Aquatic
Living
Resources

Detection of changes in time-series of indicators using CUSUM control charts

Benoit Mesnil[a] and Pierre Petitgas

IFREMER, Département EMH, BP 21105, 44311 Nantes Cedex 3, France

Abstract – Statistical Process Control (SPC) methods are extensively used in manufacturing contexts to monitor production processes. This article illustrates their potential utility for monitoring the state of marine ecosystems, using survey indicators. We exemplify the use of one SPC tool, the cumulative sum (CUSUM) control chart, to detect persistent changes in the state of a system as new observations are collected, using simulated and real data. Practical guidelines are given on how the chart parameters can be tuned to achieve an acceptable compromise between the ability to detect anomalies quickly and keeping the risk of false alarms low. The common performance measures associated with control charts depend on some key assumptions being met, and the potential impact of their violation is indicated.

Key words: Statistical process control / CUSUM / Aquatic resource monitoring

1 Introduction

Control charts are part of the statistical process control (SPC) tools routinely used over decades to monitor manufacturing processes and signal anomalies in performance (Montgomery 1991; Whetherill and Brown 1991; Derman and Ross 1997). A process has some inherent variability and is said to be "in-control" as long as its output remains within acceptable bounds. If an anomaly occurs, causing a shift in quality beyond the baseline variability, the system is said to be "out-of-control". Control charts are graphical displays of some summary statistic of the observed data (e.g. indicators) against the order index of the sample (e.g. time), together with reference "marks" based on the in-control mean and variance, that are designed to detect whether a worrisome change in process output is indicated by the current data and action is required to fix it. Since there are costs associated with both false alarms and quality losses in commercial products, the charts' parameters are tuned to achieve a desired trade-off between the risk of false alarm and the ability to detect changes promptly.

The terms of the detection/decision problem in manufacturing contexts, as just outlined, should sound familiar to whoever is involved in monitoring the status of fish stocks or marine ecosystems for advising managers. The latter typically expect that experts raise a timely signal when a worrisome change is occurring in marine resources, warranting corrective action, while requiring that the alarm is based on strong evidence. There are thus good reasons to believe that the SPC tools are relevant for natural resources management

(Scandol 2005). This paper aims to present the general principles of control charts and to provide practical guidelines for their application to fish stock indicators derived from survey data. The focus is on the cumulated sum (CUSUM) chart (Page 1961), which is a running total of deviations from a reference value.

There is an extensive literature on control charts, and they are still a very active field of research. However, this paper is not intended to discuss the current theoretical findings, but simply to pursue awareness raising among marine scientists as initiated by Nicholson (1984) or Scandol (2003). For that purpose, only the "decision interval" (DI-CUSUM, also known as "tabular") form of the CUSUM is considered. That form is advocated in SPC textbooks (e.g. Hawkins and Olwell 1998) for the type of data considered in the survey indicators context. Control charts can be designed to monitor changes in mean level (location charts) or in variance (scale charts) of process outputs; explanations are only given for location charts here.

2 Method description

Suppose a suite of observations x_i collected at time $i = 1, \ldots m$ and assume that their in-control mean μ and standard deviation σ are known from a pilot study or for a reference period. In the following, it is considered that the data are first standardised through the transformation $z_i = (x_i - \mu)/\sigma$. Deviations from μ that will be monitored are then in units of σ, which is convenient.

The decision-interval CUSUM works by recursively accumulating positive and negative deviations separately with two

[a] Corresponding author: Benoit.Mesnil@ifremer.fr

statistics:

$$S_i^+ = \max\left[0, S_{i-1}^+ + z_i - k\right]$$

for positive deviations ("one-sided upper CUSUM"), and

$$S_i^- = \min\left[0, S_{i-1}^- + z_i + k\right]$$

for negative deviations ("one-sided lower CUSUM"), with starting values normally set as $S_0^+ = S_0^- = 0$. A CUSUM chart is obtained by plotting these statistics against i. If measurements tend to stay above the in-control mean, the upper CUSUM S^+ develops an upward trend; likewise, the lower CUSUM S^- shows a downward trend if observations are consistently below the mean.

