Hrvatski meteorološki časopis – Croatian Meteorological Journal, 41, 2006., 3–19.

UDK: 551.501.45
Izvorni znanstveni rad

GENETIC ALGORITHMS: TOWARDS THEIR USE IN THE HOMOGENIZATION OF CLIMATOLOGICAL RECORDS

Genetski algoritmi: o njihovu korištenju u homogenizaciji klimatoloških zapisa

ALEXANDER JANN

Central Institute for Meteorology and Geodynamics
Hohe Warte 38, A-1190 Vienna, Austria
Alexander.Jann@zamg.ac.at

Prihvaćeno 7. prosinca 2005, u konačnom obliku 24. studenog 2006.

Abstract: The prerequisite of the homogenization of climatological time series is to find the points of time where the series underwent abrupt shifts in the climate variable. Since normally no information about the (potentially large) number of such shifts is available beforehand, this is a highly challenging optimization problem for the resolution of which a genetic algorithm approach was proposed. In the present paper, important steps towards its applicability to real climatological data are made. An outlier test is incorporated in the genetic algorithm framework. Another typical climatological component that can be fully integrated into the genetic algorithm scheme is the exploitation of metadata containing information about a station's history. In a conceptually similar manner, the simultaneous investigation of several, possibly interrelated, time series is made feasible.
The advantages of online analysis, where the change-point detection is carried out not only for the entire *n*-element series but also for the partial series of its first *n-1*, *n-2*, … elements, are discussed. Such supplementary analyses can regularly be crucial when change-point signals are detected in the analyses of partial series while random noise accidentally obscures these signals in the full series.

Key words: Climatological data processing, time series analysis, homogenization, genetic algorithm

Sažetak: Svrha homogenizacije klimatoloških vremenskih nizova jest određivanje vremenskih točaka kada se javlja nagli skok vrijednosti klimatske varijable. Budući da obično unaprijed ne raspolažemo podacima o broju (potencijalno velikom) takvih promjena, suočeni smo s vrlo zahtjevnim optimizacijskim problemom, za čije se rješenje predlaže primjena genetskog algoritma. U ovom su radu učinjeni znatni koraci za primjenjivost metode na stvarne klimatološke podatke. U okvir genetskog algoritma ugrađen je test za podatke koji značajno odstupaju od skupa (test *outlier*). Još jedna tipična klimatološka komponenta koja se u potpunosti može unijeti u shemu genetskog algoritma jest korištenje podataka o povijesti postaje (metadata). Na konceptno sličan način, omogućuje se istovremeno proučavanje više vremenskih nizova, možda međusobno povezanih.
Prikazane su prednosti on-line analize, gdje se pronalaženje točaka promjene obavlja ne samo za cijeli niz *n* elemenata, već i za parcijalne nizove njegovih prvih elemenata *n-1*, *n-2,* .... Takve dodatne analize uvijek su presudne kad se u analizama parcijalnih nizova otkriju signali točaka promjena (*change points*) dok u potpunim nizovima bijeli šum može prikriti takve signale.

Ključne riječi: obrada klimatoloških podataka, analiza vremenskih nizova, homogenizacija, genetski algoritam

1. INTRODUCTION

The proper interpretation of climatological records requires the search for (multiple) *change points*, i.e. points of time where the statistical character of the considered meteorological variable underwent a significant change. The statistical parameters being normally investigated are mean or (less frequently) variance. Numerous tests are available for checking the stationarity of these quantities: Herzog and Müller-Westermeier (1998) compiled a list of more than 30 relevant test methods. The tests that have become comparatively popular in climatology for investigating sample means are the *Standard Normal Homogeneity Test* invented by Alexandersson (1986) and the bivariate test introduced by Potter (1981). Both assume that the population mean remains strictly constant in-between the change points; other tests can detect shifts in the mean even if trends are superimposed on the data (e.g. Buishand, 1984; Alexandersson and Moberg, 1997). Once the change points are identified, it is necessary to assign each of them to one of the two classes: *artificial change* (caused by changes in instrumentation or observers, station relocation, etc.) or *climate change*. Homogenization, i.e. correction for those shifts in the mean being classified as *artificial*, is a prerequisite for sound statements about trends in the climate system.

Many of the tests in use are of the *two-sample type*, i.e. two samples are compared in order to infer whether they originate from the same population or not. A well-known parametric representative of this group of tests is the classical Student's *t*-test dealing with normally distributed populations. The formula for the test statistic *t* is

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s^* \sqrt{1/n_1 + 1/n_2}} \tag{1a}$$

with

$$s^* = \sqrt{\frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}}\,, \tag{1b}$$

where $\bar{x}_1$ and $\bar{x}_2$ are the two sample means, s_1 and s_2 are the respective standard deviations, n_1 and n_2 are the sample sizes. When using the *t*-test as a change-point detector, the considered series is split at the individual observations, the basic statistical parameters $\bar{x}$ and s are computed for the two resulting sub-samples, and t is derived. The associated cumulative Student's probability distribution function $F(-|t|)$ with n_1+n_2-2 degrees of freedom can be computed, and the probability of equality of the population means μ is given by

$$p = \Pr(\mu_1 = \mu_2) = 2F(-|t|)\,. \tag{2}$$

The location for which the associated p is lowest can be considered as the most likely position of a change point (Buishand, 1982; Szinell, 1997), which is defined as the position *after which* the observations with the *new* probability distribution function follow.

Having the common situation in climatology in mind, the interesting question emerges how such a two-sample test can be generalized to handle the case of a time series constituted by numerous populations, with the number of change points being unknown. In previous works (e.g. Easterling and Peterson, 1995; Lanzante, 1996; Herzog and Müller-Westermeier, 1997; Moberg and Alexandersson, 1997), it has been attempted to use two-sample tests in a sort of *sequential* manner. First, all configurations with one change point were investigated. Then, taking the first change point for granted, the most likely position of a possible second change point was sought under this precondition. This procedure of increasing the number of change points sequentially was repeated until a termination criterion was reached.

Reservations against such schemes root in the fact that – if there are actually several populations involved – the considered samples will normally represent mixtures of populations. Naturally, there exists no test which is so robust to yield a correct change-point position for any arbitrary mixture. Therefore, starting with the search for a single change point and irrevocably fixing it at the identified *optimum* position carries the risk of leading the procedure astray already at the very beginning. Eventually, such considerations lead to the conclusion that it is highly recommendable to treat the problem of multiple change-point detection as one of global optimization. Unfortunately, a full evaluation of all possible change-point configurations would consume an enormous amount of computing time for typical climatological series (Caussinus and Mestre, 1997; 2004). Hence, necessarily, the term global optimization normally means devoting a (considerably) increased amount of time in order to achieve a (considerably) better approximation to the true solution. For doing this, Jann (2000) proposed a genetic algorithm (GA) as an optimization tool.

In the paper quoted, a couple of experiments were described where the genetic algorithm was successfully applied to simulated data. It is almost imperative to base the assessment of the merits of a new concept (or an intercomparison of methodologies, cf. Ducré-Robitaille et al., 2003) on data that are known to conform to the prerequisites of the tests involved. Consequently, we defer application to real climatological records to future studies, and by virtue of the numerical simulation approach at several instances will benefit in the following from knowing the true change points by design. Nevertheless, the present paper demonstrates important progress towards climatological practice as it goes beyond the consideration of perfect series in isolation: Outlier detection is addressed in Section 3, the integration of metadata records is the issue of Section 4, and the simultaneous analysis of several time series is considered in Section 5. Section 6 discusses aspects of real-time monitoring. Before dealing with these issues, the following section provides an overview of the employed GA. More details, in particular on the motivation behind some of the features introduced, can be found in the original article (Jann, 2000).

As the focus of the present paper is on selected aspects of change-point detection, the subject of homogenization will not be covered in full depth. Readers interested in a more comprehensive treatise on the general aspects of homogenization are referred to the review article by Peterson et al. (1998). A wealth of additional information about mathematical concepts with respect to the detection of abrupt changes can be found, for example, in the book by Basseville and Nikiforov (1993).

