Norwegian Meteorological Institute

No. 11/2021
ISSN 2387-4201
Free

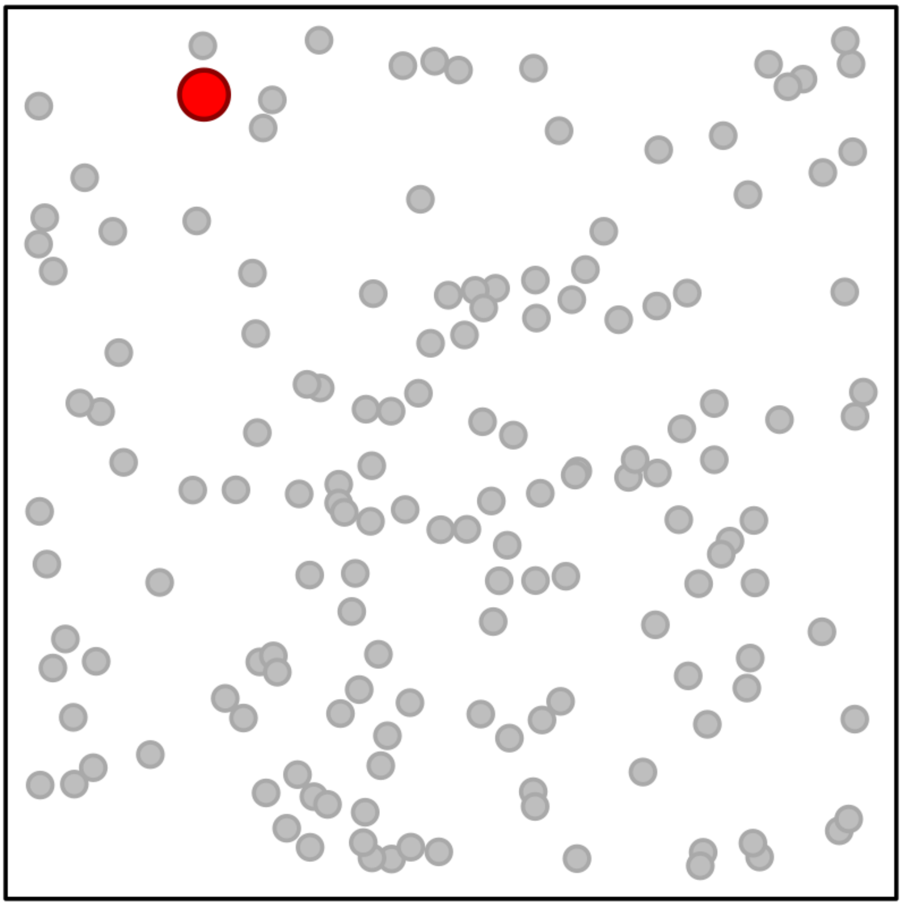
METreport

A spatial consistency test for the quality control of meteorological observations

Part I: Methodology

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 I: Methodology.	October 22, 2021
Section	**Report no.**
Division for Climate Services	11/2021
Author(s)	**Classification**
Cristian Lussana and Line Båserud	● Free ○ Restricted

Abstract

This is the first 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 methodology applied for spatial quality control is described in this document. The second 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 MET Norway.

Keywords

data quality control, in-situ observations, spatial statistics, automatic procedures

Disciplinary signature
Hans Olav Hygen

Responsible signature
Cecilie Stenersen

Abstract

This is the first 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 methodology applied for spatial quality control is described in this document. The second 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 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

Quality control of meteorological data is the *art* of identifying observations that are not representative *enough* of the actual atmospheric state. Observations are values quantifying the state of an atmospheric variable (e.g. two-metre temperature, precipitation), they can be either direct measurements or aggregated values over time and/or space. We say that non representative observations are affected by gross measurements errors, abbreviated as GEs. A reference for the definition of GEs that is relevant for this document is *Gandin* (1988). For convenience, so as not to burden the text with too long sentences, observations affected by GEs are sometimes indicated as *bad* observations. Observations that are representative of the true atmospheric state are *good* observations. The word "art" is used on purpose because quality control, on the one hand, is usually regulated by well-documented procedures but, on the other hand, it relies also on the skill and mastery of expert technicians, observers and scientists. Note that we have deliberately not established strict criteria of accuracy or precision for our definition of quality control, instead we have used the rather generic expression "representative enough", because the tolerable uncertainty of an observation depends on the application. The knowledge required to carry out quality control of observations over a specific region, with sufficient confidence in all situations, is acquired over years of practice and requires a constant updating on new measurement systems. The final judgments on data quality are archived as synthetic codes describing their status. These codes will be called *flags* in this document. Then, statements like ”observations are flagged as good (bad) ones" are used in the text to refer to the action of assigning a judgement to a set of observations.

