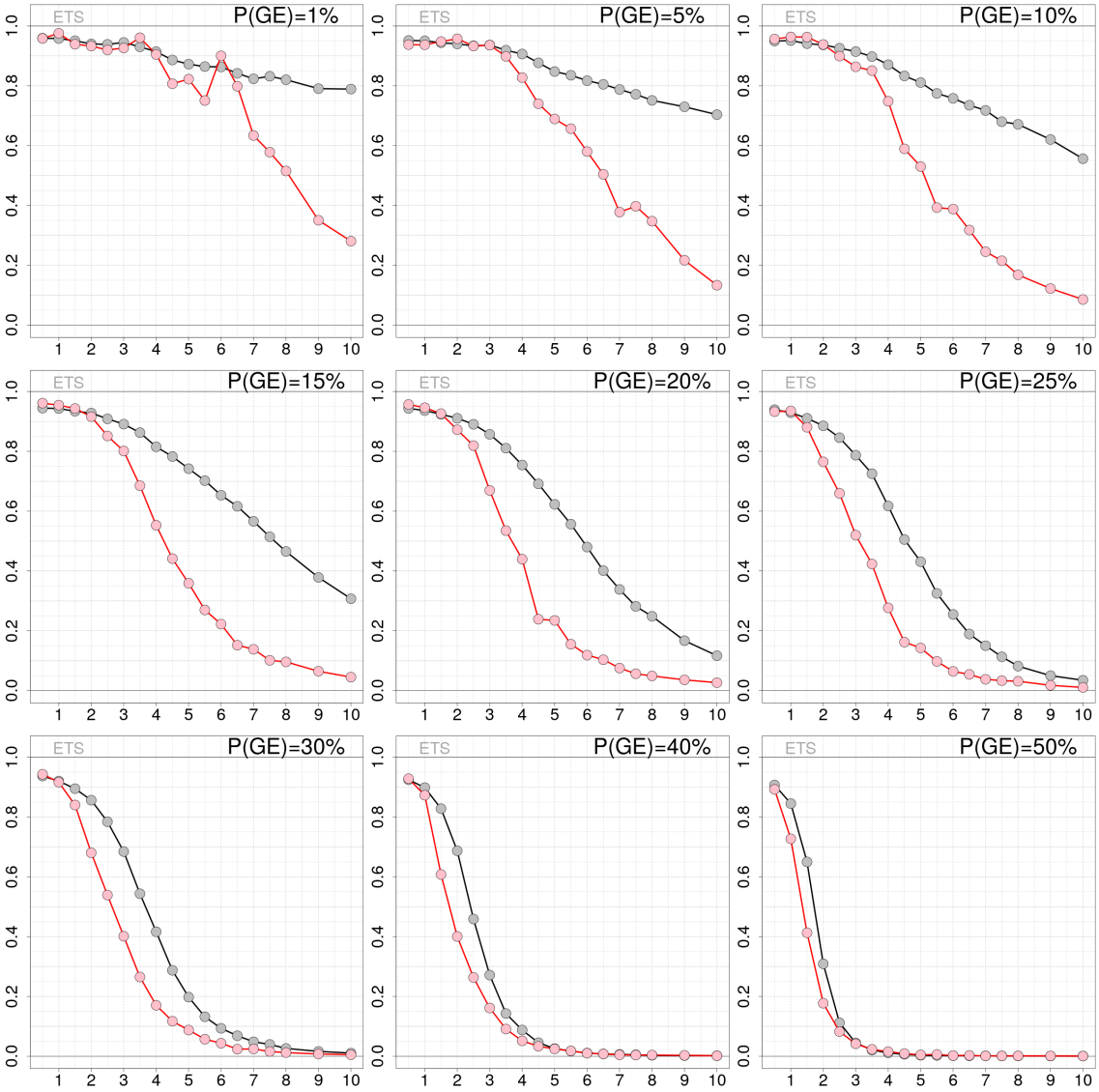


Figure 43: Experiments on synthetic data of temperature generated with $a = 0°C$ (see Fig. 1). Equitable threat score (ETS) as a function of the SCT threshold for different values of the prior probability of GE occurrence (P(GE)), as indicated on the panels. The dataset is based on 500 simulations. The pink lines show the results when considering observations in Folldal only (i.e. white squares in Fig. 1). The gray lines show the results when considering all observations.

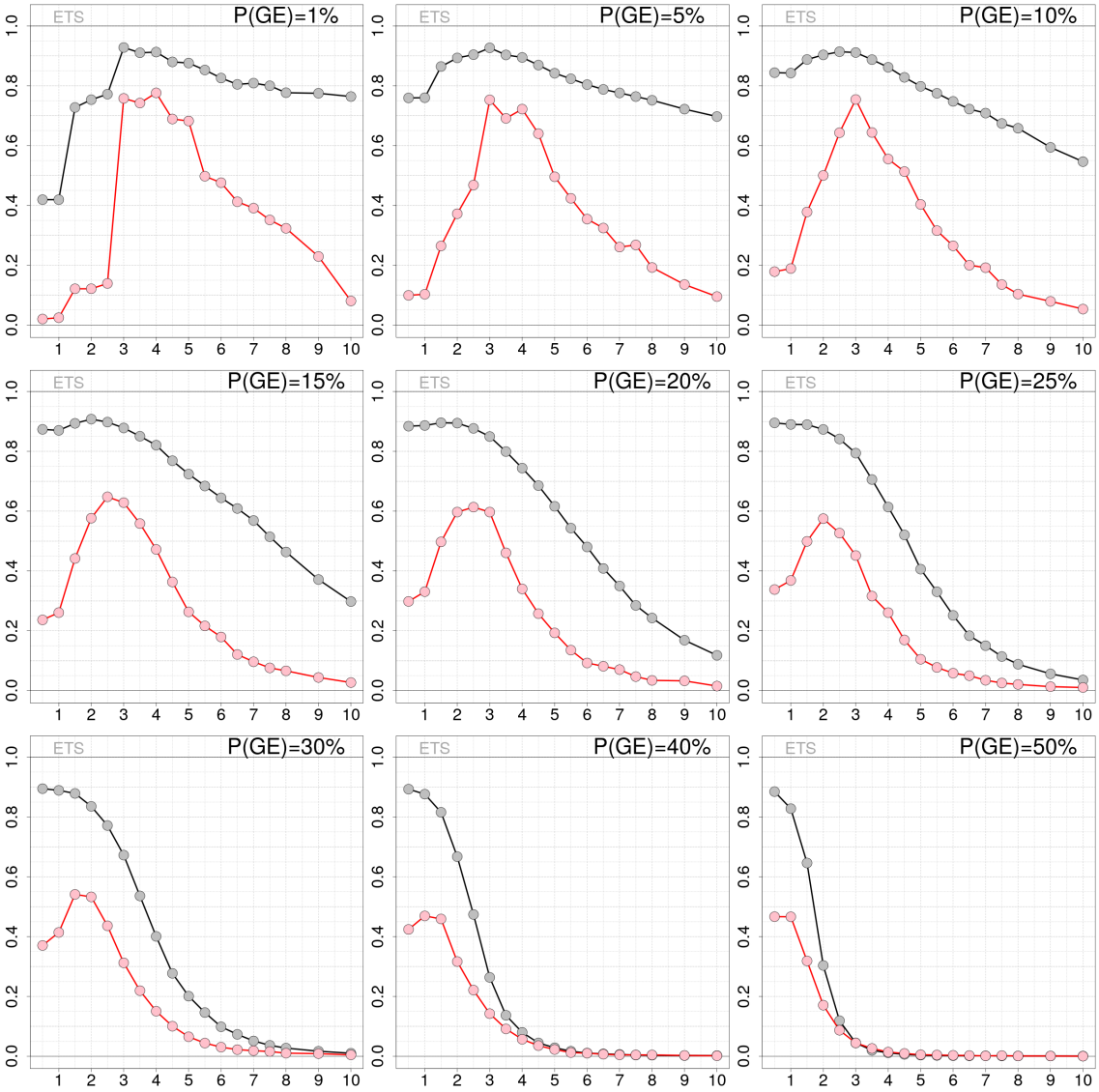


Figure 44: Experiments on synthetic data of temperature generated with $a = 4°C$ (see Fig. 1). Equitable threat score (ETS) as a function of the SCT threshold for different values of the prior probability of GE occurrence (P(GE)), as indicated on the panels. The dataset is based on 500 simulations. The pink lines show the results when considering observations in Folldal only (i.e. white squares in Fig. 1). The gray lines show the results when considering all observations.