2. FUNDAMENTALS OF THE OPTIMIZATION PROCEDURE

2.1. The genetic algorithm

The material on which GA operations are applied here are samples of bit strings, where the number of bits is given by the number of observations in the investigated climatological series and the symbol '1' is reserved for those positions where a change point is assumed. For example, for a 20-element series with one change point at position 10, the solution of the change-point detection problem is described by the sequence 00000000010000000000. In the optimization problem under consideration, the bit strings represent the independent variable, and the cost function C (to be specified in the subsequent section) is the dependent expression to be minimized. The optimization is carried out by a systematic random walk through the bit pattern space using the standard mechanisms of GAs: selection, crossover and mutation. The sequence of these three mechanisms is repeated until presumed convergence.

The computations start with the production of a certain number Q of parental bit sequences (Q=200 throughout the paper unless indicated otherwise). In our implementation, a tenth of the bits were randomly selected and set to 1. Bits with value 1 were set back to 0 if one of the four preceding bits was 1, in order to avoid too short sub-series. Judging bit sequences to be *illegal* when the distance between assumed change points falls below a minimum of 5 observations was motivated by the findings of Karl and Williams (1987) and Easterling and Peterson (1995) that at least five data items of each involved population are required to arrive at reliable statistical conclusions. When undesired bit sequences with change-point markers being separated by less than four zeros were produced through the mechanisms described below, such sequences were immediately excluded from any further processing. After computing the cost functions C of the parents (the largest value within the sample being designated $C_{\max}$), the *selection step* chooses Q surviving candidates from the parent generation, with probabilities of selection being proportional to

$$P_j = \alpha C_{\max} - C_j . \tag{3}$$

The parameter α ($\geq$1) is to be set on an empirical basis; it was kept constant at 1.5 throughout the study.

The selected parents are paired and subject to an exchange of sequences with their respective companions (this is the *crossover step*). The crossings took place strictly in the order in

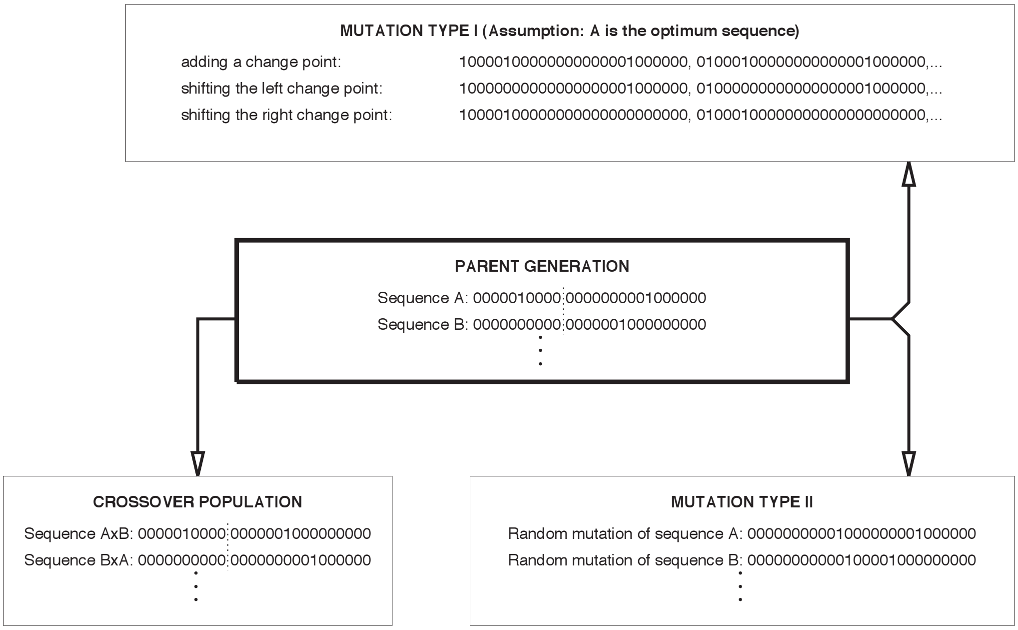

Figure 1. Illustration of the mechanisms constituting the genetic algorithm. The crossover is assumed to take place at the position marked by the dashed line.

Slika 1. Ilustracija mehanizama koji tvore genetski algoritam. Presjek se nalazi na položaju označenom isprekidanom crtom.

which parents were selected in the preceding step, i.e. the first was paired with the second, etc. For each pair, the position of the splitting was determined by drawing from a uniformly distributed random variable. Figure 1 gives an example.

The *mutation step* alters the value of some selected bits, leaving the rest of the sequence unchanged. Two variants of this step were implemented: In the first iteration, after the finding of a new best sequence, sequences closely resembling this (provisional) optimum were investigated, which should normally be the best basis for further progress (=mutation type I):

- The effect of introducing one additional change point was investigated for each possible position;
- The second mechanism producing alternative bit sequences was to shift each of the indicated change points.

However, if the previous iteration yielded no improvement in the minimum C, a mutation of *type II* was carried out. This was a random mutation applied to those parents which had survived the selection step. The number of mutations was chosen to be Q, i.e. on average, each string underwent a single change. $3Q/4$ mutations were reserved to take place from 0 to 1 whereas the other $Q/4$ mutations were changes from 1 to 0 provided that a sufficient number of change points were present in the considered population. Figure 1 offers illustrations of both types of mutation.

The cost functions C were calculated for both the crossover population and the mutated sequences, and the best $Q/2$ mutually different representatives of each group were elected to enter the next parent generation (which was then subject to the selection step, and so forth). If there were not enough candidates to fill the Q places, a random draw was made from the already investigated bit sequences in order to complete the new parent generation.

Convergence was assumed and the procedure was terminated after a certain number of iterations (consisting of selection, crossover, and mutation) without any progress in lowering the minimum C. This threshold number of unsuccessful iterations, n_{it}, was set to 30 in the experiments described, unless indicated otherwise.

2.2. The cost function

The cost function that was chosen to be minimized writes

$$C = \frac{s_r^2}{s^2} - \beta_1 \nu - \beta_2 \tag{4}$$

where s^2 is the variance of the original series and s_r^2 is the variance of the adjusted series (where for each supposed population the derived sample mean is subtracted from its elements). The factor β_1 is set to 1 if for all assumed change points the values p derived from the t-test satisfy a prescribed significance criterion $p \leq p_t$ (or, equivalently, $F(-|t|) \leq p_t/2$), otherwise $\beta_1=0$. The symbol ν designates the number of indicated change points, hence the term $-\beta_1\nu$ reflects the philosophy of detecting the maximum number of significant discontinuities as C decreases when the number of significant change points goes up.

However, prompted by the unsatisfactory results found for certain configurations (albeit those would hardly ever be encountered in climatology), it was decided to take only those shifts into account for the determination of ν where the sense of the preceding jump (i.e. increasing or decreasing) is different from the current one, i.e. the proper detection of back-and-forth jumps is supported through the bonus term whereas the recognition of patterns of two consecutive positive (or negative) shifts is left to the term s_r^2/s^2. Though in this counting scheme some of the indicated shifts may not contribute to the value of ν, they must still, of course, be significant in order to make $\beta_1=1$. The third term of the cost function is a technical one introduced to support convergence towards the optimum solution by favouring those bit sequences which contain some interesting change-point positions: β_2 is set to 1 when at least one change point meets the significance criterion, otherwise $\beta_2=0$. Summing up, three groups of bit sequences are separated by means of cost function (4):

1) The sequences containing entries of '1' but not a single significant change point. These sequences are characterized by positive values of C.
2) Those with both significant and insignificant change points indicated. The cost functions are negative, but $C>-1$.
3) Finally, the sequences where each entry of '1' marks a significant change point; here, $C<-1$.