Considering the growing number of observations collected today in national meteorological and hydrological services (NMHS) around the world, it becomes more and more difficult to apply the art of quality control on all data and within the required time frame. For instance, the production of automatic weather forecasts can benefit from real-time access to all observations and flags. The development of automatic quality control procedures, which is the subject of this document, goes hand in hand with the development of computers and it has forced people to write algorithms and define criteria to assign specific flags to the observations. The application of automatic procedures does not automatically make the process more scientifically based, rather it makes it based on objective criteria, and its results are totally reproducible by anyone using the same programs. In any case, flags have a life of their own, in constant evolution, and can never be considered written in stone, hence also the importance of metadata. In fact, automatic procedures require calibration and optimization of their parameters over datasets encompassing years of measurements. In this sense, automatic procedures may change their judgement on the data depending on when they have been applied to a dataset. In an analogous way, an expert can change opinion on a flag when new information arrives. The real benefit of applying automatic procedures to the massive flow of measurements entering the databases of NMHS is in the systematic testing of all measurements, such that they can be used for any application, even for those working on a real-time basis. The ideal situation to ensure the best possible quality assessment for the observations stored in an archive is in the combination of automatic procedures and decisions from experts. The computer programs take care of screening and validating the bulk of the observations, while the experts are free to carefully evaluate the most important and controversial cases.

The key concept of spatial quality control is the comparison of simultaneous observations of a weather phenomenon taken at different locations in space. The general concepts of the spatial quality control methods that we will consider in the present document have been described in the article by *Båserud et al.* (2020). The spatial methods have been implemented in titanlib `https://github.com/metno/titanlib`. The test considered in this document is the spatial consistency test resistant to outliers in observations, which is abbreviated here as SCT, while the titanlib function is *sct_resistant*().

The SCT is based on inverse problem theory (*Tarantola*, 2005) and builds up onto the SCT presented by *Lussana et al.* (2010). In a nutshell, the test is designed to decide the most likely option between an observation being representative of the atmospheric state or being drawn from a random number generator. When the last option is favored, the observation is flagged as bad. The estimate of the true atmospheric value at a location is made by means of neighbouring observations and it can be made as complex as desired. The innovative part of the presented SCT consists of a better method of identifying observations that are affected by GEs because of a refined distinction between GEs and representativeness errors (REs).

2 Spatial Consistency Test (SCT) resistant to outliers

The SCT algorithm is presented in Algorithm 1. The mathematical notation is reported in Tab. 1. The applied strategy is that of breaking down a large problem into smaller problems by sequentially applying the test across the domain in a moving window fashion. Each window is centered over an observation, which is then called the centroid observation, and defines a sub-region where observations are tested simultaneously. The core of the SCT over a sub-region is the Optimal Interpolation (OI *Gandin and Hardin*, 1965) described in Algorithm 2. If the centroid observation is not isolated, there are two possible outcomes, either all observations tested are good ones or the observation most likely to be bad is flagged as such and the remaining observations do not get a decision on their flags. Note that observations from isolated stations do not get a quality flag because the distinction between RE and GE is deemed as not reliable there.

Several SCT sweeps yield the proposed algorithm. The procedure stops when no bad observations are found. Note that the user is allowed to specify a maximum number of iterations. In this case, the SCT may leave some observations without quality flags, therefore an additional loop is performed using only those observations without flags as centroids. A final round of SCTs is performed using the bad observations as centroids and testing only those observations. The flags may change in this round. It is worth remarking that the opposite can never happen in Algorithm 1, since a good observation cannot be reassigned as bad. This final round is used to prevent the random order in which the test is applied has a decisive influence on the bad flags.

As shown in Fig. 1, the elements defining the sub-region are: the centroid observation; the inner and the outer circles. A centroid observation is isolated when there are no other observations inside the inner circle or the number of observations in the outer circle is less than $p_{\mathrm{out,mn}}$. Once a sub-region has been defined, the observations considered in the test are the good ones and those that have not yet been assigned a quality flag. All observations inside the outer circle (gray dots), up to a predefined maximum of $p_{\mathrm{out,mx}}$, are used for the test, although the observations actually tested are only those inside the inner circle which have not yet received a flag (red dots plus the centroid). In regions where particularly dense observational networks are available, some of the observations in the outer circle may not be used (light gray dots) because they are not among the closest $p_{\mathrm{out,mx}}$ observations. This choice has been made to optimize the computational resources. The outer circle provides the boundary conditions for the spatial analysis used in the

SCT, such that the evaluation inside the inner circle is more reliable. All the observations inside the inner circle (with the exception of the bad ones), whether they are to be tested or not (yellow dots), are used to estimate adaptively the SCT parameters. A score χ (see Algorithm 2) is assigned to each observation tested, similarly to the SCT described by *Lussana et al.* (2010). The SCT-scores quantify the likelihood of having GEs: the higher the score, the more likely is a GE. However, our test, for practical reasons, assumes some simplified hypotheses, such as that GEs are not spatially correlated and this might not be true when significant REs are present in several observations close together. In those cases, a whole group of observations may be flagged as bad, although it is probably correct to assume that all observations are good ones and they are simply measuring the same small-scale phenomenon (i.e. "small" scale with respect to the local observation density). In order to improve the spatial test, in this version we have introduced the spatial outlier detection (SOD) score, $\boldsymbol{z}$. The SCT-score of an observation is compared against the areal statistics of the SCT-score, then an observation is flagged as bad when its SCT-score is an outlier. In this way, neighbouring observations of the same small-scale phenomenon will borrow strength from each other. The thresholds T specified in the setup of SCT (it is worth mentioning again that this is *sct_resistant()* in titanlib) for deciding whether an observation is good or bad are actually thresholds for the SOD-score.

The paragraphs above describe the main steps of Algorithms 1 and 2. In the text that follows, we will focus on some important details of the method.