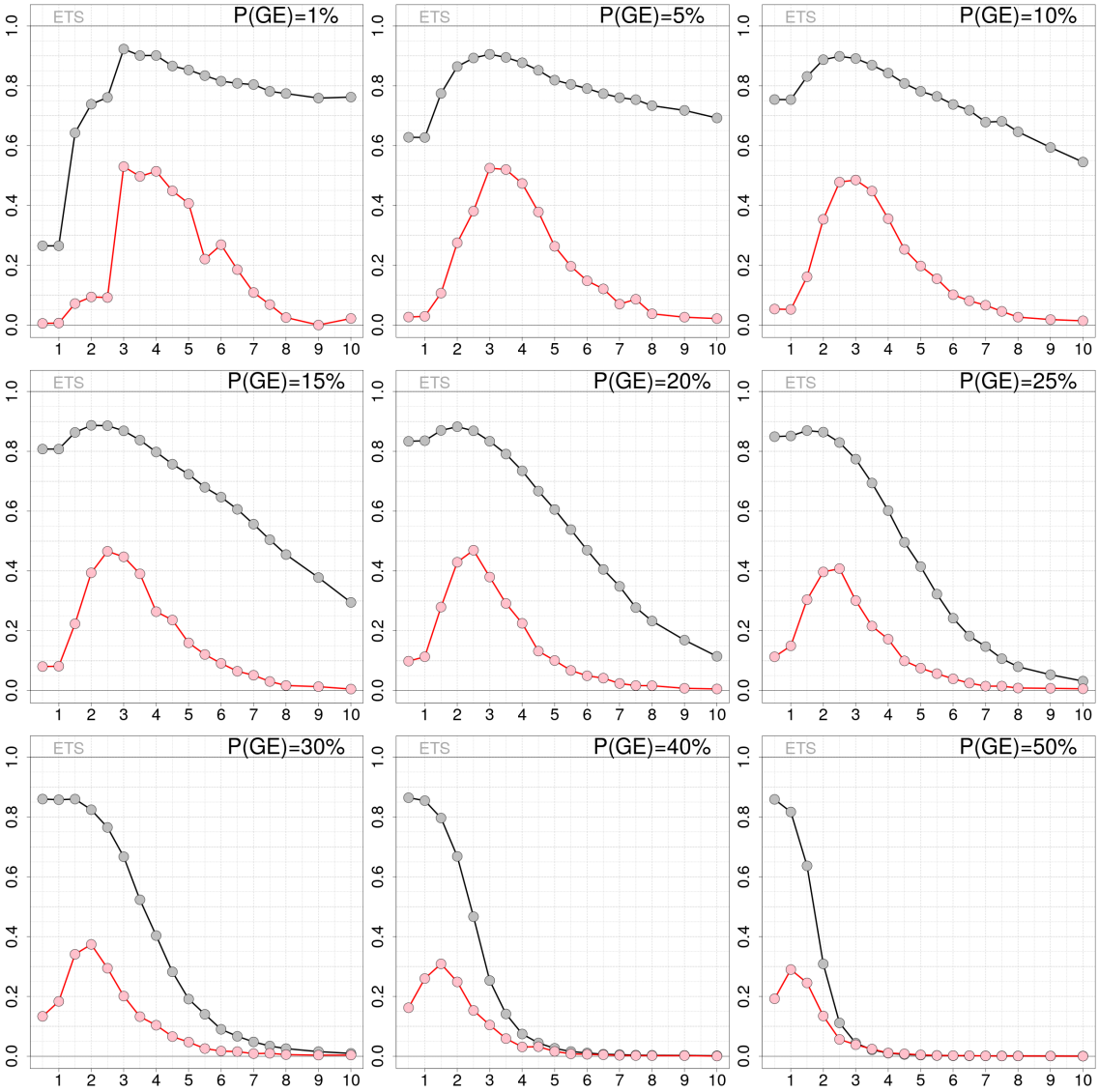


Figure 45: Experiments on synthetic data of temperature generated with $a = 9°C$ (see Fig. 1). Equitable threat score (ETS) as a function of the SCT threshold for different values of the prior probability of GE occurrence (P(GE)), as indicated on the panels. The dataset is based on 500 simulations. The pink lines show the results when considering observations in Folldal only (i.e. white squares in Fig. 1). The gray lines show the results when considering all observations.

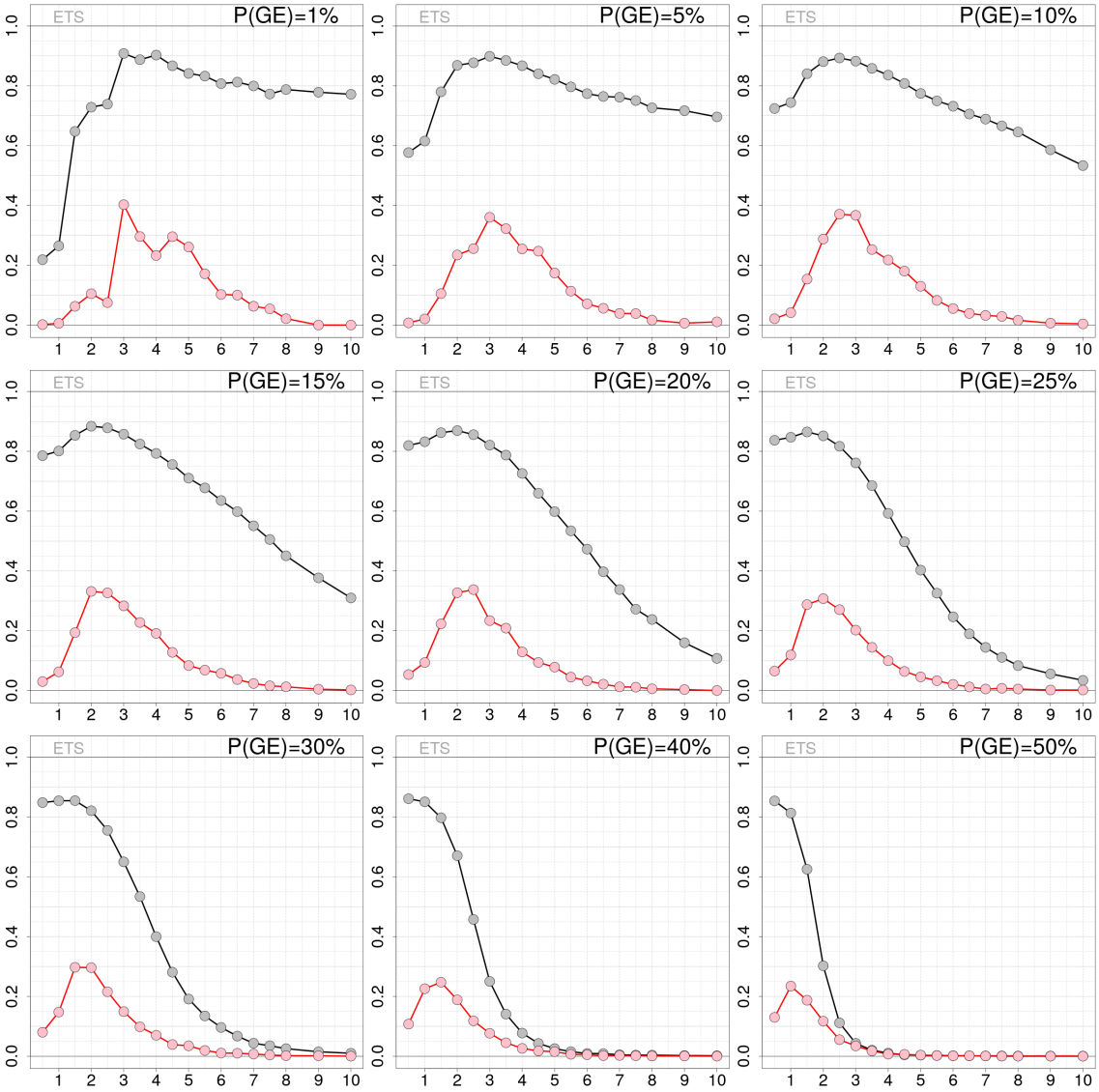


Figure 46: Experiments on synthetic data of temperature generated with $a = 14°C$ (see Fig. 1). Equitable threat score (ETS) as a function of the SCT threshold for different values of the prior probability of GE occurrence (P(GE)), as indicated on the panels. The dataset is based on 500 simulations. The pink lines show the results when considering observations in Folldal only (i.e. white squares in Fig. 1). The gray lines show the results when considering all observations.