The null sequence (=assumption of a homogeneous series) has been duly included into the scheme by defining its cost function to be -1.
With respect to the significance testing, a two-step procedure was implemented, as motivated by Jann (2000). After the GA had converged for a relaxed significance threshold, i.e. a higher value of p_t, the Q best bit sequences were picked from a ranking for the desired stricter limit. They constituted the initial parent generation for a repetition of the procedure, now with a tighter threshold. All runs dealing with cost function (4) started with p_t=0.05 and then switched to a significance threshold of p_t=0.01.

3. HANDLING OF OUTLIERS

Outliers, i.e. isolated values of a time series which (in a statistical sense) do not belong to the population(s) surrounding it, may, for example, appear in climatological records as a result of incorrect readings or errors during the encoding of data. Occasionally, even a correct value may constitute an outlier and bear interesting climatological information (e.g. in El Niño situations or during sudden stratospheric warmings). However, in the context of homogenization, where the overall characteristics of the series is to be judged, it is probably appropriate to disregard single extreme values of whatever origin (cf. González-Rouco et al., 2001).

The basic concept of detection and subsequent elimination of outliers need not be complex: Gille (1997) and Wolter (1997) considered a value an outlier when it was outside a range

$$(\textit{mean} \text{ or } \textit{median}) \pm f \times (\textit{standard deviation}),$$

with f being a subjectively prescribed factor which in the quoted papers ranged from 3 to 4.5. This test will be used hereafter, with f (arbitrarily) set to 3.5.

Lanzante (1996) strongly argued against the use of the traditional mean and standard deviation for the determination of the confidence interval. Using standard formulas, present outliers distort the statistical quantities to such a degree that outlier detection may be outright impossible. The solution proposed by Lanzante (1996) is the use of robust estimates called biweight mean and standard deviation, respectively, which are less prone to distortions by outliers. These new parameters are symbolized hereafter by $\hat{x}$ and $\hat{s}$; the relevant formulas are given in Lanzante (1996), Appendix B.

For the inclusion of outlier probing into the change-point detection process, a scheme may be envisaged where the GA is essentially unchanged, but for each assumed population the members have to pass the test of lying within $[\hat{x} - f\hat{s}, \hat{x} + f\hat{s}]$. After the elimination of any indicated outliers, the common mean and the standard deviation are derived for the *cleaned* sample and eventually used for computing cost function (4).

The statistical dilemma faced here is that an outlier should be disregarded during change-point testing but, on the other hand, the very same outlier may be difficult to identify as long as the main statistical parameters of the populations are not known, when the series is not yet split into its components. Thus there are two problems and, in principle, each requires the other to be solved beforehand. The proposed scheme sets outlier detection first by carrying out the outlier test even on mixtures of populations. In doing so, the standard deviation can increase considerably and the confidence interval of the outlier test thus widens. Outliers (that would be recognized if the assumed change-point positions were more appropriate) may then be overlooked (Fig. 2a). Fortunately, the numerical effect for the change-point test in this case is that two factors, the mixing of populations and ignorance about the outlier, tend to act in the same direction, namely towards an increase of the cost function. Hence, the GA selection step will likely eliminate the affected bit sequence with incorrectly set change-point positions since it is inferior to other sequences where the change points fit better. The latter are encountered despite yet undetected outliers with their uncertain impact on the test statistic (hence, in principle, having the potential to guide an algorithm away from the right change-point positions). Due to the GA mechanisms, breaks at

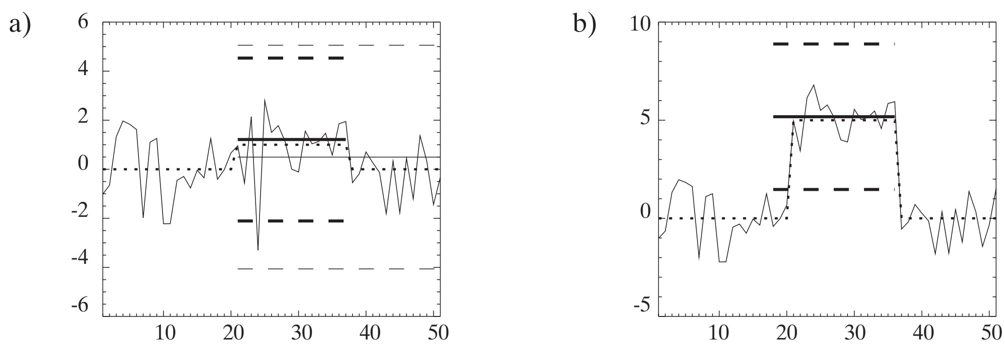

Figure 2. Illustration of potential difficulties when combining change-point and outlier detection. (a) Example of outlier recognition failure caused by incomplete partitioning of series into populations. The outlier is located at position No. 24. The dotted line shows the course of the true population mean; bold horizontal lines indicate the biweight sample mean (solid) and the confidence interval of the outlier test $[\hat{x} - 3.5\hat{s}, \hat{x} + 3.5\hat{s}]$ (dashed) for the central population, if change points are set correctly. Thin horizontal lines show the biweight mean and confidence interval for a mixture of the central and right-hand population, i.e. if only the left change point would be correctly identified. (b) Example of incorrectly flagging a couple of values from another population as outliers because of inexact separation of populations. The dotted line shows the course of the true population mean; bold horizontal lines indicate the sample mean (solid) and the confidence interval of the outlier test (dashed) for a mixture of the central population and three elements of the left-hand population (the latter would erroneously be flagged as outliers if no provisions were taken).

Slika 2. Ilustracija mogućih poteškoća prilikom kombiniranja točke promjene i pronalaženja podataka koji značajno odstupaju od skupa (*outlier*). (a) Primjer neuspjela prepoznavanja *outliera* zbog nepotpunog dijeljenja niza u populacije. *Outlier* se nalazi na položaju br. 24. Točkasta crta prikazuje hod stvarnog srednjaka populacija; podebljane horizontalne crte prikazuju dvostruko otežani srednjak uzorka (puna crta) i interval povjerenja *outlier* testa $[\hat{x} - 3.5\hat{s}, \hat{x} + 3.5\hat{s}]$ (isprekidana crta) za središnju populaciju, ako su točke promjene ispravno postavljene. Tanke horizontalne crte označavaju dvostruko otežani srednjak i interval povjerenja za kombinaciju centralne i desne populacije, t.j. samo kad bi se lijeva točka promjene ispravno identificirala. (b) Primjer neispravnog označavanja nekoliko vrijednosti neke druge populacije kao *outliera* zbog netočne podjele populacija. Točkasta crta prikazuje hod stvarnog srednjaka populacije, podebljane horizontalne crte prikazuju srednjak uzorka (puna crta) i interval povjerenja testa *outlier* (isprekidana crta) za kombinaciju centralne populacije i tri elementa lijeve populacije (te bi posljednje bile netočno označene kao *outlieri* da se o tome nije vodila briga).

the actual change-point positions are investigated even if there has been no previous indication that such a location is promising.

A minor modification of the procedure was nevertheless necessary: if a mixture investigated during the optimization process is composed in such a way that population A contributes many more members than the adjoining population B, the members originating from B may all together be flagged as outliers. This is the case in Figure 2b where the three leftmost members of an assumed sample lie below the lower threshold of the acceptance interval. Though the test is right in flagging the three observations as not belonging to the sample population, the decision to eliminate these observations from the change-point testing would be incorrect. Such a false elimination may (albeit not necessarily) lower *C*, which could finally deliver a wrong result of the combined outlier/change-point detection procedure. Hence, it is advisable to take the additional provision that for outliers indicated at the edge of an investigated sub-sequence, the outlier criterion be checked with respect to both adjacent populations.