The observed values and their locations are not the only information needed to perform the SCT. We also need to specify our a-priori level of confidence in each observed value. A-priori means before considering the surrounding observations into our spatial analysis. There are many ways to specify the observation uncertainties, such as by means of observation error variances (*Uboldi et al.*, 2008). For the SCT presented here, we have decided to specify two ranges of values around each of the observed values. The two ranges are: the range of valid values ($\boldsymbol{v}_{\text{mn}}$, $\boldsymbol{v}_{\text{mx}}$) and the range of admissible values ($\boldsymbol{a}_{\text{mn}}$, $\boldsymbol{a}_{\text{mx}}$). The situation is illustrated in Fig. 2 for the cross-section drawn in Fig. 1. The ideas behind the introduction of those two ranges are the following. Even before considering any neighbours or any areal statistics, we may expect that if a predicted (reconstructed) value at a location is close to the observed value, then the observed value is likely to be valid. On the other hand, when the reconstructed value deviates significantly from the observed value, the observation is likely to be bad. The quantification of statements like "close enough to" and "deviates enough from" largely depends on the application and the experience of the people working on it. For example, when dealing with the quality control of temperature, one may assume that an observation is good when the leave-one-out cross-validation (cv-) analysis is within the range of $\pm 1°$C from the observed value, while when the cv-analysis deviates from the observed value for more than 20°C, then there might be issues with that observation. The range of valid values may then be set as $\pm 1°$C, while the range of admissible values to $\pm 20°$C around the observed values. If we consider precipitation, then we are facing a variable where uncertainties follow a multiplicative error model (i.e. the larger the observed value, the larger its uncertainty). In this case, the range of valid values may be set to ± 1 mm for the smaller values and $\pm 10\%$ of the observed values for larger values. The range of acceptable values could be set to $\pm 100\%$ of the observed values, of course taking into account that negative precipitation does not make sense. Note that we can specify different levels of confidence in the observations depending on the instrumentation used to measure a quantity, for example. The observations we trust more should get larger ranges, because we are more confident in their observed values even when they deviate significantly from the cv-analyses.

As written in Algorithm 2, the SCT-score at the generic ith location is:

$$\chi = \sqrt{(y_i^{\mathrm{o}} - \check{y}_i^{\mathrm{a}})(y_i^{\mathrm{o}} - y_i^{\mathrm{a}})} \tag{1}$$

a schematic representation of the relationships between analysis, background and observations is shown in Fig. 3. Compared to the CV-analyses, the analyses stay closer to the observed values (because the analysis uses the ith observations, while the cv-analysis doesn't), and this is especially true for isolated observations where there are fewer nearby observations to ”pull" the analysis away from the observed value. In this way, it is more difficult to flag isolated observations as bad and this is done on purpose since these observations provide information where there is none. In the definition of χ we have included a square-root, such that χ has the same units of the observed variable.

With reference again to Algorithm 2, the estimated areal average of the SCT-score is μ, that is the median of a selection of χ values (all belonging to observations in the inner circle). The estimator of the dispersion is the inter-quartile range σ (IQR, i.e. the difference between the 75th and the 25th percentiles), plus an additional term σ_μ that takes into account the uncertainty in the estimate of the areal average. σ_μ is equal to σ divided by the sample size of the χ values used. This expression is inspired by the law of large numbers, see for instance *Taylor* (1997), its value is inversely proportional to the squared root of the number of observations used and it constitutes a “penalty” term on the uncertainty when only few observations are available. Note that we use median and IQR as operators to obtain the estimates because they are more resistant to ouliers in the sample data, if compared to other estimators, such as the average and the standard deviation (*Lanzante*, 1996).

It might happen that the adaptive estimate of the dispersion σ becomes unrealistically small. For example, in the case of precipitation, if most of the observations report no precipitation, then σ becomes exactly equal to 0 mm and we can flag as bad an observation of 0.1 mm, for example. A safe-check on the SCT-score has been included, just to avoid cases like this. Given a set of observations and the SCT settings, we can roughly estimate the spread of the observed values which is considered acceptable, regardless of the actual observed values. In the case of completely isolated observations, the following equation is valid for the generic ith location (*Lussana et al.* (2010), Eq.(23)):

$$(y_i^{\mathrm{o}} - \check{y}_i^{\mathrm{a}})(y_i^{\mathrm{o}} - y_i^{\mathrm{a}}) = (y_i^{\mathrm{o}} - y_i^{\mathrm{b}})^2 \frac{\varepsilon^2}{1+\varepsilon^2} \tag{2}$$

where ε^2 is the ratio between observation and background error variances (*Uboldi et al.*, 2008). The range of valid values is set to $(v_i^{\mathrm{mx}} - v_i^{\mathrm{mn}})$. Then, in the hypothetical case of: i) isolated observation locations and ii) background values within the range of valid values, an alternative form of the SCT-score at the ith location is:

$$\chi' = \sqrt{\varepsilon^2/(1+\varepsilon^2)}(v_i^{\mathrm{mx}} - v_i^{\mathrm{mn}}) \tag{3}$$

this score does not depend on the observed values and it does not quantify the likelihood of GE, instead it transforms our confidence in the observations (i.e. range of valid values) onto the space of χ values. The definitive σ value is the highest value between the IQRs of χ and χ'. The IQR of χ' is linked to the precision specified for the application. The broader the range of valid values, the higher the IQR of χ'. Or in other words, the more we trust the observations, the smaller the final values of z would be.