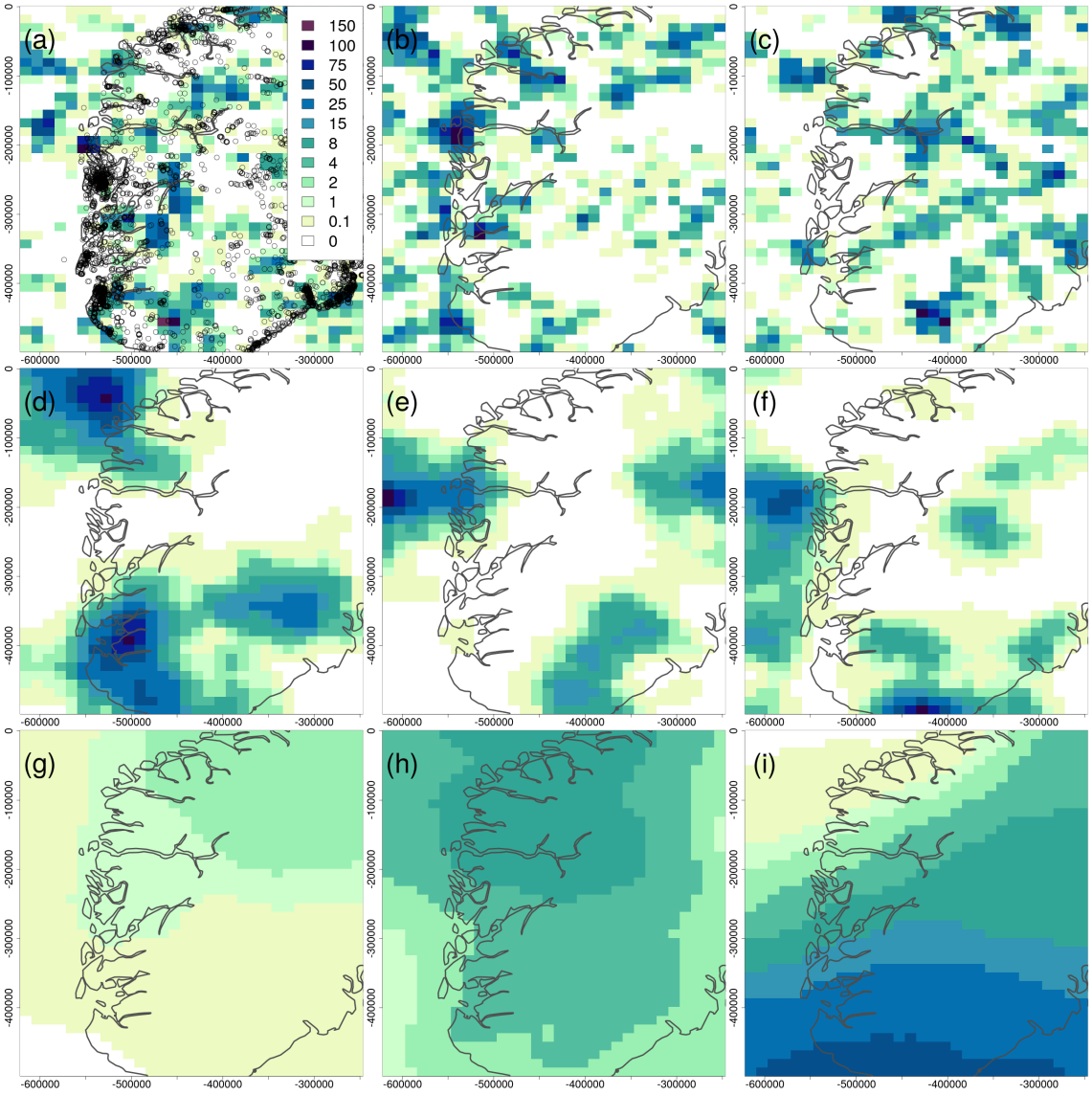


Figure 47: Precipitation synthetic gridded fields created with a stochastic weather generator (similar to Fig. 7 but for $L = 50$ km and $L = 400$ km). All values are in mm. The panels a, b and c in the top row are generated with a spatial length scale of $L = 10$ km (Eq.(1)). The panels d, e and f in the middle row have $L = 50$ km. The panels g, h and i in the bottom row have $L = 400$ km. In panel a, the actual observational network of in-situ weather stations available at MET Norway in July 2020 is shown (black dots). The Gaussian anamorphosis parameters are Gamma shape=0.1328493, rate=0.0285855. The units of easting and northing coordinates are metres.

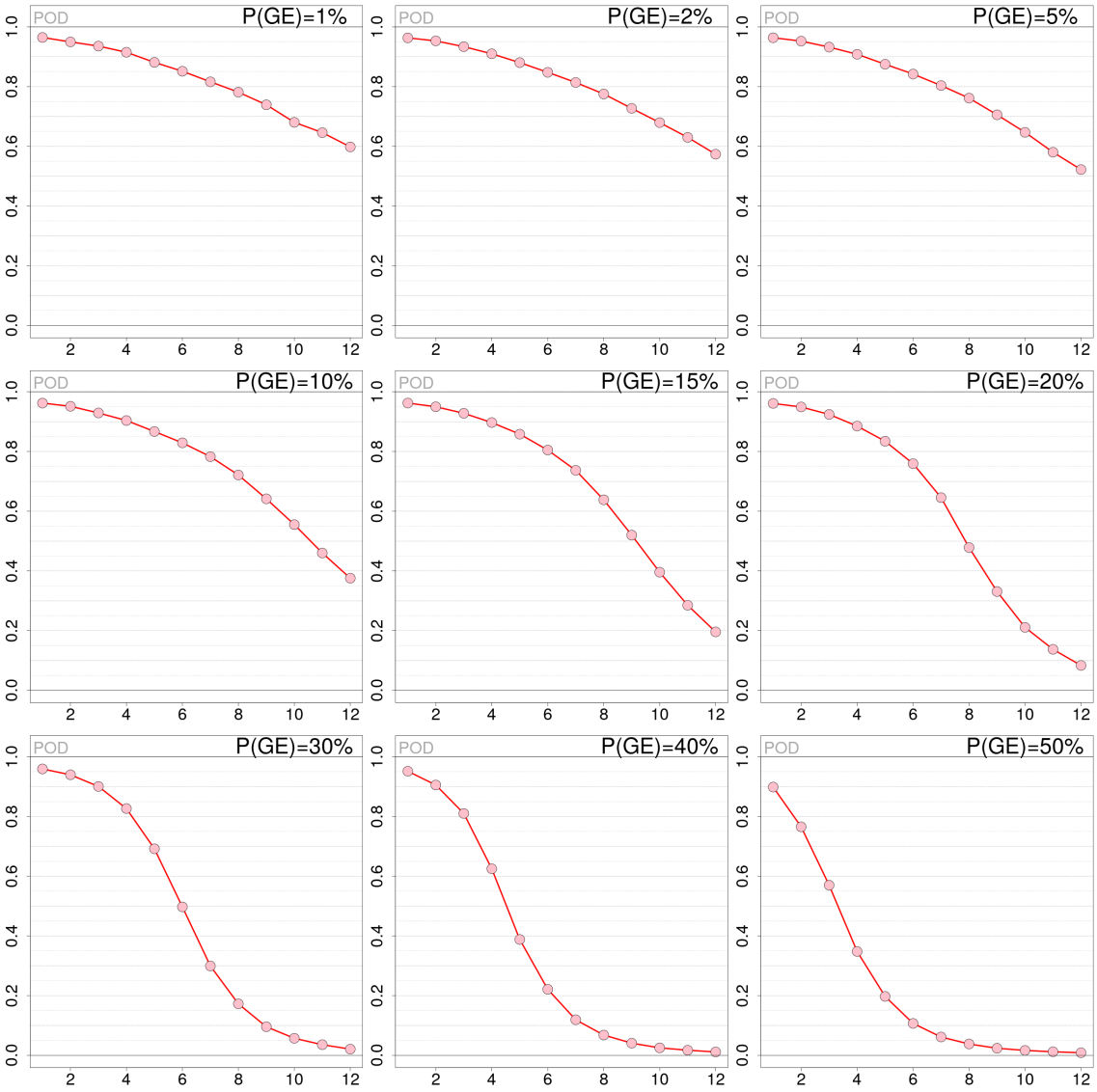


Figure 48: Experiments on synthetic precipitation data produced with $L = 10$ km. Probability of detection (POD) as a function of the SCT threshold T for different values of the predefined probability of gross error occurrence, which are indicated on the top right corner of each panel. The dataset analyzed is based on 100 simulated fields for each threshold.