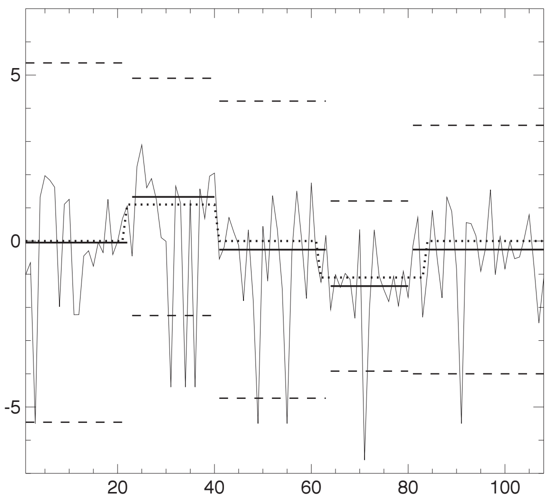

Figure 3. A representative result of the combined change-point/outlier detection procedure. The dotted line indicates the course of the true population mean. Bold solid lines indicate the biweight sample means computed for the change-point configuration that finally turned out to be the optimum. Dashed horizontal lines define the confidence intervals [$\hat{x} - 3.5\hat{s}$, $\hat{x} + 3.5\hat{s}$] corresponding to each assumed population. The eight introduced outliers lie below the lower thresholds and can be eliminated before computing *C* (Eq. 4). The result is eventually the same as if the outliers were simply not present.

Slika 3. Reprezentativni rezultat kombiniranog postupka pronalaženja točke promjene/*outliera*. Točkasta crta označava hod pravog srednjaka populacija. Podebljane pune crte označavaju dvostruko otežan srednjak izračunat za konfiguraciju točke promjene koja se konačno pokazala optimalnom. Isprekidane horizontalne crte definiraju intervale povjerenja [$\hat{x} - 3.5\hat{s}$, $\hat{x} + 3.5\hat{s}$] koji odgovaraju svakoj pretpostavljenoj populaciji. Osam unesenih *outliera* zapravo se nalazi ispod donjih pragova i oni se mogu eliminirati prije izračunavanja *C* (jednadžba (4)). Rezultat je konačno isti kao da *outlieri* nisu ni bili prisutni.

The positive assessment of the method through theoretical reasoning was substantiated by its practical application to simulated time series. Experiments were run where outliers were introduced in such a way that their identification would have been theoretically possible had the change points been known. This was done by analysing the series before any outliers were added. Sample means and standard deviations were computed for the identified components and outliers being larger than $\hat{x} + 3.5\hat{s}$ or smaller than $\hat{x} - 3.5\hat{s}$ were then randomly inserted somewhere into these subseries. Of course, the information about the change-point configuration found for the original series was withheld from the combined outlier/change-point detection procedure applied subsequently. The latter, if working properly, should identify and eliminate the outliers so that the analysis ultimately is the same as for the original series where no outliers were present. This was found to be virtually always the case. Figure 3 provides an example of such a successful verification.

4. INCLUSION OF METADATA

Metadata, i.e. records about the history of observing stations, may help in clarifying the nature of change-point signals found in climate series. A significant shift occurring at a single station is unlikely to be attributable to climatological causes when a simultaneous change in instrumentation or a station relocation is noted in the annals for that very station.

Moreover, metadata can serve as an ingredient to an improved positioning of the change point. Often, change points obtained through statistical analysis are contrasted with metadata records. If there is a (subjectively) small discrepancy between the two sources, a linkage between a specific *statistical* change point and a physical cause documented in the metadata record is normally inferred. With a metadata record considered to be trustworthy, the final positioning is determined in accordance with the supplementary data. Thus, the statistical analysis is overruled, which is perfectly acceptable if all the indicated shifts are still statistically significant for the *adjusted* configuration. Because of the impacts of random noise, the objectively determined positions can never be considered sacrosanct and a slight modification on a sound basis appears permissible.

A direct numerical inclusion of this philosophy into cost function (4) proved difficult since one cannot determine on a scientifically reasonable basis how much an agreement between the analysis and metadata record is worth in numerical terms, weighed against a decrease in statistical significance. Nonetheless, an experiment with a variant of Equation 4 with a large relative weight on metadata agreement compared to the other terms gave an instructive result: For the series shown in Figure 4a, significant change points could be forced far away from positions 26–50, where the two present breaks are actually located. For example, one could postulate a (single) change point at position 84 if metadata show that a shift at that position might be explained by station history. As Figure 4a features a simulated series, where the truth is known, that hypothesis can be rejected immediately. A basically correct analysis has been effectively destroyed due to an overvaluation of metadata, which can normally only indicate a larger probability for a shift in the series at that point of time; an actual shift is, however, not assured. (If, for whatever reason, a specific change point is considered to be sure, the implications on the algorithm are trivial and need not be scrutinized here. One would simply constrain the GA to investigate only bit sequences with a '1' in the right position.) The warning signal in the case of Figure 4a is the reduction in the number of change points, indicating that one has abandoned the original concept of merely repositioning change points. Hence, a desirable feature to be implemented in the algorithm is that not a single significant change point should be sacrificed for reaching an agreement between metadata and another change point.

Therefore, a two-step procedure is proposed. The first step is to carry out the GA procedure as described above, without considering the metadata record. Then, the GA minimizes the cost function

$$C = \frac{s_r^2}{s^2} - \beta_3 (N+1) \tag{5}$$

where N stands for the number of coincidences between change points and explanations found in metadata. The t-test is again used to verify the significance of all supposed shifts in μ; β_3 is set to 1 if all assumed change points are found significant, 0 otherwise. With the lesson learned from Figure 4a, another prerequisite for β_3 to assume the value of 1 is that the number of change points be not reduced compared to the optimum solution detected during the first run of the GA; otherwise the bonus cannot be awarded.

It is acceptable to express uncertainty about a datum contained in the metadata. For example, if an available source reports a change at the station "that took place at some time in the year 1942" and a series of monthly averages is tested for homogeneity, the change points between December 1941 and December 1942 may be considered consistent with the metadata information. Hence, whenever a change point is indicated in that period, N is increased by one (if two change points during this year are indicated, the bonus for being in compliance with the metadata is to be awarded only once).

Figure 4b provides an example of change-point detection taking into consideration (imprecise) metadata information.

5. PARALLEL ANALYSIS OF SEVERAL SERIES

Besides metadata, the use of reference series is another prominent concept in the distinction between climatic and artificial changes in a meteorological time series. A reference series is designed in such a way that it contains the climatic variability prevailing in an area (i.e. the variability found at the majority of stations) and yet is devoid of the inhomogeneities present in single station records. Subtracting the proper reference series from a climatological record should leave only the artificial inhomogeneities which can thus be unambiguously revealed.

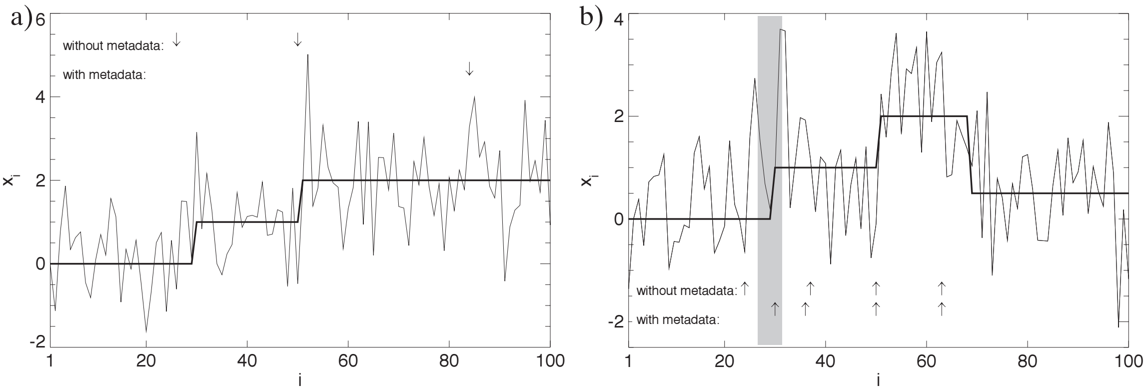

Figure 4. Metadata inclusion into the GA procedure. a) Simulated time series superimposed by the population means (bold line). The first row of arrows indicates change points found by the GA procedure without metadata information. The arrow in the second row points to element no. 84, where a significant change point could be positioned if metadata indicate a relevant event there, along with a reduction in the number of change points from two to one. b) Simulated time series superimposed by populations' means (bold line). Arrows indicate change-point positions derived from the proposed two-step procedure. First row: without metadata information. Second row: with metadata information, using Equation 5, i.e. taking care that the number of change points is not reduced. It is assumed that metadata provide a somewhat vague information that a change point somewhere between observations 27 and 31 (region shaded in the graphics) appears likely.