The chart's parameter k is usually called the *reference value*, or the *allowance*, and is related to the size of the smallest shift in the level of z that one is wishing to detect quickly. Note that deviations smaller than k are ignored in the recursions above. The decision rule is to declare an out-of-control state whenever S^+ exceeds the *decision interval h* or S^- falls below $-h$. The values chosen for the parameters h and k, both being measured in standard deviation units, determine the performance of the control chart. There is no theoretical objection against setting different h–k pairs for upper and lower CUSUM's if changes in one direction matter more than in the other. The CUSUM chart together with its k and h parameters define an SPC monitoring scheme.

The performance of control charts is generally evaluated in terms of their run length. A run is the number of sampling events that elapse between the start of the monitoring and the first alarm. The run length is a random variable whose probability distribution depends on the process and the chart parameters. Its expectation – called Average Run Length (ARL) – is commonly used as a summary measure of performance. The notation ARL(δ) is used to designate the ARL of an SPC scheme for detecting a change of size δ (in σ units) occurring in the process mean level. Thus, ARL(0) is the ARL of a scheme when the process actually stays in-control all the time (in-control, or IC ARL); yet, due to its inherent variability, an alarm may be raised by chance alone when the chart is updated with a new datum. In other words, ARL(0) is the average time until a false alarm is raised, which should ideally be large. Conversely, if the mean of the process distribution shifts from μ to $\mu + \delta$, due to an anomaly, the chart should detect this quickly, implying a short ARL(δ). Note that the run length distribution is often very broad and skewed, even when the underlying processes is strictly gaussian, and it may be misleading to only consider its mean. Therefore, the experts (e.g. Luceño and Puig-Pey 2002) recommend to also look at other percentiles, whenever the distribution can be computed. Run-length distributions can be obtained by simulations (e.g. Jun and Choi 1993) or by formal integration (e.g. Hawkins and Olwell 1998) for various statistical distributions of the process belonging to the exponential family. Chart parameters can be tuned to achieve a desired compromise between a long ARL(0) and a short ARL(δ) – i.e. low false alarm rate vs. fast detection abilities. Procedures for tuning a CUSUM scheme are explained hereafter in the section on practical guidelines.
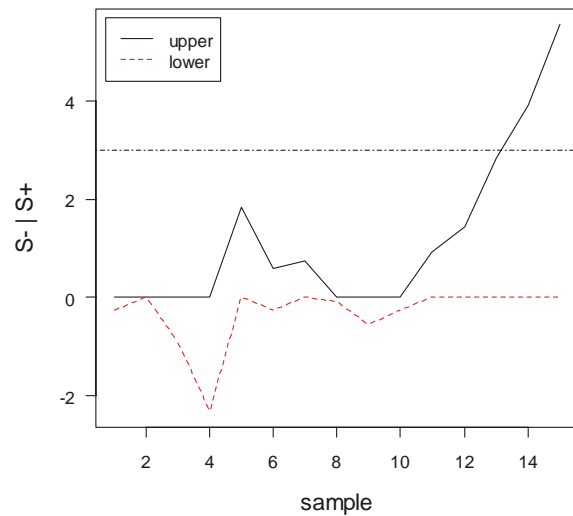

Fig. 1. Example of a DI-CUSUM control chart using simulated data. A shift in process mean was supposed to occur starting from sample 11. S^+ and S^- are the upper (solid line) and lower (dashed line) CUSUM. The horizontal dash-dot line indicates the decision limit h: an alarm is triggered on taking sample 13 which causes the CUSUM to trip that limit.

3 Example

Using simulated data will serve the purpose well. Let us assume that the in-control mean of the fake process is 5 and ten random samples are drawn from a normal distribution $N(5,1)$. From sample 11 onwards, an anomaly is supposed to cause a shift in the mean to 6 and samples are drawn from $N(6,1)$. A value of 0.5 is adopted for the allowance k and the S statistics are computed as new data occur. The CUSUM control chart obtained after sample 15 is shown in Fig. 1. For illustration, the decision limit h was set rather high (3) to guard against false alarms (ARL(0) with these settings is about 120). Nevertheless, it can be seen that the upper CUSUM trips that limit at sample 13, only two time steps after the occurrence of the anomaly, which can be regarded as a good performance. Of course, h could be reduced to speed up detection, but at the cost of higher chance of false alarms. As can be seen on the figure, an h smaller than 2.0 would have resulted in (false) alarms in the earlier period although, by construction of the simulation, the system was in control. It is important at this point to emphasise that "in-control" does not mean "perfectly stable"; in that state, a system may vary due to random fluctuations, but within limits that are considered to be natural in the judgement of the assessor.