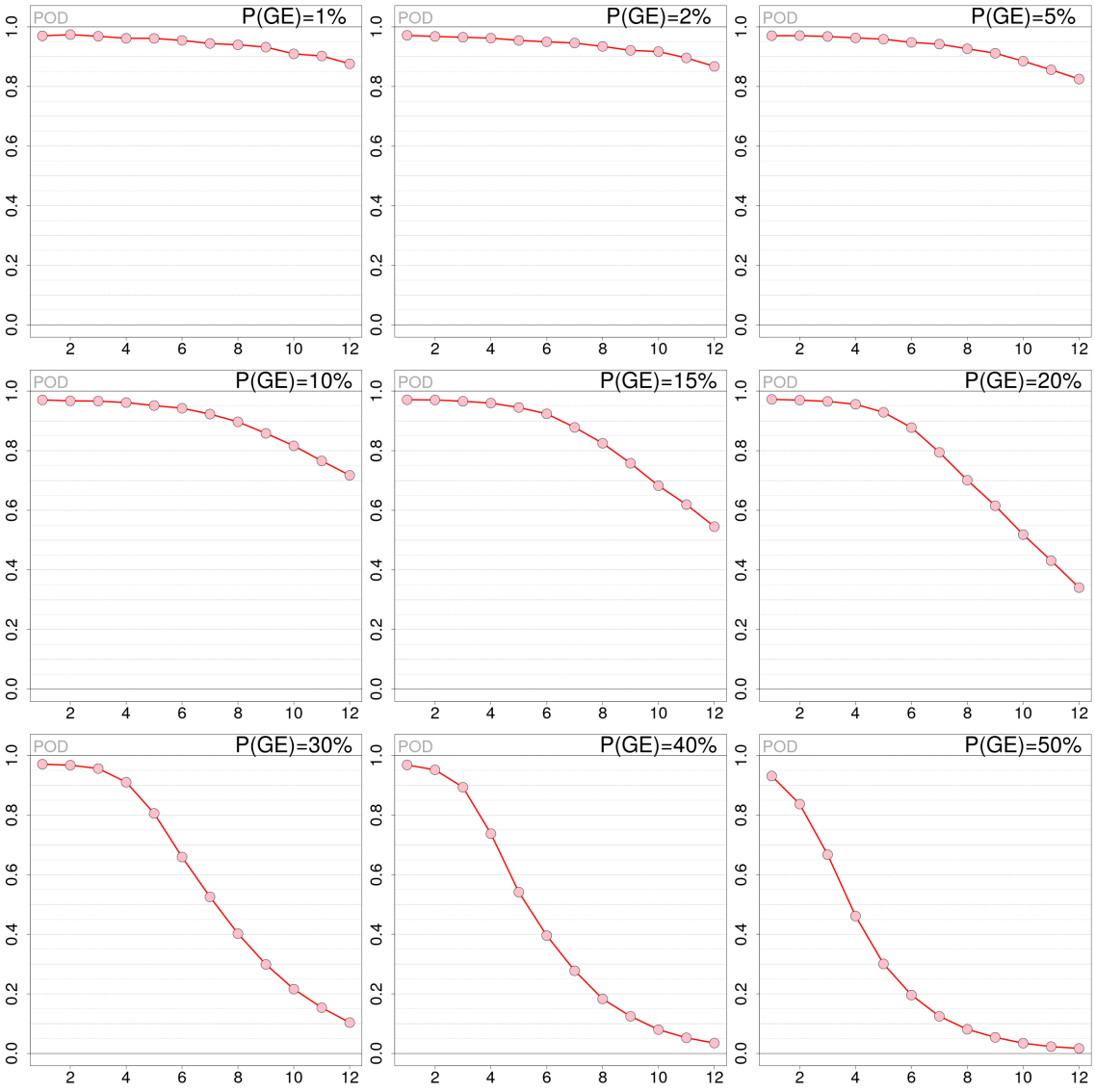


Figure 49: Experiments on synthetic precipitation data produced with $L = 100$ km. Probability of detection (POD) as a function of the SCT threshold T for different values of the predefined probability of gross error occurrence, which are indicated on the top right corner of each panel. The dataset analyzed is based on 100 simulated fields.

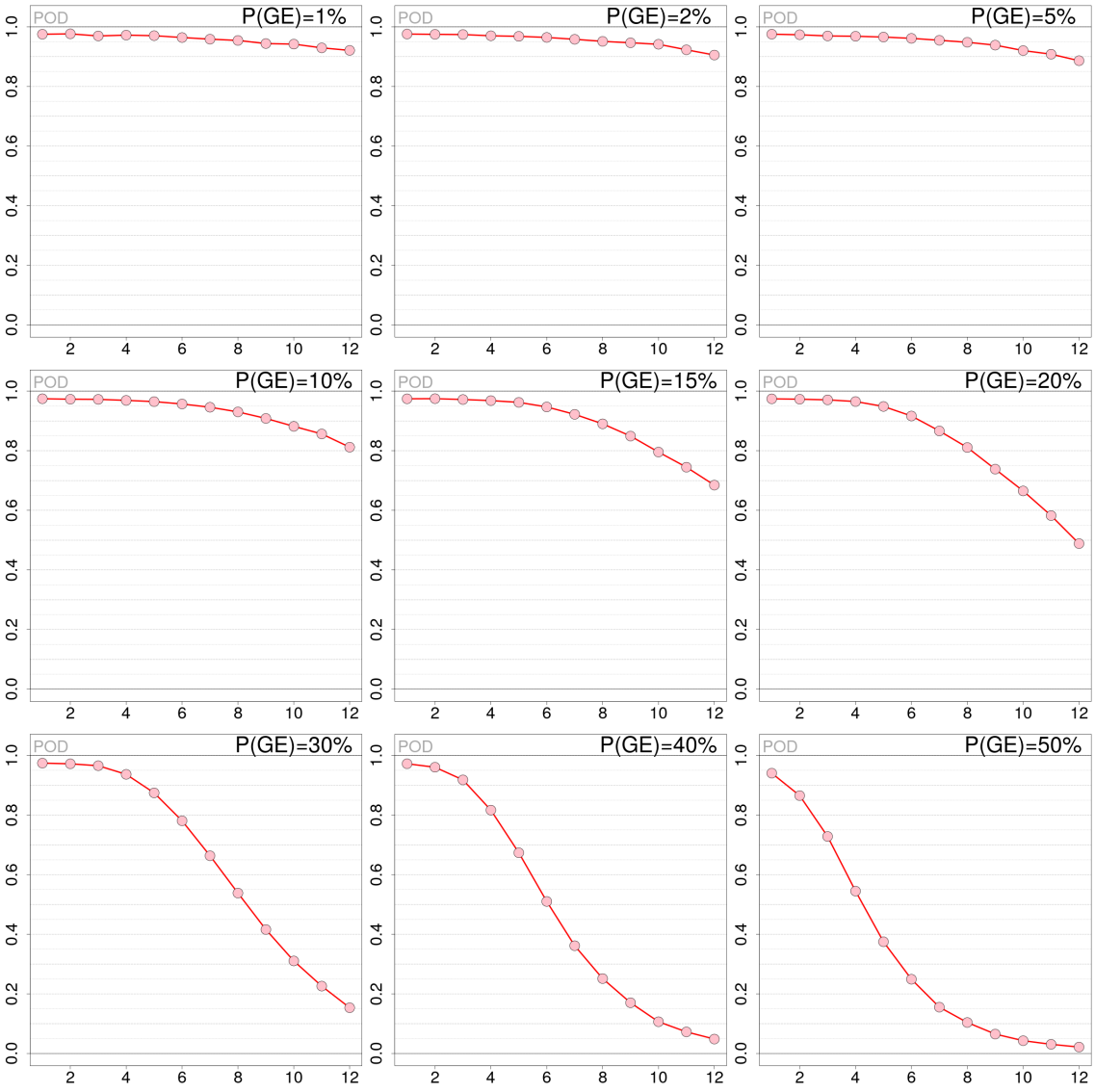


Figure 50: Experiments on synthetic precipitation data produced with $L = 200$ km. Probability of detection (POD) as a function of the SCT threshold T for different values of the predefined probability of gross error occurrence, which are indicated on the top right corner of each panel. The dataset analyzed is based on 100 simulated fields.

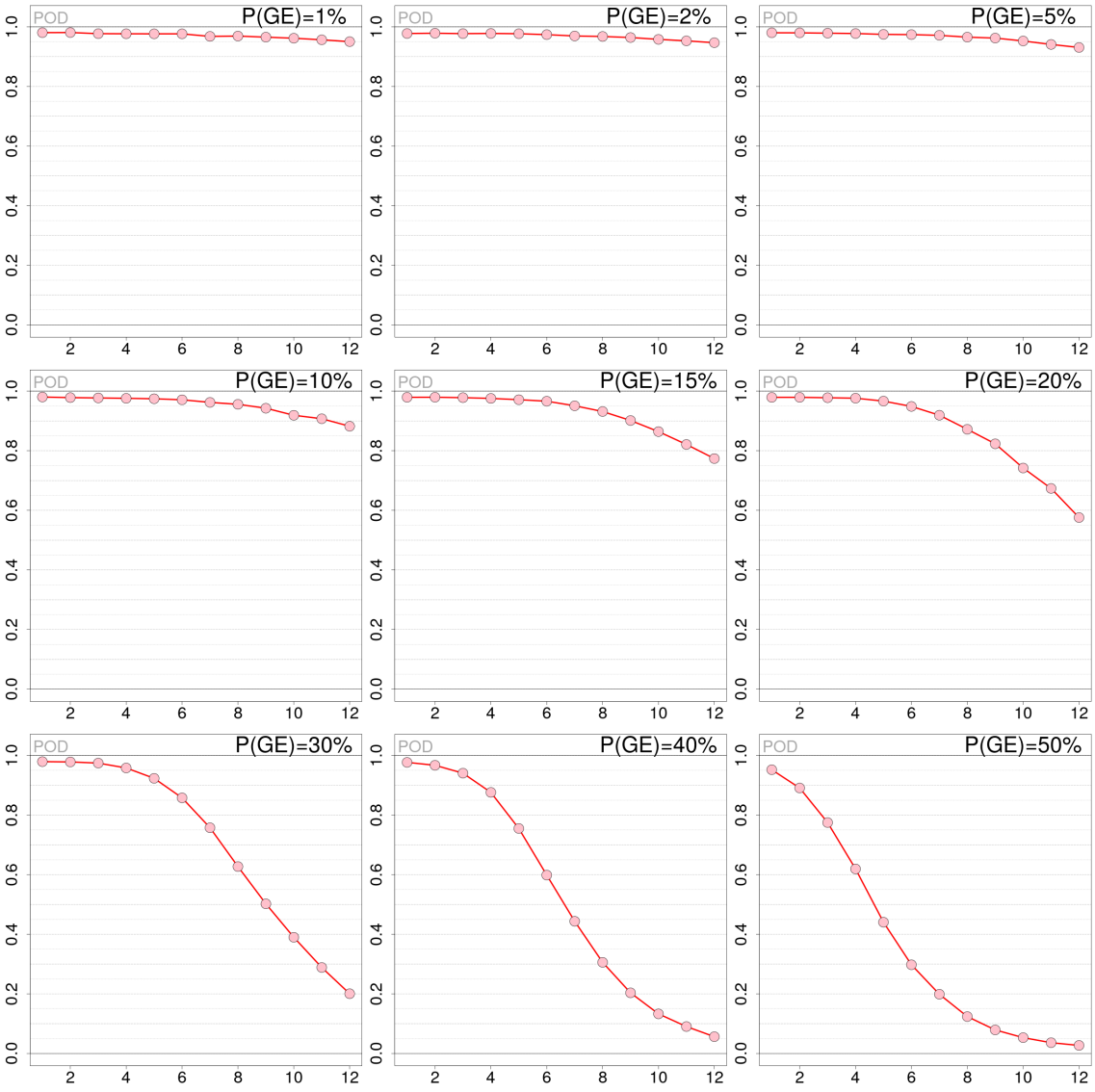


Figure 51: Experiments on synthetic precipitation data produced with $L = 400$ km. Probability of detection (POD) as a function of the SCT threshold T for different values of the predefined probability of gross error occurrence, which are indicated on the top right corner of each panel. The dataset analyzed is based on 100 simulated fields.