Slika 4. Uključivanje podataka o postaji (*metadata*) u postupak GA. a) Simulirani vremenski nizovi s pripadajućim srednjacima populacija (podebljana crta). Prvi red strelica označava točke promjene pronađene postupkom GA bez informacija iz podataka o postaji. Strelica u drugom redu pokazuje element br. 84, gdje bi se mogla nalaziti signifikantna točka promjene ako podaci o postaji upućuju na značajni događaj na tom mjestu, uz smanjenje broja točaka promjene s dvije na jednu. b) Simulirani vremenski nizovi s pripadajućim srednjacima populacija (podebljana crta). Strelice pokazuju položaje točaka promjene dobivene predloženim dvostupanjskim postupkom. Prvi red, bez informacija iz podataka o postaji. Drugi red, s informacijama iz podataka o postaji, korištenjem jednadžbe (5), tj. vodeći računa da se broj točaka promjene ne smanjuje. Pretpostavlja se da podaci o postaji daju donekle neodređenu informaciju da bi se jedna točka promjene mogla vjerojatno nalaziti između motrenja 27 i 31 (osjenčano područje na slici).

The interesting question is, of course, how to construct an average climate series when the available material is comprised of possibly inhomogeneous series with artificial breaks being yet unidentified. Literature has documented diverse reactions to the paradox. Some developed procedures to "diminish the effect of an inhomogeneity in the candidate station" (Peterson and Easterling, 1994). Others (e.g. Caussinus and Mestre, 2004) entirely rejected the use of reference series for which the alleged absence of *artificial* inhomogeneities cannot actually be proven.

There are a couple of ways to contrast series without the explicit construction of a reference series. Menne and Duchon (2001) consider difference series between adjacent stations and test whether these series show the characteristics of white noise. If the test fails at a certain station for several of its neighbours, a non-climatological inhomogeneity in the series is likely. Caussinus and Mestre (1997) also consider difference series, but apply the change-point detection procedure and look for congruent change points. If a certain station is involved in a considerable number of difference series with a shift at (about) the same time, this points to an *artificial* inhomogeneity at this station. The simplest example is the case of three series A, B and C. If a change point is common to the difference series A-B and B-C (but absent in A-C), it eventually leads to the conclusion that series B alone exhibits an inhomogeneity there which, in turn, is probably *artificial*.

The latter approach is of particular interest here since it can be integrated fairly easily into the GA framework. In fact, the procedure has by and large already been devised in the previous section when considering the inclusion of metadata. Now, however, it is not coherence with a reference event that is sought but the simultaneity of change points where not a single one is given at the beginning.

Figure 5 shows an example of three series to which inhomogeneities were introduced, namely:

- series A, with change points imposed at positions 20 and 44;
- series B, change points at positions 33 and 58;
- series C, one change point at position 71.

No assumption had to be made about superimposed climatic variability because it would have been eliminated anyway through the use of difference series in the analysis process. Precautions were taken, however, to ensure that the three simulated series had a reasonable amount of correlation between them (before adding the artificial jumps), because mutual correlation between the series is a prerequisite of the Caussinus&Mestre method (as well as of the method of reference series). To view poorly correlated series in conjunction could well mean mixing climatologically incomparable regions, thus invalidating all subsequent conclusions.

Figure 5 documents the results of the optimization of the cost function (4) undertaken for the three difference series A-B, B-C and A-C. Building upon the preliminary analyses for the three difference series, the GA then minimized the cost function

$$C = \overline{s_r^2 / s^2} - \beta_3 (N + 1) \tag{6}$$

where N stands for the number of concurrent change points found in any two series (generally, the two difference series with a common change point need to have one original series in common in order to contribute to N; in the simple three-series case, this is always fulfilled). The overbar in Equation 6 indicates averaging over all difference series. The factor β_3 has the same function as in Equation 5: it is set to 1 if all supposed shifts in μ are found significant (again using the t-test), 0 otherwise. Another prerequisite for awarding the bonus is that, by analogy to the previous section, for each series, the number of change points is not reduced compared to the optimum solution detected during the runs for the single series.

For this application, a natural algorithmic extension to be added to mutation type I of the GA is a simultaneous shifting of those change points which contribute to N. This is a promising mechanism for improving C, since the bonus of coincidence could be retained while a somewhat better positioning is sought. It is of course much more efficient to include this search explicitly into the procedure rather than relying on the random processes to incidentally yield the same modifications.

When performing the optimization of C (Eq. 6) for the case in Figure 5, the change-point indications are somewhat shifted. Coinciding, statistically confirmed change points are reported at:

(a) position 22 for A-B and A-C (consequently the inhomogeneities are expected to originate from series A);
(b) position 34 for A-B and B-C (probable origin: series B);
(c) position 40, where the situation is similar to (a);
(d) position 50, where the situation is similar to (b) (Interestingly, the original analyses indicated a coincidence at position 62 which has now been shifted 12 positions to the left. With the truth known, one would prefer the positions at No. 62, but statistics decided otherwise. If such a pattern were observed for a real climatological series, one would consult metadata to seek clarification.);
(e) position 78 for A-C and B-C; inhomogeneity in series C as probable cause.

The good consistency of these results with the designed change-point configurations is obvious.

The increase of change points in the analysis of series B-C may require explanation: For cost function (4), the term decided against the assumption of three change points (by a very narrow margin). Since a staircase configuration added to s_r^2/s^2 in (4) at most 1, the bonus term provided no stimulus towards such a solution, either. With the cost function (6), however, the decision between the two statistically possible solutions was clearly in favour of the change-point triplet since all three change points coincide with the counterparts of the other two difference series.

6. ONLINE ANALYSIS

Change-point detection becomes an even more challenging task if the application demands the identification of the change point