Each element of the vector $\mathbf{z}$ of the SOD-scores is obtained as the deviation of the corresponding SCT-score from its areal average, normalized by the dispersion. It is written as:

$$\mathbf{z} = \frac{\chi - \mu}{\sigma + \sigma_\mu} \tag{4}$$

Figure 4 shows the graph of CV-analysis residuals against the analysis residuals for just one of the many applications of the SCT (i.e. one of the circular regions moving around through the domain where the SCT is applied), a similar graph has been presented in the paper by *Lussana et al.* (2010) and its shape resembles that of an hourglass, hence the nickname used in the caption of the figure. The figure intends to provide a graphic complement to Algorithm 2. The gray dots lie between the 1:1 line and the blue line, with slope $1+1/\varepsilon^2$. Redundant observations lie on the diagonal line, where the analysis and the CV-analysis are equivalent. Because of the OI scheme implemented, completely isolated observations lie on the blue line. The curve of constant χ^2 values are shown as dashed gray lines. Observations on the same curve have the same quality, in the sense that they will have the same value of the SOD-score. Note that the CV-analysis residuals of isolated observations are much higher than those of the redundant observations with the same quality. Once again, this emphasizes the value of isolated observations as valuable sources of information, given that for isolated stations the test is more tolerant. The curve of the average data quality, corresponding to $\chi = \mu$, is shown as the dashed black line. The dashed red line depends on the dispersion of the gray dots and it marks the boundary between good and bad observations, as indicated in the figure. Only those observations that lie above the red line are considered bad observations.

3 Outlooks and future plans

The SCT presented aims at flagging the bad observations even when they constitute a significant part of the dataset to check. In order to do that, the SCT has been modified and it has been made less sensitive (or more resistant) to misbehaviour of the data than the one presented by *Båserud et al.* (2020); *Lussana et al.* (2010). The solutions adopted includes: taking into account the local spread of the SCT-scores, through the definition of the SOD-score, therefore making the distinction between GEs and REs less affected by local weather phenomena observed by few stations; the use of more robust and resistant statistics, such as median and interquantile range instead of mean and variance. Furthermore, the algorithm has been revised to ensure that the order in which the bad observations are flagged does not have impacts on the final results.

The future plans are to include the spatial checks in the quality assurance system of MET Norway for in-situ observations. This step is not straightforward and it would require -for instance- some work to develop tools that make the titanlib tests ready to read and write from the operational databases.

The way the background in the outer circle is estimated plays an important role in the algorithm, both for temperature -where the elevation plays a key role- and for the other variables. The idea is to test further methods to obtain the background values, such as implementing a linear regression based on "three-group resistant regression" (*Lanzante*, 1996).

The optimization of the SCT thresholds and parameters is another active area of work. We are thinking of a *titan-tuner*, that is a tool which helps the user to estimate the parameters to use in the SCT.