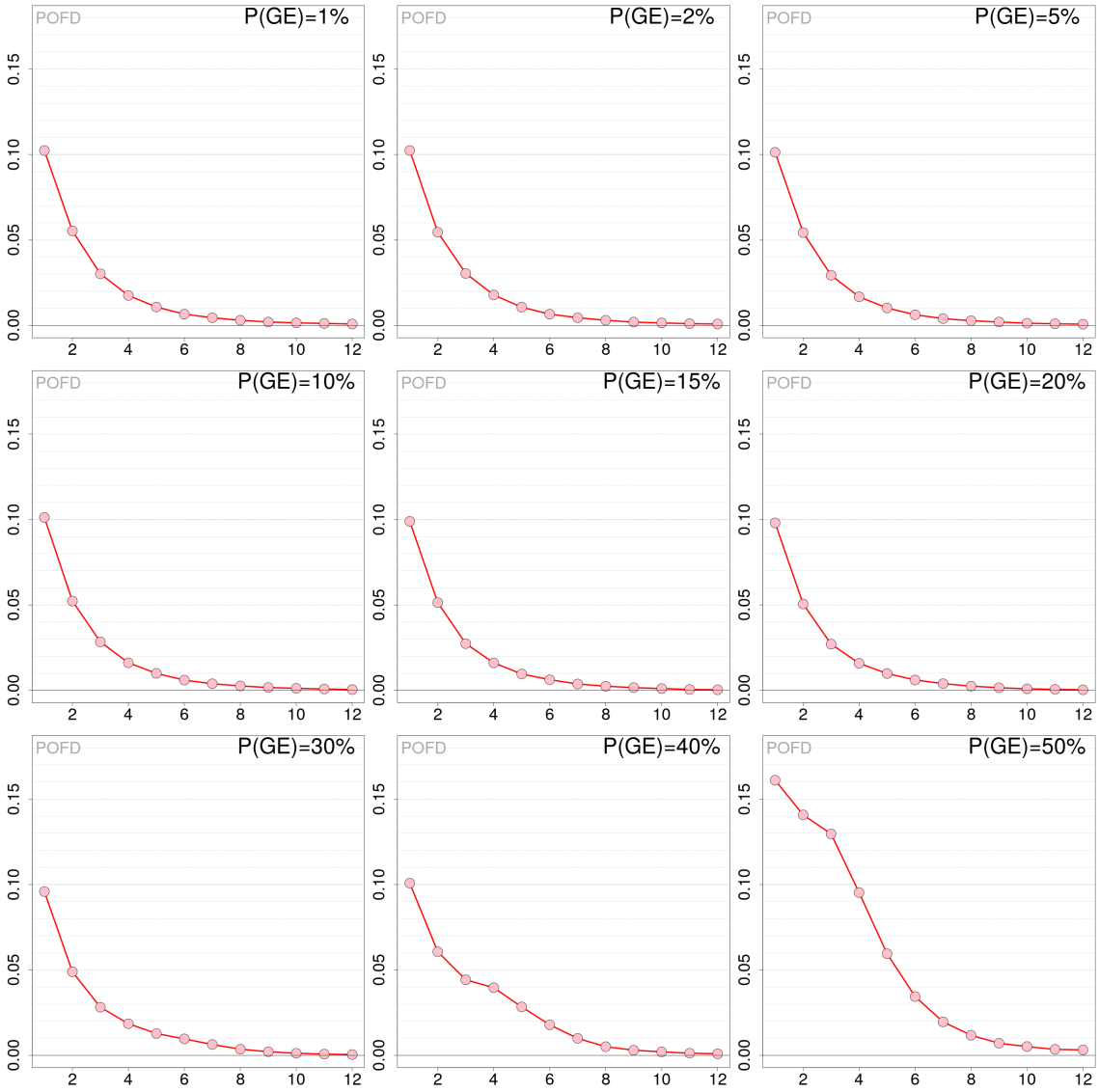


Figure 52: Experiments on synthetic precipitation data produced with $L = 10$ km. Probability of detection (POD) as a function of the SCT threshold T for different values of the predefined probability of gross error occurrence, which are indicated on the top right corner of each panel. The dataset analyzed is based on 100 simulated fields for each threshold.

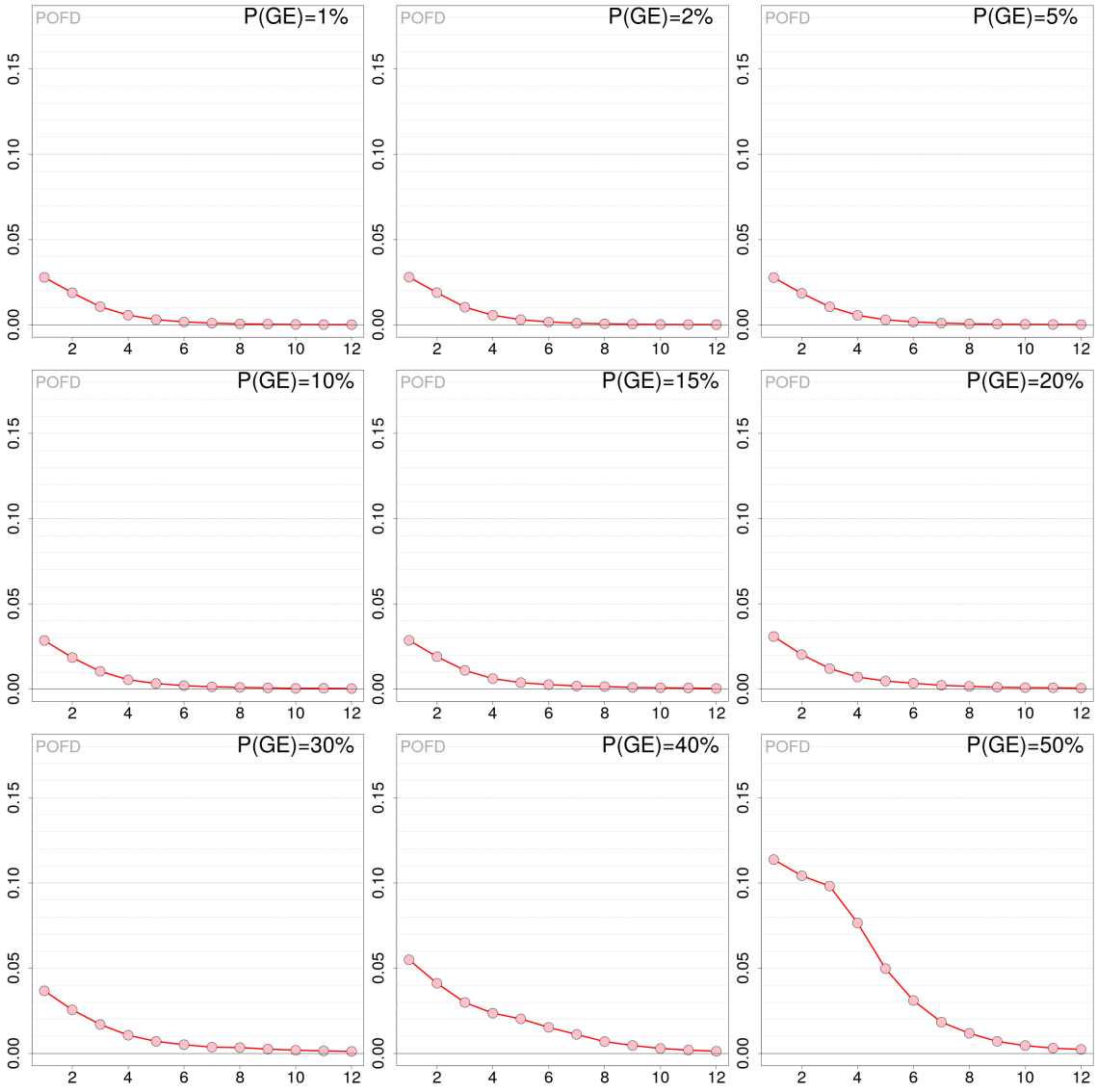


Figure 53: Experiments on synthetic precipitation data produced with $L = 100$ km. Probability of detection (POD) as a function of the SCT threshold T for different values of the predefined probability of gross error occurrence, which are indicated on the top right corner of each panel. The dataset analyzed is based on 100 simulated fields.