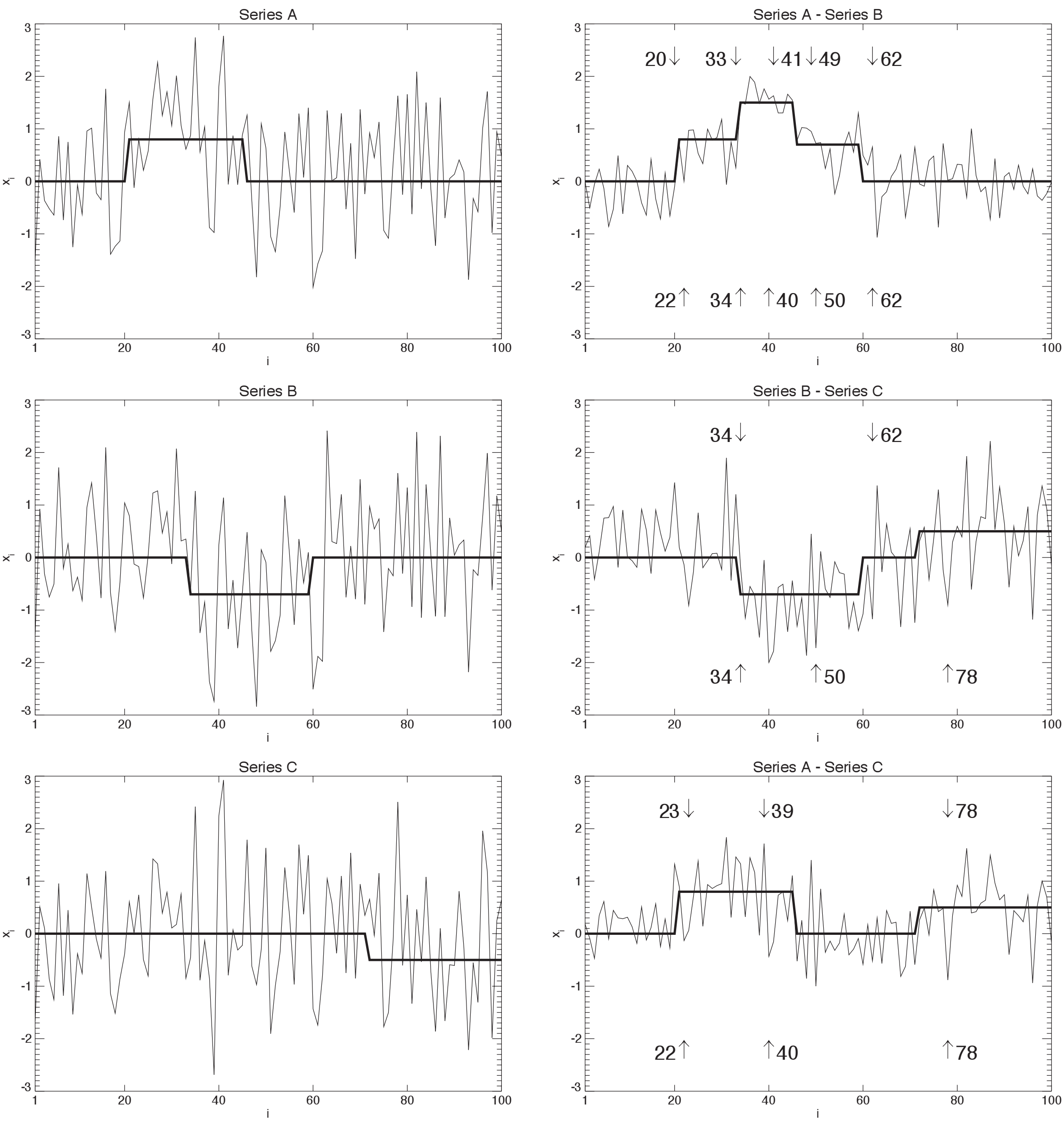

Figure 5. Left column: Simulated time series A, B and C with imposed change points; bold lines show courses of the populations' means for each series. Right column: Analyses of difference series A-B, B-C and A-C (bold lines show the corresponding differences of populations' means = populations' means of the difference series). The arrows and numbers in the upper halves of the panels indicate the change-point positions obtained when analysing each difference series in isolation. The positions indicated in the lower halves were obtained from a subsequent coupled analysis of the three series, using cost function (6).

Slika 5. Lijevi stupac: Simulirani vremenski nizovi A, B i C s definiranim točkama promjene. Podebljane crte prikazuju hodove srednjaka populacija za svaki niz. Desni stupac: Analize nizova razlika A-B, B-C i A-C (podebljane crte prikazuju odgovarajuće razlike srednjaka populacija = srednjake populacija nizova razlika). Strelice i brojevi na gornjoj polovini slike prikazuju položaje točaka promjene dobivene pri analizi svakog niza razlika posebno. Označeni položaji na donjoj polovini slike dobiveni su naknadnom analizom kombinacije tri niza podataka, upotrebom *cost* funkcije dane u jednadžbi 6.

soon after its occurrence. The performance of procedures in a real-time framework is critical in fields such as bacteriological infections (Whittaker and Frühwirth-Schnatter, 1994), short-range forecasting of severe weather by means of meteorological radars (e.g. Browning and Collier, 1989) and industrial quality control. Quality assurance is also an area where climatological variables may become the subject of continuous real-time inspection (a procedure which is often termed 'online analysis'). Menne and Duchon (2001) emphasize how important it is to detect potential inhomogeneities quickly so that corrective measures can be soon applied to the affected station. Another situation, where the focus is on the most recent sections of a climatological series and quick answers are desirable, is when public interest so dictates, with the global warming issue probably being the most prominent example.

Interestingly, however, there is no need to justify the set-up of an *online mode* with reference to the nature of the application. As will be shown shortly, even the analysis of a climatological series that terminated, say, 50 years ago could benefit from performing the analysis as if it were a real-time application.

6.1. Aspects of practical implementation

Figure 6 shows an example where the GA optimized the cost function (4) not only for the entire 100-element series but also for the partial series comprising the first 10, 11, ... 99 elements. The visualization of the results of the increasing length of the series confirms the expectation that any run can greatly benefit from the preceding run. Of course, if the opposite were true – with change-point positions fluctuating pronouncedly with any newly included observation – one would have to question either the robustness of the statistical procedure or the solvability of the statistical problem as a whole. For the (representative) example in Figure 6, it is obvious that, usually, using the previously obtained optimum as a starting point and adding a 0 at the end of the bit sequence, the solution is either given already at the very beginning or found in the first iteration through the mechanisms of mutation type I. Hence, computing time can be considerably reduced by exploiting the available information rather than starting the procedure after each observation from scratch.

It is almost trivial that, also in terms of the quality of the final result, an unaltered GA profits from preceding optimization runs for shorter sub-series. If the solutions of the previous runs resemble the final solution, they strongly facilitate finding the latter; if the optimum change-point configurations of the precursors are completely different from the solution for the entire series, the situation is the same as for an uninformed GA. With the GA performing a random walk, inappropriate starting sequences are very quickly eliminated and do not impact the algorithm negatively.

In order to find out whether the additional computational effort incurred by investigating the partial series can be compensated by such a measure as lowering Q from 200 to 30, for example, a small simulation experiment was launched. The test material consisted of 100 series, each comprising 100 random numbers representing samples from a normal distribution with mean 0 and standard deviation 1, shortly written as N(0,1). Each series was modified through adding an increment $\Delta\mu$ to its elements No. 30 to 50, i.e. two change points were imposed at events 29 and 50 so that the distribution first changed to N($\Delta\mu$,1) and returned to N(0,1) later. The analyses of the entire series (Q=200) were compared with the final results of the online analyses with Q=30, n_{it}=20. The used increments were $\Delta\mu$=1.0 and $\Delta\mu$=1.5, and for both values the outcome showed 93 of the 100 series to have been analyzed identically by the two approaches. Evaluating the few instances with differing outcomes gave an ambiguous result, as – besides several cases where the expected benefit was observed – there were a couple of situations identified where results for the shorter series apparently suggested a certain suboptimal configuration, and with a low Q, the number of investigated alternatives was too small to find the path to the existing better solution (for Q=20, the problem aggravated to such an extent that the online mode appeared to be of questionable value).