The case of the North Sea cod stock provides another example, using real data. The case study considered in the FIS-BOAT project used survey data from the International Bottom Trawl Survey in the first quarter in 1985-2005. Experts agreed to select the first ten years (1985-1994) as the reference period, for computation of the in-control μ and σ. The indicator shown in this example is the survey index of total abundance (all sizes/ages of cod combined); it is log-transformed to approach normality. Following the procedure described in Sect. 5.1, the allowance k was set at 1.3 and the decision limit h at ±1.0 (see Petitgas et al. 2009). The resulting control chart
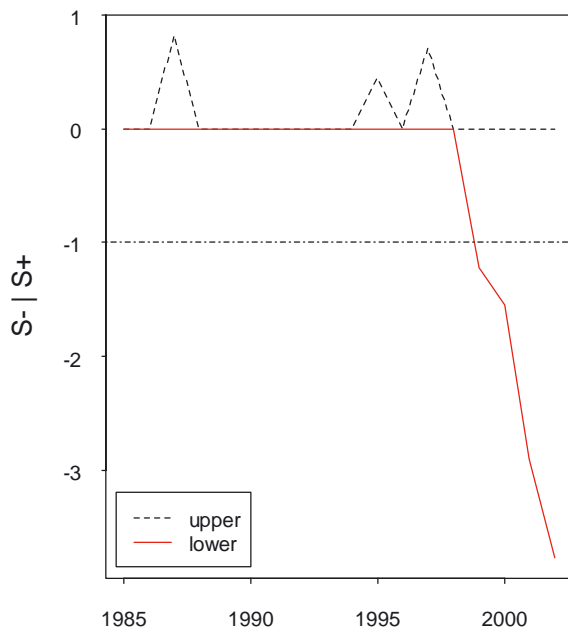
Fig. 2. DI-CUSUM of total-number indices of North Sea cod from the International Bottom Trawl Survey. The decision limit was set at –1 (dash-dot line) and an alarm might have been raised in 1999. The graph is clipped at 2002 to improve legibility, as the lower CUSUM plunges steeply until 2005.

is shown in Fig. 2. The lower CUSUM gives a clear signal that stock abundance is decreasing from 1999 onwards (the steep dive continues to 2005). The upper CUSUM only shows transient and inconsequential fluctuations. The state of stock indicated by the CUSUM is consistent with assessments carried out by ICES (e.g. 2003), based on catch-at-age analyses, which show that the spawning stock has consistently been below the limit reference point since 1999. However, due to arguments about misreporting of catches and about potential biases due to the inclusion of commercial catches per unit effort in the analyses, it is only in 2001 that ICES advised a closure of the directed fishery. Perhaps, consideration of the survey data in the formal framework of control charts might have permitted strong corrective action to be advised earlier.

4 Software

Attached to this paper on-line are two scripts in the R language (R Development Core Team 2005), developed for the purpose of the EU project FISBOAT, which may help readers to implement a CUSUM monitoring scheme:

- CusumTutorial.r is generic, for exploring CUSUM charts with "free-format" time series vectors;
- FBCusumCharts.R is designed to automate the production of standard tables of results for the FISBOAT report ("traffic light template"), with input from the standard case studies files (the data must comply with a specific format with standard names etc.).

Both use a set of functions stored in the separate file Cusum-Funcs.r that must be sourced into the user's R workspace (on

first use) as instructed in the scripts. The scripts are meant to be run in a stepwise fashion (select a line or a block and submit to R console) and are amply commented to guide the user. The scripts only use basic R commands and do not require any special library.

FBCusumCharts.R proposes to run CUSUM monitoring schemes on a set of indicator time series so as to construct a table of CUSUM alerts with set risks of false and non-alarms. The top part of the script deals with each indicator in turn. Note that a logarithm transformation is applied to the Survey and Recruit indices (Cols. 5 and 6 of data sets); the agreed reference period for each case study is "hard-wired" but can be edited if needed; an indication of an appropriate value for the allowance k, based on the mean deviation from the reference mean outside the reference period, is proposed but is not coded as a default value. Once the full set of indicators has been processed, the bottom part of the script gathers the individual resnam.# objects to produce the table of alarms (signed CUSUM values above h or below $–h$) and the table of CUSUM parameters and saves them to files.