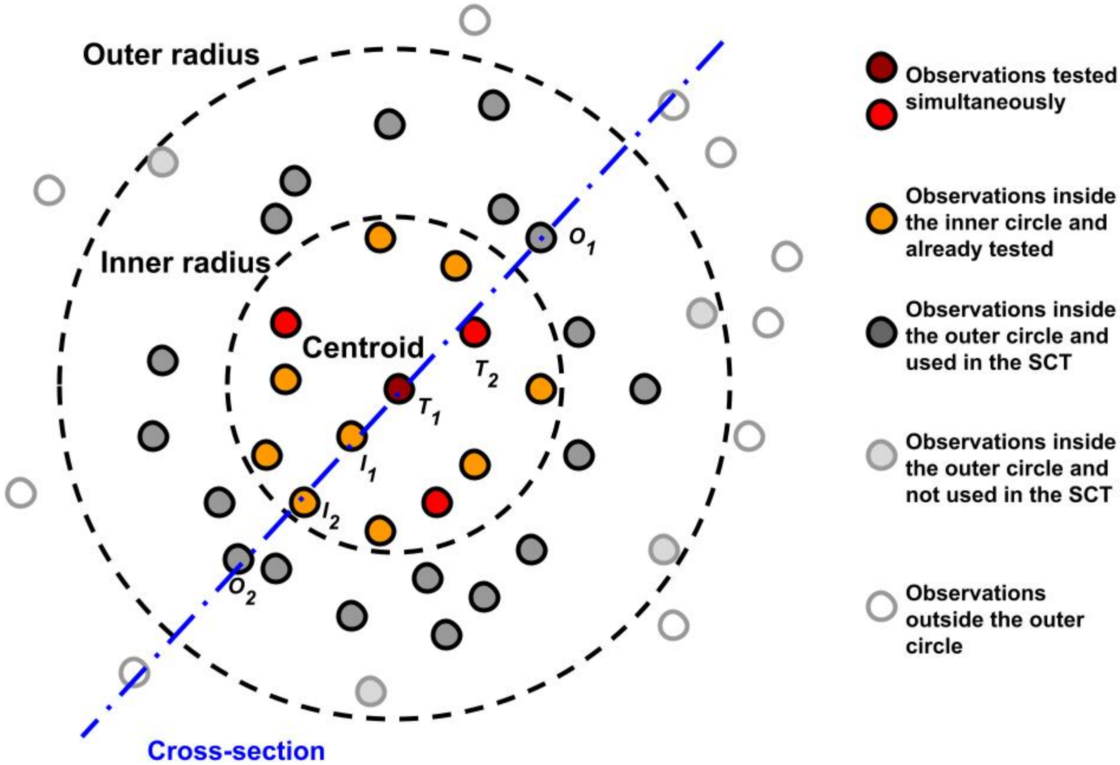


Figure 1: Schematic of the spatial consistency test (SCT) applied in a moving window fashion (inspired by Thomas Nipen). The elements defining the region to test are: the centroid observation (dark red dot), the inner circle (defined by the inner radius r_{in}), the outer circle (defined by the outer radius r_{out}). The meanings of the colors are described in the legend. The dot dashed blue line marks the cross-section shown in Fig. 2 and the observation labels serve as links to the graph in Fig. 2.

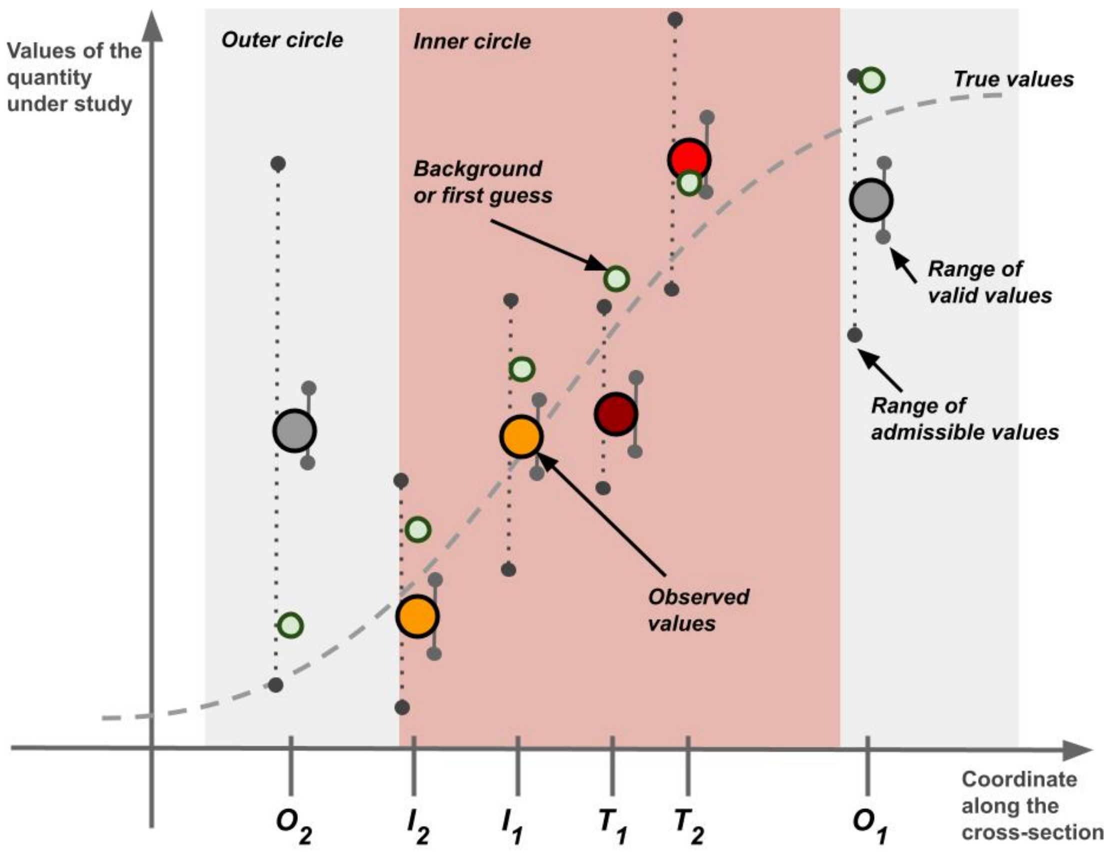


Figure 2: Cross-section drawn in the hypothetical test of Fig. 1 with the observed values and the a-priori information required by the SCT. The labels on the axis of abscissas are the same as in Fig. 1 and they are used to link the observation points in the two figures. The axis of ordinates has the units of the quantity to quality control. The points on the graph are the observed values along the cross-section and the shaded regions identify the inner and outer circles, as clearly indicated on the graph. The observations sample an unknown continuous field of the same quantity, which is shown on the graph as the dashed line of true values. The colors of the dots with the observed values are the same as in Fig. 1. Each of the observed values is shown with two additional pieces of information: the range of valid values ($\boldsymbol{v}_{\mathrm{mn}}$, $\boldsymbol{v}_{\mathrm{mx}}$) and the range of admissible values ($\boldsymbol{a}_{\mathrm{mn}}$, $\boldsymbol{a}_{\mathrm{mx}}$).