In climatology, the ultimate analysis, i.e. the one for the entire series, normally has highest priority. For such applications, experience suggests that the valuable information from the analyses of sub-series can, indeed, often be obtained from a 'quick mode', i.e. from runs with smaller Q. To be on the safe side, a higher

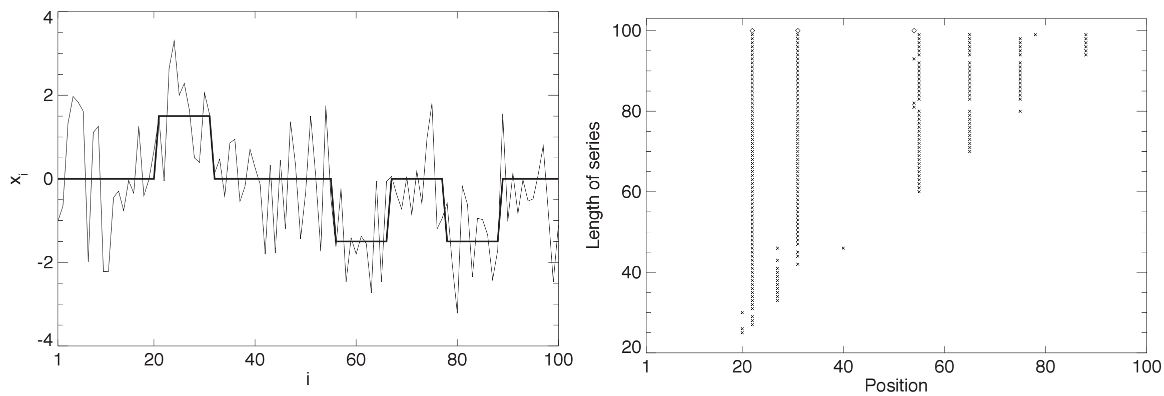

Figure 6. Left panel: Simulated time series with the course of population means superimposed as a bold line. Right panel: Results of change-point detection (via optimization of cost function (4) by the GA) for the increasing length of the series. The crosses mark the positions of the detected change points; for the sake of clarity, the result for the full series is depicted by diamonds.

Slika 6. Lijeva slika: Simulirani vremenski nizovi s pripadnim hodom srednjaka populacija (podebljana crta). Desna slika: Rezultati pronalaženja točaka promjene (optimizacijom *cost* funkcije (4) pomoću GA) za produženje niza. Križići označavaju položaje pronađenih točaka promjena. Da bi se dobilo na jasnoći, rezultat za cijeli niz označen je rombovima.

Q should be used in the last run, where the complete series is analyzed. As already argued, the process *certainly* benefits – or is at least not impacted negatively – from such a proceeding in terms of the final result (compared to the analysis of the complete series alone).

6.2. The benefit of online analyses from a theoretical statistical viewpoint

By increasing the length of the series stepwise, it is assured that the change points appear one by one. This situation is fundamentally different (and much more favourable) compared to the immediate analysis of the whole series where each inspected sample is potentially a mixture of several populations: it is reasonable to expect that a test like the t-test, designed for dealing with a single (additional) change point, can, under these circumstances, be employed in much better accordance with its actual capabilities.

For schemes sequentially setting change points one by one, this opens up an option to approach the global optimum much better than otherwise possible. Figure 7 illustrates this. The classical sequential approaches perform poorly for the examined series, which is the same as that in Figure 6 (only the first 99 elements of the series were used, though, because the full series exhibits a statistical "anomaly" which will be discussed shortly). In the first step, the best single change point from minimization of either s_r^2 or p (Eq. 2) is found to be at No. 39. To find the second change point, one has the 'classical sequential' options to either

- investigate the two sub-series of the first 39 elements and the other 60 elements
- or to adjust for the differences in the two sub-series' means and to repeat the test on the adjusted series.

Both tactics failed to identify a second significant change point (for the significance threshold p_t=0.01), which is to be attributed to the unfortunate positioning of change point No. 1 in an area where actually none is present.

By performing a kind of online analysis, a clear indication of six change points could be obtained for the series in Figure 7, even without resort to a global optimizer. The change-point positions do actually not deviate very much from the results of the GA shown in Figure 6. The procedure employed used the bit sequence representing the solution of the preceding run and investigated C (Eq. 4) for all possible change-point configurations with one or two additional change points on the right side of the last but one change point. The proceeding was motivated by the patterns in Figure 6, indicating that it is advisable to allow the last identified change point to vary slightly in its position or even to be withdrawn in favour of a better configuration. On the other hand, when a solution with *m* significant change points first appeared, the position of the *m*–1-th change point was definitively fixed. When, temporarily, no solution for these fixed change points could be found, no change-point configuration was plotted in Figure 7; the once accepted change points were nevertheless retained. This heuristic approach worked well in the case shown, which serves to illustrate how significant the benefits can be compared to classical sequential techniques. Yet, obviously, certain further refinements are necessary if such an algorithm were to be employed for operational applications.

A similar motivation is behind the sometimes followed practice of investigating partial series of prescribed length, hoping that in the shorter series only single change points will have to be dealt with. Depending on the relation between the chosen sub-series length and the distances between actual change points, such a procedure might either still have multiple change points in an investigated sample or neglect information on the involved populations, which would be useful for setting the change points precisely. It may therefore be necessary to try several sub-series lengths before arriving at firm conclusions (compare Moberg and Alexandersson, 1997). The approach introduced above is of a more dynamic character, adjusting itself to the actual lengths of sub-series with constant means.

Nevertheless, it is, of course, not the intention here to advocate this methodology too much since the GA shall still yield better solutions. Yet, even a perfect global optimizer, if it could be constructed, would derive benefit from online analyses, which comes from documentary diagrams such as those in Figure 6, right panel. In the case shown, it is particularly important to plot the change points as a function of the length of the series since in the final run the last included value pushed the rightmost change point below the significance threshold p_t=0.01. Merging the two populations also affected the significance level of the next change point to the left, and, eventually, even a third change point vanished. If one had only the analysis for the entire series, the signals of three change points – indicated fairly consistently in the analyses of the partial series – would be overlooked. In fact, by virtue of the simulation approach, we can state with certainty that the signals lost for length 100 *are* coupled with true change points.

Unfortunately, the temporary disappearance of signals shown is not a rare or particularly curious effect of random noise. In fact, the very same series yielded two more instances of such a pattern at lengths 81–82 and 93, but the GA, with its receptiveness to the addition of any number of change points at any stage, was in both instances capable to master the situation. It reinstated the temporarily lost change points soon afterwards. These examples alert that even when a statistically optimal result for the full series is known, the picture could nevertheless be incomplete. Valuable insight into the uncertainty inherent in the analysis is obtainable from the scrutiny of partial series.

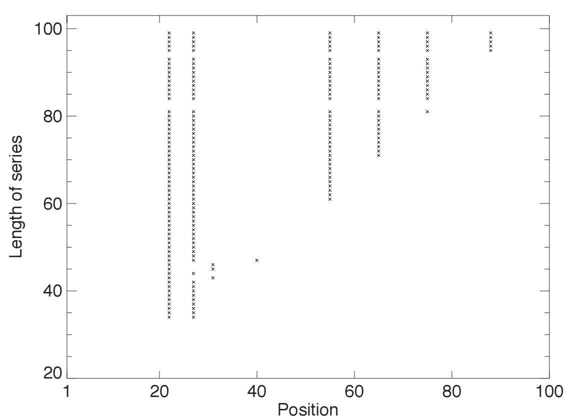

Figure 7. Results of change-point detection for the increasing length of the series; a sequential technique that adds one or two change points and allows for the disappearance of the rightmost change point of the preceding solution. The crosses mark the positions of the detected change points. The investigated series is the same as in Figure 6. For lengths where no solution is given, the significance of one or more change points temporarily fell below the prescribed threshold.

Slika 7. Rezultati pronalaženja točaka promjene za produženi niz; sekvencijalna tehnika koja dodaje jednu ili dvije točke promjene i uzima u obzir nestanak krajnje desne točke promjene iz prethodnog rješenja. Križići označavaju položaje pronađenih točaka promjene. Proučavani niz je isti kao onaj na slici 6. Za dužine kod kojih nema rješenja, signifikantnost jedne ili dviju točaka promjene je privremeno palo ispod praga.

7. CONCLUSION

The main argument behind the introduction of the GA method is that it targets on global optima and thus can be expected to yield results closer to the true solution than the simpler methods proposed in earlier works. That the GA matches the expectations has been demonstrated in the previous paper (Jann, 2000) for simulated, *perfect* data sets. It is, however, undoubtedly true that any method would have little value in climatology if it could not cope with deviations from the ideal case (outliers) or if it were incompatible with the wish to re-position change points on the basis of metadata records. Hence, these issues had to be addressed before application in practice is envisaged. It showed that a relatively straightforward inclusion of concepts dealing with outliers and metadata into the GA framework can be accomplished.