This implementation includes functions to compute in-control or out-of-control ARLs and run length distributions of one-sided CUSUM for normal data without auto-correlation, adapted from a Fortran code by Gan (1993) found on the StatLib JQT archive (http://lib.stat.cmu.edu/). They have been checked against the values tabulated in various SPC textbooks and articles, and the results compare well.

5 Practical guidelines

5.1 CUSUM design: tuning the chart parameters *k* and *h*

With fisheries survey data, the task is usually to analyse time series of one or several indicators of population status (control variables). Often, only one value per indicator and per year is available (individual data), sometimes with the precision on the indicator in each year. We distinguish two phases: Phase I for defining the in-control (IC) state; and Phase II for designing the CUSUM to signal changes from the in-control state with desired performance.

Phase I

The task in Phase I is to set the IC or reference process parameters μ and σ. This is a critical phase in that the values adopted for these parameters will condition signals from the analysis. Normally, this is an experimental phase where the state of the system is closely checked, many measurements are taken and scrutinised, to retain only those that can be safely assumed to correspond to a well behaved process. In our case, we will often start with existing data collected in the past, and Phase I will essentially consist in the definition/choice of an in-control (or reference) period, and then using the data in the selected years to estimate in-control parameters μ and σ. The IC period can be defined on various criteria, including an inspection of patterns in the times series ("eyeballing"). However, it is best defined collectively, on expert knowledge, as

the period in which the population was in a satisfactory state and/or showing satisfactory dynamics. Thus, the IC period may not necessarily comprise consecutive years. Sensitivity to the IC period should be analysed and the IC period may be also re-defined a posteriori. This is consistent with the iterative and rejection procedures described in SPC textbooks (e.g. Montgomery 1991).

Phase II

In Phase II, the task is to design (or tune) the CUSUM scheme to signal a specified deviation from the IC mean with a desired performance, i.e. this is where the chart parameters k (allowance) and h (interval) are determined. The choice of k is based on the magnitude of the shift δ in the mean that makes "a meaningful impact" on the system, resulting in out-of-control state. The value of h determines whether an alarm is raised or not (an alarm is triggered when the CUSUM plot crosses the horizontal line at h, or at $-h$ or $+h$ for a two sided CUSUM). The rationale for choosing h is primarily based on minimising the risk of false alarm, but the ability to promptly detect shifts that matter should also be preserved. Setting h, once k is chosen, involves run-length (RL) considerations. Once the indicator series is standardised, a four-step procedure is suggested for tuning the monitoring scheme:

1. *Define the allowance parameter k.* If δ is the shift of interest (in standard deviation units), there is a broad support in the literature for setting k at half the value of that shift (formal demonstration in Chap. 6 of Hawkins and Olwell 1998), and this rule can be safely adopted. The "meaningful" shift δ can be set after analysing the deviations from μ outside the IC period. For instance, the shift to be detected can be set to a percentile of these deviations or to their mean. For fish population indicators based on surveys, k will generally take a value between 0.5 and 1.5; too small values of k should be avoided (Hawkins and Olwell 1998, p. 33).
2. *Define the decision parameter h.* Using published tables or our R software with a zero value for the shift δ, search for an h that gives desirably large IC ARL(0) given k, and thus a low risk of false alarm. Larger values of h (and k) lead to larger ARLs.
3. *Check the IC run- length distribution.* Because the RL distribution may be quite skewed (Fig. 3), consideration of the average RL alone may be misleading and, using the function arldis.f in our R scripts, the full distribution of the in-control RL should be checked. For example, if k and h are chosen to aim for a "large" IC ARL of 100, and a "small" value of 10 samples or less for the 25th percentile is observed, it is likely in the actual application of the scheme that more false alarms will occur than the large ARL(0) would indicate. In such cases, h should be increased.
4. *Check the OC ARL.* The h value arrived at in the previous steps may have provided a reasonable bound on the probability of false alarm, but you then need to ascertain that the scheme is able to quickly detect the shifts you are interested in, i.e. that its out-of-control ARL is small enough. Return to the ARL tables or software with a value of $2k$ for
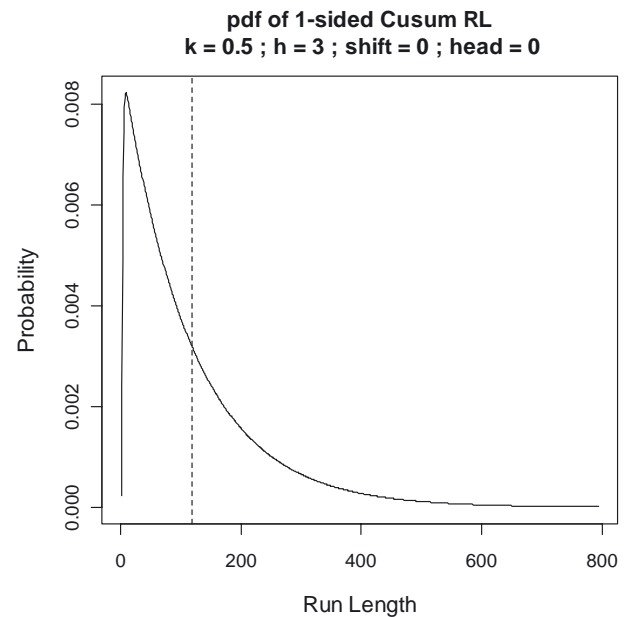
Fig. 3. Example of an in-control run length (RL) distribution showing its typical skewness and long "tail". The parameters h and k were set at the same values (in standard deviation units) as for the CUSUM chart in Figure 1. An average of 120 (= ARL(0), dashed vertical line) is expected with these settings; yet, the chance of a false alarm occurring within 36 samples or less is 25%.