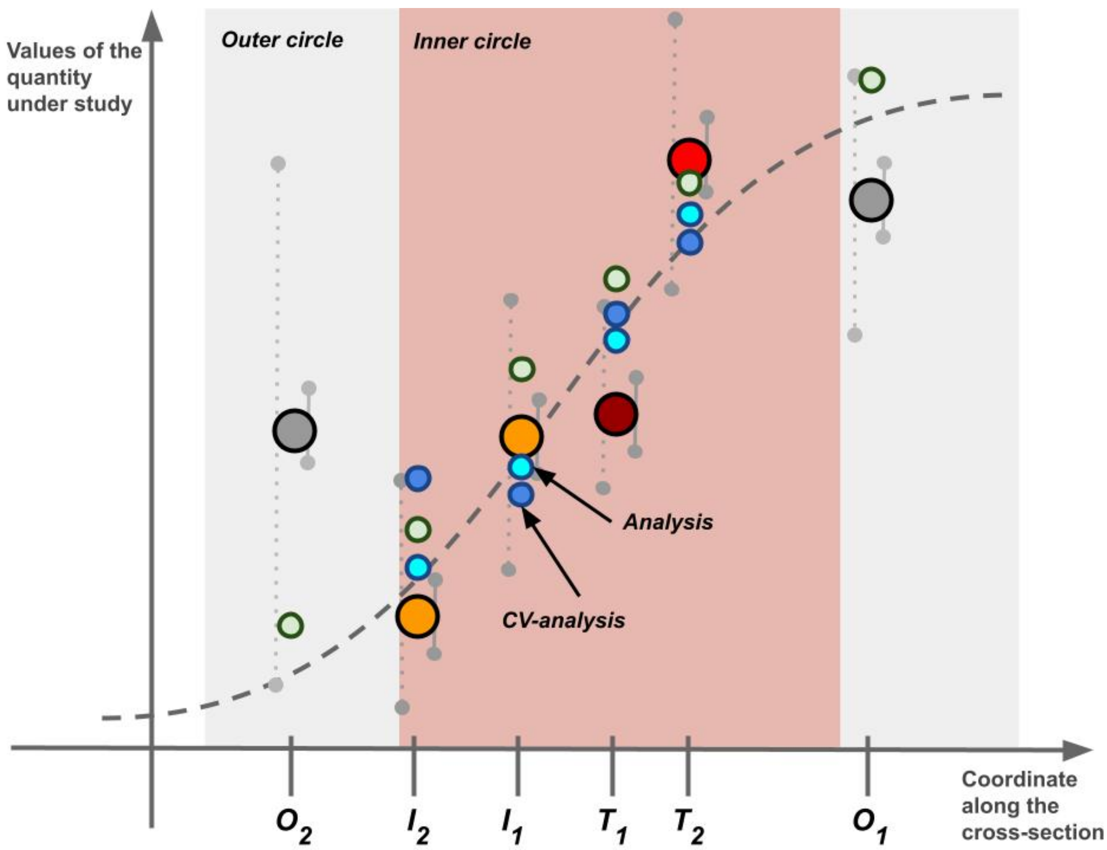


Figure 3: Same layout as in Fig. 2, with the inclusion of analyses (cyan dots) and cv-analyses (blue dots).

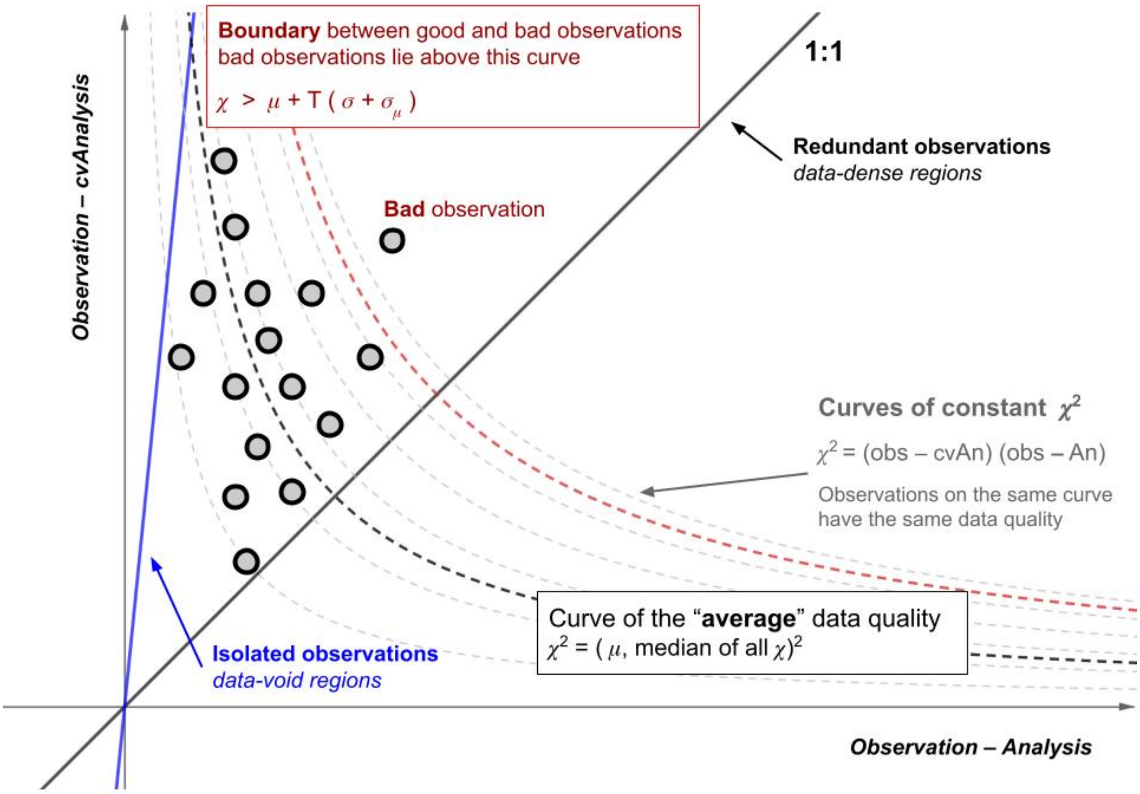


Figure 4: Schematic representation of an hourglass graph, that is CV-analysis residuals versus analysis residuals. The significant elements of the graph are described within the graph itself.