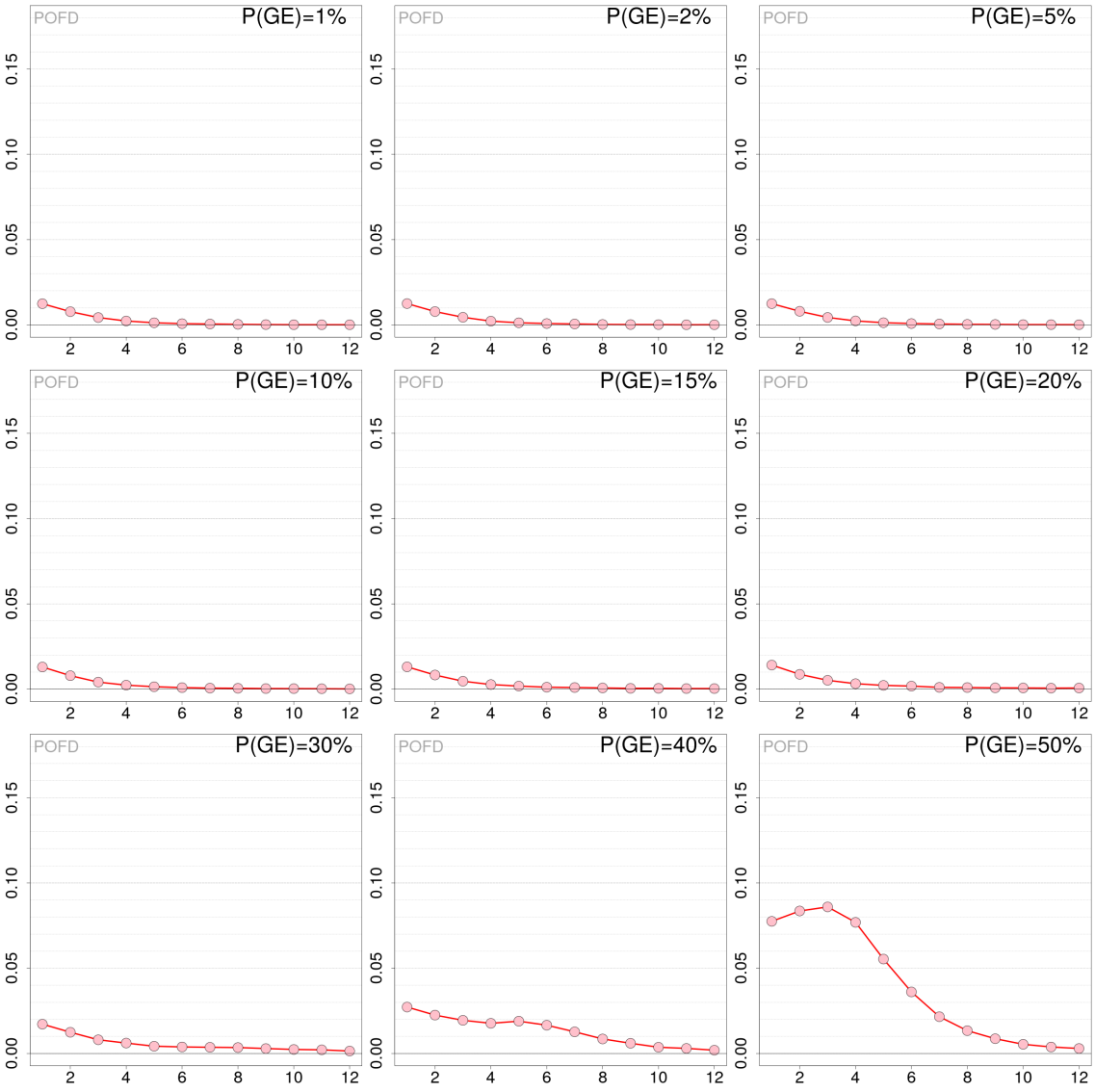


Figure 54: Experiments on synthetic precipitation data produced with $L = 200$ km. Probability of detection (POD) as a function of the SCT threshold T for different values of the predefined probability of gross error occurrence, which are indicated on the top right corner of each panel. The dataset analyzed is based on 100 simulated fields.

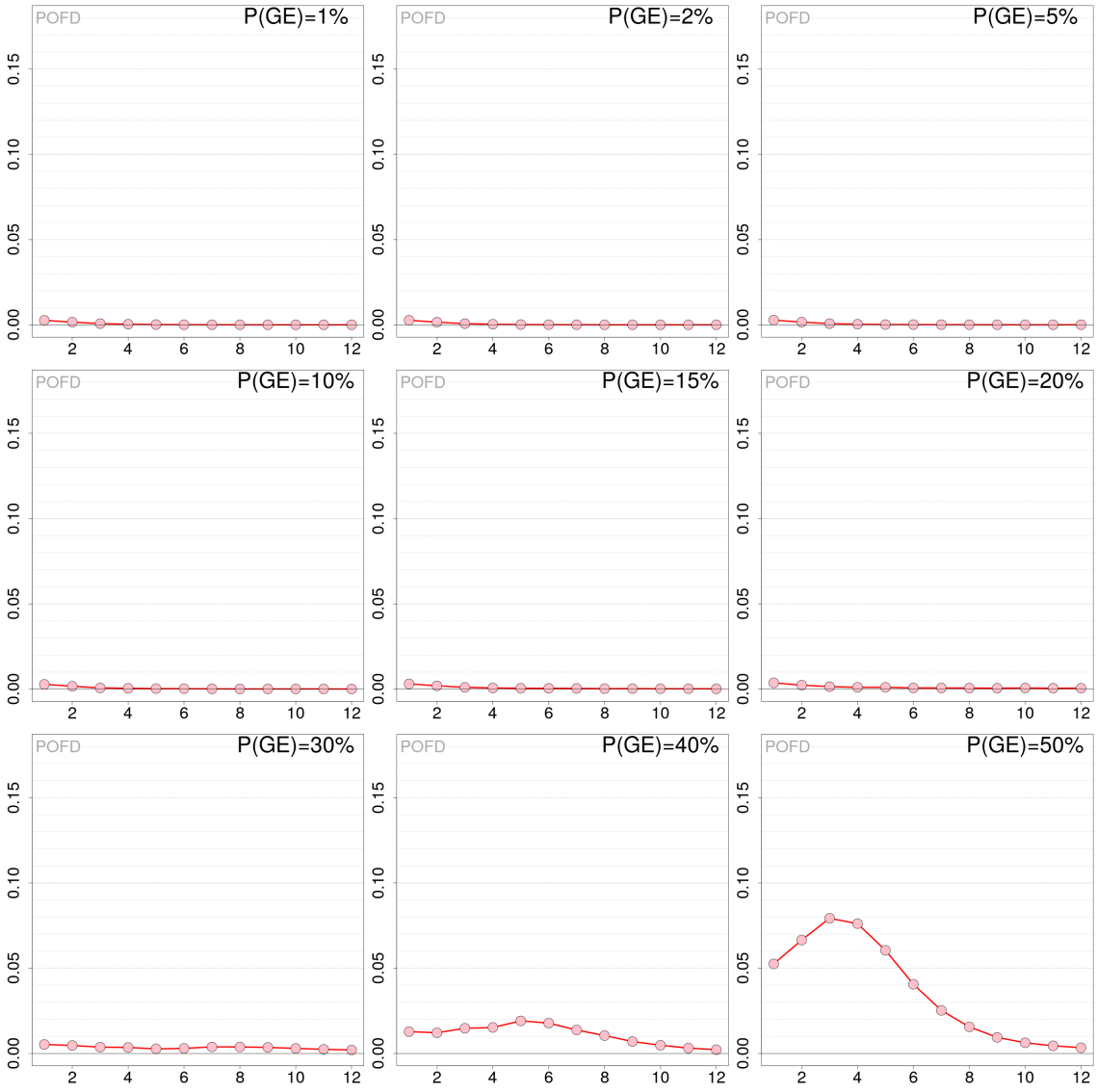


Figure 55: Experiments on synthetic precipitation data produced with $L = 400$ km. Probability of detection (POD) as a function of the SCT threshold T for different values of the predefined probability of gross error occurrence, which are indicated on the top right corner of each panel. The dataset analyzed is based on 100 simulated fields.