the shift δ, and check that the OC ARL(δ) corresponding to the envisaged h is adequately small. In general, it is possible to find values of k and h such that the OC ARL does not exceed 3 years.

As pointed out in the introduction, the basic challenge of using and tuning SPC schemes is to find an acceptable compromise between the risk of false alarm and the ability to detect shifts that matter in the state of the system, and it is often necessary to iterate through these 4 steps to arrive at that compromise. The notions of "meaningful impact", "acceptable risk" and "desired performance" are very much policy issues and have to be decided in partnership with managers and stakeholders.

5.2 Assumptions and effects of violations

The main assumptions underlying the statistical properties of CUSUM charts are:

(i) that the monitored variable has a distribution from the exponential family; in particular, the run length characteristics commonly tabled in textbooks or computed with our R software are only valid for normally distributed data;
(ii) that the in-control process parameters are known rather than estimated; and
(iii) that the time series of residual variation is not auto-correlated.

The effects of violating these assumptions are all in the same direction: the in-control ARL(0) experienced in practice is shorter than the value computed for the perfect case, i.e. the chances of false alarms are larger than expected (e.g. Sect. 3.7

in Hawkins and Olwell 1998; Jones et al. 2004; Lu and Reynolds 1999; Reynolds and Stoumbos 2004). Smaller values of k (also large h) enhance the robustness to non-normality, but increase the impact of estimating the reference μ and σ from the data. An encouraging note: even though a CUSUM tuned with a given k is optimal for detecting shifts of $2k$ standard deviations, its performance remains high for actual shifts that are 'not too far' (Hawkins and Olwell, p. 54).

Time series of population indicators from surveys are often short (< 20 years) with marked variations and sometimes show correlation or trend. It is advised to check the distribution of the indicator variable as well as its correlation in time. It may be necessary in some cases to transform the variable into a Gaussian or to detrend the time series. Short series imply that data for the reference period are even shorter, and we use noisy data to estimate the IC process parameters. Since all departures from the assumptions will result in effective RLs being very different (in general shorter) than values publicised for the "clean" case, an *ad hoc* remedy is to take relatively large h values. A conservative advice is to use (k, h) parameters giving large IC RLs: ARL > 20 years and 25th percentile of RL distribution > 10 years. When some deviations from μ outside the reference period are large in comparison to σ, it may be telling that the variance has changed or that the indicator variable is skewed. In that case, starting Phase II with a large value of k is advisable. When the value of h is small in comparison to an increasing (decreasing) CUSUM deviation, it is likely that there is auto-correlation in the indicator time series.