References

Båserud, L., C. Lussana, T. N. Nipen, I. A. Seierstad, L. Oram, and T. Aspelien (2020), Titan automatic spatial quality control of meteorological in-situ observations, *Advances in Science and Research*, *17*, 153–163, doi:10.5194/asr-17-153-2020.

Gandin, L. S. (1988), Complex quality control of meteorological observations, *Monthly Weather Review*, *116*(5), 1137–1156, doi:10.1175/1520-0493(1988)116<1137:CQCOMO>2.0.CO;2.

Gandin, L. S., and R. Hardin (1965), *Objective analysis of meteorological fields*, vol. 242, Israel program for scientific translations Jerusalem.

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., F. Uboldi, and M. R. Salvati (2010), A spatial consistency test for surface observations from mesoscale meteorological networks, *Quarterly Journal of the Royal Meteorological Society*, *136*(649), 1075–1088.

Tarantola, A. (2005), *Inverse Problem Theory and methods for model parameter estimation*, SIAM, Philadelphia.

Taylor, J. R. (1997), *An introduction to error analysis. The study of uncertainties in physical measurements*, 327 pp., University Science Books, Sausalito, Californi (USA).

Uboldi, F., C. Lussana, and M. Salvati (2008), Three-dimensional spatial interpolation of surface meteorological observations from high-resolution local networks, *Meteorological Applications*, *15*(3), 331–345.

Symbol	Description	type
r_{in} (r_{out})	radius of the inner (outer) circle	s
p	global number of observations	s
p_{in} (p_{out})	number of observations in the inner (outer) circle	s
p_{test}	number of observations to test in the inner circle	s
$p_{\mathrm{out,mn(mx)}}$	minimum (maximum) number of observations needed in the outer circle	s
$\boldsymbol{y}^{\mathrm{o}}$	observations	v
$\boldsymbol{y}^{\mathrm{b}}$	background	v
$\boldsymbol{y}^{\mathrm{a}}$	analysis	v
$\check{\boldsymbol{y}}^{\mathrm{a}}$	cross-validation (cv) analysis	v
$\boldsymbol{v}^{\mathrm{mn(mx)}}$	lower (upper) limits of the first-guess of the valid predicted values	v
$\boldsymbol{a}^{\mathrm{mn(mx)}}$	lower (upper) limit of the first-guess of the admissible predicted values	v
ε^2	relative precision of the observation with respect to the background	v
$\tilde{\mathbf{S}}$	Background error correlation matrix	m
$\boldsymbol{D}$	horizontal de-correlation length	v
$\boldsymbol{D}_{\mathrm{mn(mx)}}$	minimum (max) allowed value of the horizontal de-correlation length	v
$\boldsymbol{D}_z$	vertical de-correlation length	v
k	rank of the furthest observation to use in the determination of D	s
$\boldsymbol{T}$	SCT threshold	v
$\boldsymbol{T}_{+(-)}$	SCT threshold when observation minus cv-analysis is positive (neg.)	v
χ	SCT score	v
χ'	alternative SCT score	v
μ	median of the SCT scores	s
σ	dispersion of the SCT scores	s
σ_μ	confidence on μ	s
$\boldsymbol{z}$	spatial outlier detection (SOD) score	v
z_{mx}	maximum SOD score for predictions outside the valid range	s
L	length scale used in the stochastic generator of precipitation fields	s
$\boldsymbol{r}_i$	position vector (e.g. lat., long., elevation) of the ith observation location	v
$\Delta d(\boldsymbol{r}_i, \boldsymbol{r}_j)$	horizontal (radial) distance between the ith and the jth observation locations	s
$\Delta z(\boldsymbol{r}_i, \boldsymbol{r}_j)$	elevation difference between the ith and the jth observation locations	s

Table 1: Mathematical notation. The variable types are abbreviated as: “s” for scalars; “v” for vectors; “m” for matrices.

Algorithm 1 SCT algorithm. For the sake of brevity, observations affected by GEs are named *bad* observations, while those without GEs are *good* observations.

Require: see the variables in Table 1

The following loop is repeated several times, each time learning from the previous iteration and testing only those observations that are left without a decision. As a measure of precaution, during the first iteration, SCT-core is used only to flag bad observations. The iteration of the loop should end when no bad observations are flagged.

for all observations $\{i = 1, \ldots, p\}$ (sequence of SCTs over sub-regions) **do**

Definition of a sub-region: Assess if the ith location is suitable as a centroid of the concentric *outer* and *inner* circles, having radii of r_{out} and r_{in}, respectively. A location is suitable as a centroid if the corresponding observation has not previously been flagged as either good or bad.