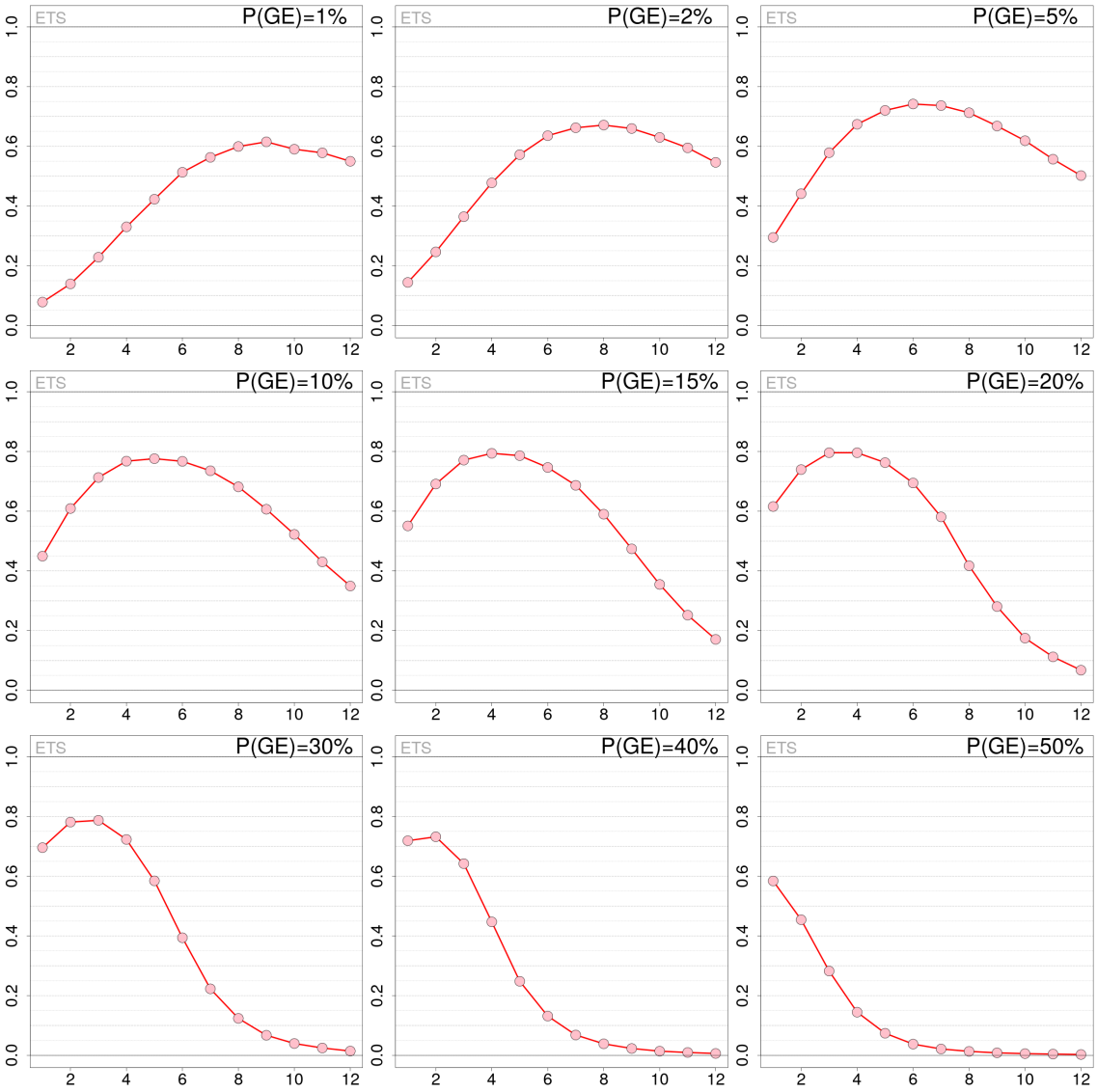


Figure 56: Experiments on synthetic precipitation data produced with $L = 10$ km. Equitable threat score (ETS) as a function of the SCT threshold T for different values of the predefined probability of gross error occurrence, which are indicated on the top right corner of each panel. The dataset analyzed is based on 100 simulated fields for each threshold.

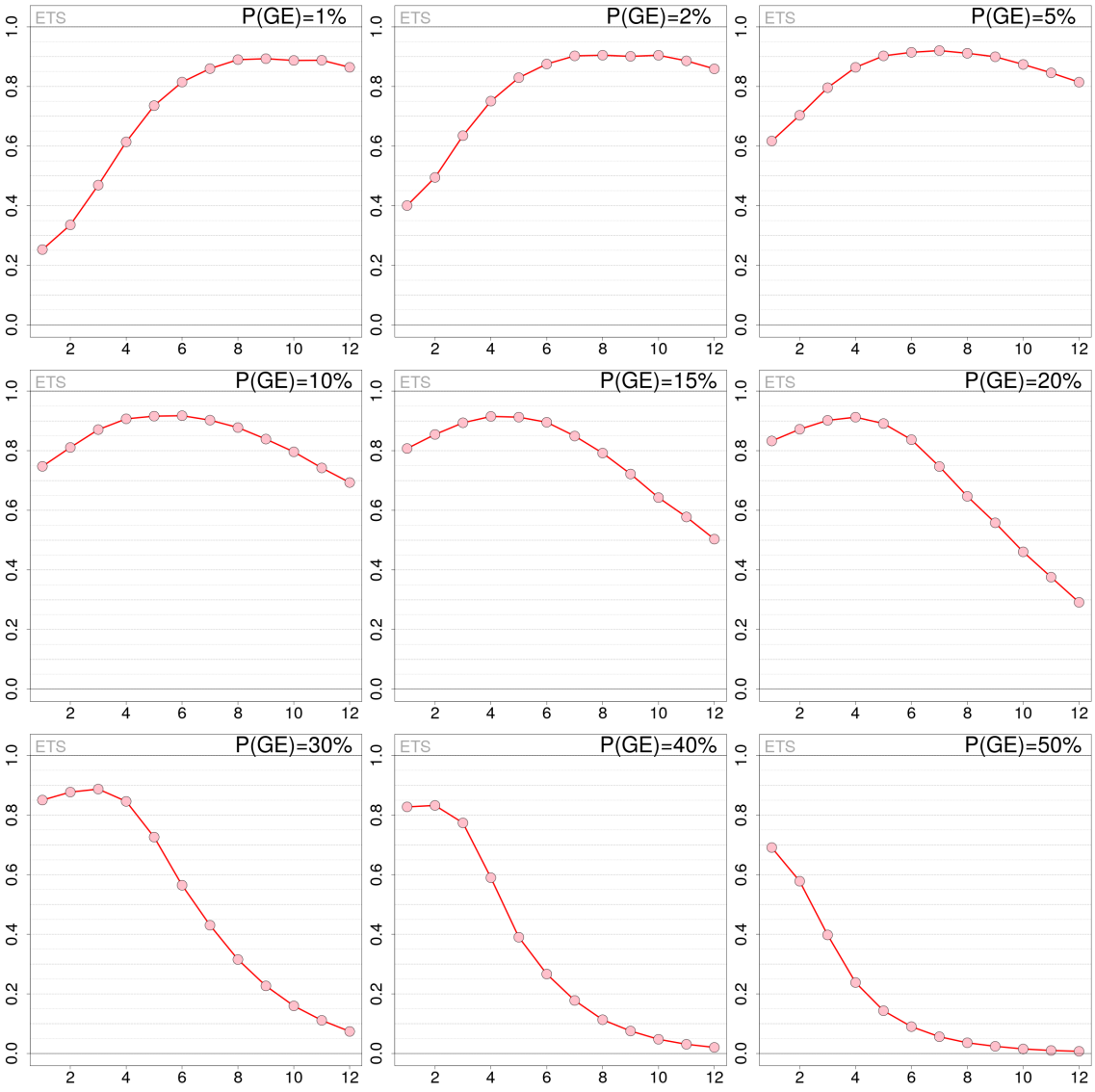


Figure 57: Experiments on synthetic precipitation data produced with $L = 100$ km. Equitable threat score (ETS) as a function of the SCT threshold T for different values of the predefined probability of gross error occurrence, which are indicated on the top right corner of each panel. The dataset analyzed is based on 100 simulated fields.