Perhaps, the universality of the GA technique should be emphasized: though the experiments presented generally made use of the *t*-test, none of the arguments in favour of the GA procedure, which is concerned with optimization aspects only, relied on the characteristics of this two-sample test. There is even no reason to limit oneself to cost functions comprising a two-sample test; successful experiments, where some genuinely global cost functions devoid of any two-sample concept were fed into the GA in a quite analogous fashion as Equations 4–6, have already been carried out:

- Rhoades and Salinger (1993) in their section 3.4 obtained the positions of a prescribed number of change points through minimization of s_r^2 (using dynamic programming for optimization). Leaving the question aside how to determine the "right" number of change points, one has almost the same expression as in Equation 4 that is minimized. Thus, results can be expected to be very similar, e.g. searching for six change points for the series in Figure 7, just one change point was shifted by three positions compared to the result for cost function (4) (that configuration exhibited one marginally insignificant *t*-value, hence the difference).
- Arising from Bayesian inference, Zurbenko et al. (1996) used a cost function

$$C = \frac{n}{2} \ln s_r^2 + \nu \ln n \tag{7}$$

(n: sample size). Like Equation 4, this expression – called the Schwarz criterion – combines the minimization of s_r^2 and a term to control the number of change points. This term is comparatively punitive on large numbers of breaks (which may be a desirable or disadvantageous feature depending on the application, cf. Menne and Williams, 2005), e.g. for the case in Figure 7, the magnitude of the shifts had to be increased from 1.5 to 2 in order to allow for a 6-breaks solution. With the cost functions being very similar in other respects, it is not surprising that the (independent) GA runs for Equation 4 and Equation 7, respectively, then yielded identical results.

- The (also Bayesian) approach of Caussinus and Mestre (1997) has a global scope, too, though in the practical implementation they confined themselves to a more limited view; later, the change-point detection part was reported to be amenable to global optimization by dynamic programming (Caussinus and Mestre, 2004). The full framework, featuring an implicit outlier detection, seems to be not approachable by this global optimization tool, though (Ibid.). The problem can, however, fairly easily be solved with the GA, by introducing a code '2' to indicate outliers and using the same GA mechanisms of mutation for the new *byte type* as for the change-point code '1'. Similar verifications of inhomogeneity/outlier mixtures as in Figure 3 were successfully carried out also for the Caussinus&Mestre cost function (jumps had to be larger and outliers more severe, however, since the penalty function is even more restrictive than the Schwarz formulation, cf. Menne and Williams, 2005). It shows that – since the GA poses virtually no requirement on the cost function – this optimization method gives the freedom to select even such an involved statistical formulation if that suits the posed homogenization problem best.

REFERENCES

Alexandersson, H., 1986: A homogeneity test applied to precipitation data. *J. Climatol.*, **6**, 661–675.

Alexandersson, H. and A. Moberg, 1997: Homogenization of Swedish temperature data. Part I. Homogeneity test for linear trends. *Int. J. Climatol.*, **17**, 25–34.

Basseville, M. and I.V. Nikiforov, 1993: Detection of abrupt changes – theory and application. Prentice-Hall, Englewood Cliffs, NJ, 556 pp.

Browning, K.A. and C.G. Collier, 1989: Nowcasting of precipitation systems. *Rev. Geophys.*, **27**, 345–370.

Buishand, T.A., 1982: Some methods for testing the homogeneity of rainfall records. *J. Hydrol.*, **58**, 11–27.

Buishand, T.A., 1984: Tests for detecting a shift in the mean of hydrological time series. *J. Hydrol.*, **73**, 51–69.

Caussinus, H. and O. Mestre, 1997: New mathematical tools and methodologies for relative homogeneity testing. In: *Proceedings of the first seminar for homogenization of surface climatological data, 6–12 October 1996, Budapest, Hungary*, 63–82.

Caussinus, H. and O. Mestre, 2004: Detection and correction of artificial shifts in climate series. *Appl. Statist.*, **53**, 405–425.

Ducré-Robitaille, J.–F., L.A.Vincent and G. Boulet, 2003: Comparison of techniques for detection of discontinuities in temperature series. *Int. J. Climatol.*, **23**, 1087–1101.

Easterling, D.R. and T.C. Peterson, 1995: A new method for detecting undocumented discontinuities in climatological time series. *Int. J. Climatol.*, **15**, 369–377.

Gille, S.T., 1997: Why potential vorticity is not conserved along mean streamlines in a numerical Southern Ocean. *J. Phys. Oceanogr.*, **27**, 1286–1299.

González-Rouco, J.F., J.L. Jiménez, V. Quesada and F. Valero, 2001: Quality control and homogeneity of precipitation data in the southwest of Europe. *J. Climate*, **14**, 964–978.

Herzog, J. and G. Müller-Westermeier, 1997: Homogenization of various climatological parameters in the German weather service. In: *Proceedings of the first seminar for homogenization of surface climatological data, 6–12 October 1996, Budapest, Hungary*, 101–111.

Herzog, J. and G. Müller-Westermeier, 1998: Homogenitätsprüfung und Homogenisierung klimatologischer Meßreihen im Deutschen Wetterdienst (in German). *Berichte des Deutschen Wetterdienstes*, no. **202.** Selbstverlag des Deutschen Wetterdienstes, Offenbach am Main, 27 pp.

Jann, A., 2000: Multiple change-point detection with a genetic algorithm. *Soft Computing*, **4**, 68–75.

Karl, T.R. and C.N. Williams Jr., 1987: An approach to adjusting climatological time series for discontinuous inhomogeneities. *J. Clim. Appl. Meteor.*, **26**, 1744–1763.

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. *Int. J. Climatol.*, **16**, 1197–1226.

Menne, M.J. and C.E. Duchon, 2001: A method for monthly detection of inhomogeneities and errors in daily maximum and minimum temperatures. *J. Atmos. Oceanic Technol.*, **18**, 1136–1149.

Menne, M.J. and C.N. Williams Jr., 2005: Detection of undocumented changepoints using multiple test statistics and composite reference series. *J. Climate*, **18,** 4271–4286.

Moberg, A. and H. Alexandersson, 1997: Homogenization of Swedish temperature data. Part II. Homogenized gridded air temperature compared with a subset of global gridded air temperature since 1861. *Int. J. Climatol.*, **17**, 35–54.

Peterson, T.C. and D.R. Easterling, 1994: Creation of homogeneous composite climatological reference series. *Int. J. Climatol.*, **14**, 671–679.

Peterson, T.C., D.R. Easterling, T.R. Karl, et al., 1998: Homogeneity adjustments of in situ atmospheric climate data: a review. *Int. J. Climatol.*, **18**, 1493–1517.

Potter, K.W., 1981: Illustration of a new test for detecting a shift in mean in precipitation series. *Mon. Wea. Rev.*, **109**, 2040–2045.

Rhoades, D.A. and M.J. Salinger, 1993: Adjustment of temperature and rainfall records for site changes. *Int. J. Climatol.*, **13**, 899–913.

Szinell, C., 1997: Methods for homogenization of data series. In: *Proceedings of the first* *seminar for homogenization of surface climatological data,* 6-12 October 1996, Budapest, Hungary, 9–17.

Whittaker, J. and S. Frühwirth-Schnatter, 1994: A dynamic changepoint model for detecting the onset of growth in bacteriological infections. *Appl. Statist.*, **43**, 625–640.

Wolter, K., 1997: Trimming problems and remedies in COADS. *J. Climate*, **10**, 1980–1997.

Zurbenko, I., P.S. Porter, S.T. Rao, J.Y. Ku, R. Gui and R.E. Eskridge, 1996: Detecting discontinuities in time series of upper-air data: Development and demonstration of an adaptive filter technique. *J. Climate*, **9**, 3548–3560.