Select the subset of p_{out} observations inside the outer circle, with $p_{\text{out,mn}} \leq p_{\text{out}} \leq p_{\text{out},mx}$

Select the subset of p_{in} observations inside the inner circle, with $p_{\text{in}} \leq p_{\text{out}}$

Select the subset of p_{test} observations inside the inner circle and not tested yet, $p_{\text{test}} \leq p_{\text{in}}$. Only observations inside the inner circle can be flagged.

Assess if the ith location is isolated, that is when there are no other observations inside the inner circle OR $p_{\text{out}} < p_{\text{out,mn}}$. Observations from isolated locations can not be checked, because the distinction between RE and GE is deemed as not reliable.

Calculate the background over the p_{out} observations (if $\boldsymbol{y}^{\text{b}}$ has not been specified)

Shortcut: if the observations and the backgrounds at all p_{test} observations are close enough, then flag all p_{test} observations as good. The condition is: $\boldsymbol{v}_j^{\text{mn}} \leq \boldsymbol{y}_j^{\text{b}} \leq \boldsymbol{v}_j^{\text{mx}}$, where j is the index over the p_{test} observations.

SCT-core: Perform SCT over the p_{test} observations, considering the p_{out} observations and based on the statistics collected on the p_{in} observations, either flag the worst observation as bad or all of them as good ones.

end for

Perform two additional iterations of the previous loop. First iteration, consider as centroids all observations without flags. Second iteration, consider as centroids all the bad observations and make use of good observations only. Bad observations are turned into good ones if they pass the final round.

Algorithm 2 OI elaboration for SCT. It is assumed OI is performed once a centroid location has been chosen, then the observations and background considered here are those within the outer circle. The scalar values for D, D_z, ε^2 and all thresholds T are the elements of the corresponding vectors at the centroid observation.

Require: see the variables in Table 1

Calculate D as the average distance between the locations in the outer circle and their kth closest observation location. Note that D is constrained by the condition $D_{\text{mn}} \leq D \leq D_{mx}$

$\tilde{\mathbf{S}} : \tilde{\mathbf{S}}_{hl} = \exp[-0.5(\Delta d(\mathbf{r}_h, \mathbf{r}_l)/D)^2 - 0.5(\Delta z(\mathbf{r}_h, \mathbf{r}_l)/D_z)^2]$, $h, l = 1, \ldots, p_{\text{out}}$

for all p_{in} observations $\{i$ is the index over observations in the inner circle$\}$ **do**

$\boldsymbol{y}_i^{\text{a}} = \boldsymbol{y}_i^{\text{b}} + \tilde{\mathbf{S}}_{i,:} \, (\tilde{\mathbf{S}} + \varepsilon^2 \mathbf{I})^{-1} \, (\boldsymbol{y}^{\text{o}} - \boldsymbol{y}^{\text{b}})$

$\check{\boldsymbol{y}}_i^{\text{a}} = \boldsymbol{y}_i^{\text{o}} - 1/(\tilde{\mathbf{S}} + \varepsilon^2 \mathbf{I})_{ii}^{-1} \, (\tilde{\mathbf{S}} + \varepsilon^2 \mathbf{I})_{i,:}^{-1} \, (\boldsymbol{y}^{\text{o}} - \boldsymbol{y}^{\text{b}})$

if $\boldsymbol{a}_i^{\text{mn}} \leq \check{\boldsymbol{y}}_i^{\text{a}} \leq \boldsymbol{a}_i^{\text{mx}}$ (j is the index over inner observations satisfying this condition) **then**

$\chi_j = \sqrt{(\boldsymbol{y}_i^{\text{o}} - \check{\boldsymbol{y}}_i^{\text{a}})(\boldsymbol{y}_i^{\text{o}} - \boldsymbol{y}_i^{\text{a}})}$

$\chi'_j = \sqrt{\varepsilon^2/(1+\varepsilon^2)}(\boldsymbol{v}_i^{\text{mx}} - \boldsymbol{v}_i^{\text{mn}})$

end if

end for

Shortcut. if all cv-analysis values are outside the range of admissible values, then flag all the observations to test as bad ones.

Compute χ statistics. $\mu = q_{0.5}(\chi)$ and σ is the greater value among $q_{0.75}(\chi) - q_{0.25}(\chi)$ and $q_{0.75}(\chi') - q_{0.25}(\chi')$. Then, $\sigma_\mu = \sigma/\sqrt{\text{size\_of}(\chi)}$

Compute $\mathbf{z} = (\chi - \mu)/(\sigma + \sigma_\mu)$

Find z_{mx} as the maximum value of $\mathbf{z}$ for those indexes i when $\check{y}_i^{\text{a}} < \boldsymbol{v}_i^{\text{mn}}$ or $\check{y}_i^{\text{a}} > \boldsymbol{v}_i^{\text{mx}}$ (if this condition is not satisfied by any observation, then $z_{\text{mx}} = 0$). Let us assume z_{mx} corresponds to the jth observation

If $(\boldsymbol{y}_j^{\text{o}} - \check{\boldsymbol{y}}_j^{\text{a}}) \geq 0$, then $T = T_+$, else $T = T_-$

if $z_{\text{mx}} > T$ **then**

set the jth observation as bad

else

set all the p_{test} observations to good

end if