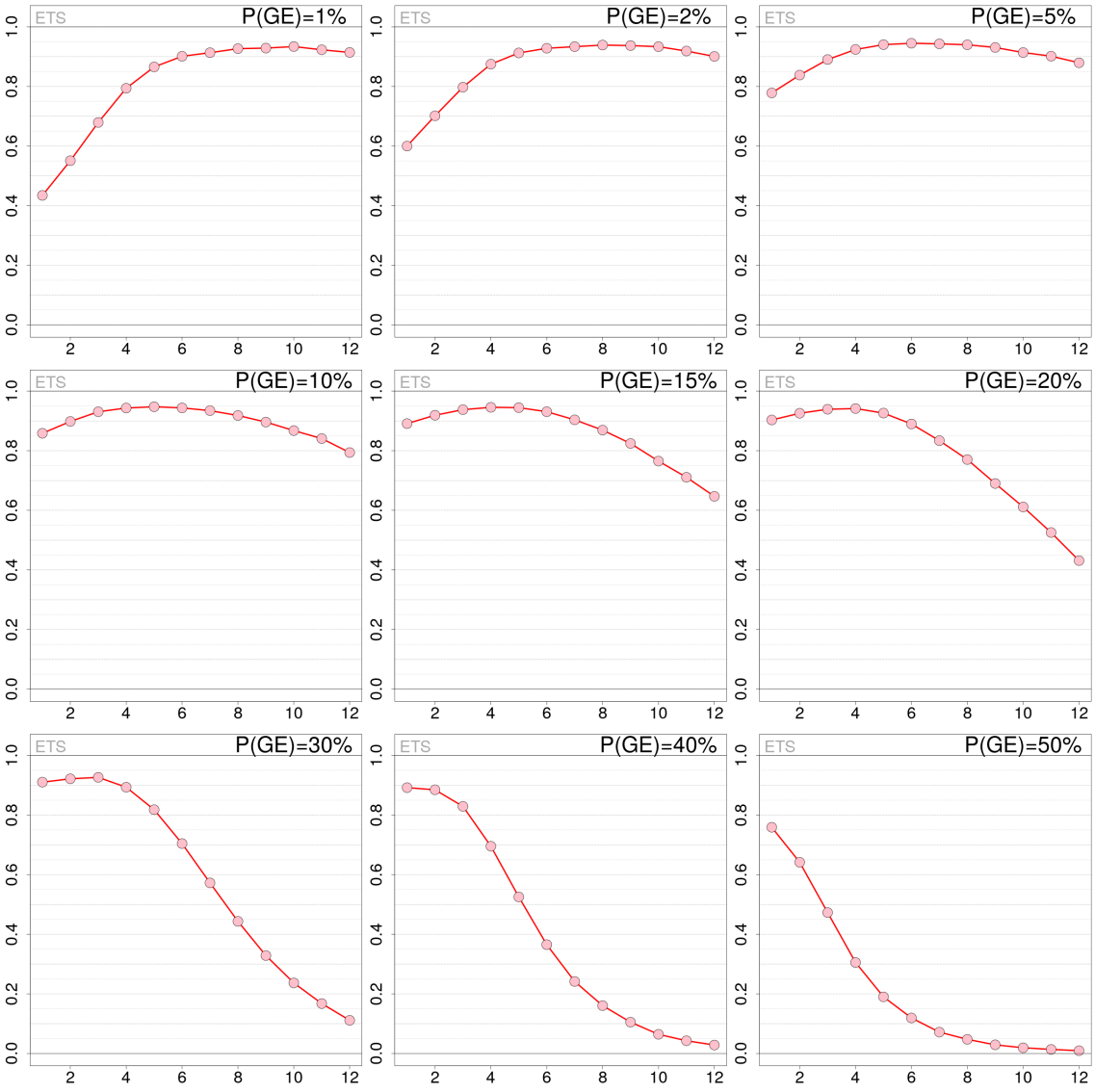


Figure 58: Experiments on synthetic precipitation data produced with $L = 200$ km. Equitable threat score (ETS) as a function of the SCT threshold T for different values of the predefined probability of gross error occurrence, which are indicated on the top right corner of each panel. The dataset analyzed is based on 100 simulated fields.

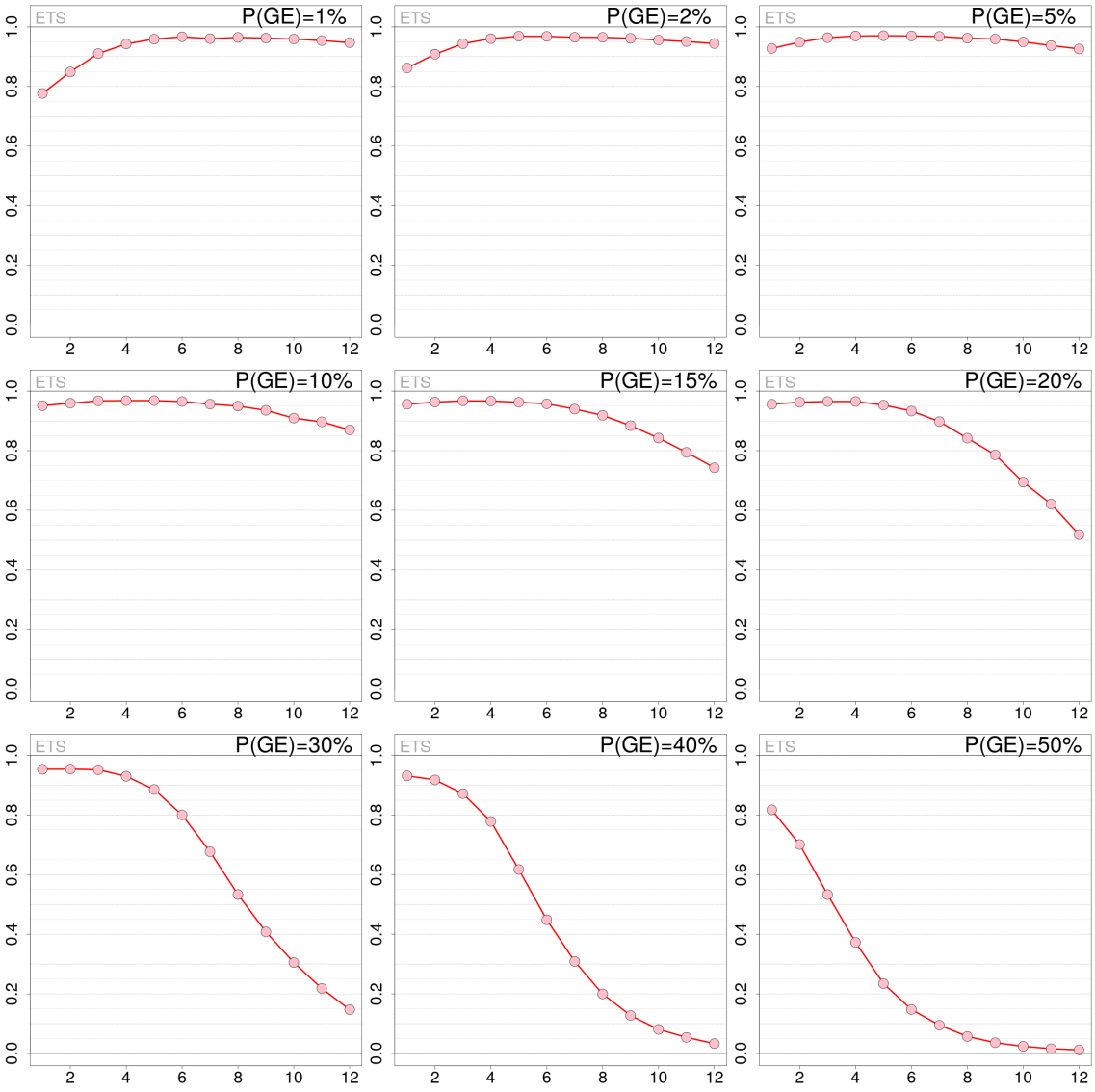


Figure 59: Experiments on synthetic precipitation data produced with $L = 400$ km. Equitable threat score (ETS) as a function of the SCT threshold T for different values of the predefined probability of gross error occurrence, which are indicated on the top right corner of each panel. The dataset analyzed is based on 100 simulated fields.

References

Frei, C. (2014), Interpolation of temperature in a mountainous region using nonlinear profiles and non-euclidean distances, *International Journal of Climatology*, *34*(5), 1585–1605.

Gaspari, G., and S. E. Cohn (1999), Construction of correlation functions in two and three dimensions, *Quarterly Journal of the Royal Meteorological Society*, *125*(554), 723–757.

Lanzante, J. R. (1996), Resistant, robust and non-parametric techniques for the analysis of climate data: theory and examples, including applications to historical radiosonde station data, *International Journal of Climatology*, *16*(11), 1197–1226.

Lussana, C., and L. Båserud (2021), A spatial consistency test for the quality control of meteorological observations, part i: Methodology, *MET report*, *11/2021*.

Lussana, C., T. N. Nipen, I. A. Seierstad, and C. A. Elo (2021), Ensemble-based statistical interpolation with gaussian anamorphosis for the spatial analysis of precipitation, *Nonlinear Processes in Geophysics*, *28*(1), 61–91, doi:10.5194/npg-28-61-2021.

Mahfouf, J.-F., B. Brasnett, and S. Gagnon (2007), A Canadian precipitation analysis (CaPA) project: description and preliminary results, *Atmosphere-Ocean*, *45*(1), 1–17, doi:10.3137/ao.v450101.

Nipen, T. N., I. A. Seierstad, C. Lussana, J. Kristiansen, and Ø. Hov (2020), Adopting citizen observations in operational weather prediction, *Bulletin of the American Meteorological Society*, *101*(1), E43–E57, doi:10.1175/BAMS-D-18-0237.1.

Soci, C., E. Bazile, F. Besson, and T. Landelius (2016), High-resolution precipitation re-analysis system for climatological purposes, *Tellus A: Dynamic Meteorology and Oceanography*, *68*(1), 29,879, doi:10.3402/tellusa.v68.29879.

Wilks, D. S. (2019), *Statistical methods in the atmospheric sciences (Fourth Edition)*, fourth edition ed., Elsevier, doi:https://doi.org/10.1016/B978-0-12-815823-4.00009-2.

Wolff, M. A., K. Isaksen, A. Petersen-Øverleir, K. Ødemark, T. Reitan, and R. Brækkan (2015), Derivation of a new continuous adjustment function for correcting wind-

induced loss of solid precipitation: results of a norwegian field study, *Hydrology and Earth System Sciences*, *19*(2), 951–967, doi:10.5194/hess-19-951-2015.