5.3 Strengths and weaknesses

Control charts have been in operation in many branches of industry since the 1930's and their statistical bases have been thoroughly investigated in a large body of literature (the references below are just a tiny sample). Applications have been extended to environmental surveillance (e.g. Manly 2001), biomedicine (e.g. Chang and McLean 2006), clinic tests, public health (e.g. Starks and Flatman 1991). The strenghts in these domains are that the in-control state is well defined (often with reference to agreed norms), the monitoring involves numerous samples taken at high frequency through rigorous sampling designs, and measurement errors are often small.

In contrast, this defines the weaknesses in fisheries applications. The main limitation is our poor ability to characterise the reference state of fisheries (or of ecosystems) with survey data that just span the recent decade(s) in a background of large variability compounded by substantial sampling variance. The IC state is "estimated" based on a finite set of existing data, without possibility of replication, instead of being determined through planned experiments (industry) or with reference to norms (pollution, health). However, the reference state does not imply perfect stability and the process may show substantial variability even when deemed well behaved. The goal of control charts is to pinpoint those events where the state of the system deviates beyond the domain of its inherent variability.

The CUSUM approach does not presume the nature of the change (linear, trend or otherwise) and treats positive and negative deviations equally. The CUSUM chart is best suited

to detecting small persistent changes. Anomalies in the system can take the form of shifts in the mean and/or changes in the variance of the distribution. Specific control charts can deal with both situations. It is common to combine location and scale charts to enhance the detection performance for both small and large shifts (Reynolds and Stoumbos 2004). Hawkins and Olwell (1998, p. 67) suggest an approximation to obtain scale charts with individual data, which can be plotted together with location charts.

Moustakides (1986) demonstrated that, among the SPC schemes that have similar in-control ARL(0), the CUSUM scheme has the smallest expected time until a change is detected when it occurs. This is the basis of the rationale for tuning the chart, with priority given to achieving large ARL(0). The emphasis on low risk of false alarm, combined with the charts' property of filtering natural variability out, has some practical advantage in our application to fisheries management: casting assessment noise straight into fisheries regulation has been damageable to the credibility of scientists, so a method that explicitly aims at avoiding this is welcome, at least in the ICES context of advising the EU annual management scheme.

Lastly, the biggest advantage of a CUSUM monitoring scheme is that it is very simple to implement, using a variety of indicators individually or in combination. Yet, it provides a formal and statistically sound framework to monitor the status of systems (e.g. fish stocks or ecosystems) in a transparent way.

Supporting information

Text S1
CUSUM Software
CusumTutorial.r
CusumFunct.r
FBCusumCharts.R

Supporting Information is only available in electronic form at www.alr-journal.org.

References

Chang W.R., McLean I.P., 2006, Cusum: a tool for early feedback about performance? BMC Medical Res. Methodol. 2006, 6:8 (http://www.biomedcentral.com/1471-2288/6/8)

Derman C., Ross S.M., 1997, Statistical aspects of quality control. London, Academic Press.

Gan F.F., 1993, The run-length distribution of a cumulative sum control chart. J. Qual. Technol. 25, 205–215.

Hawkins D.M., Olwell D.H., 1998, Cumulative sum charts and charting for quality improvement. New York, Springer Verlag.

ICES, 2003, Report of the ICES Advisory Committee on Fishery Management, 2003. ICES Coop. Res. Rep. 261, 3 volumes.

Jones L.A., Champ C.W., Rigdon S.E., 2004, The run length distribution of the CUSUM with estimated parameters. J. Qual. Technol. 36, 95–108.

Jun C.-H., Choi M.S., 1993. Simulating the average run length for CUSUM schemes using variance reduction techniques. Communications in Statistics B: Simulation 22, 877–887.

Lu C.-W., Reynolds M.R., 1999, Control charts for monitoring the mean and variance of autocorrelated processes. J. Qual. Technol. 31, 259–274.

Luceño A., Puig-Pey J., 2002, Computing the run-length probability distribution for CUSUM charts. J. Qual. Technol. 34, 209–215.

Manly B.F.J., 2001, Statistics for environmental science and management. Boca Raton, Chapman & Hall/CRC.

Montgomery D.C., 1991, Introduction to statistical quality control. 5th Edition 2005. New York, Wiley.

Moustakides G.V., 1986, Optimal stopping times for detecting changes in distributions. Ann. Stat. 14, 1379–1387.

Nicholson M.D., 1984, Some applications of CUSUM techniques in fisheries research. Int. Counc. Explor. Sea, CM 1984/D:5, 10 p.

Page E.S., 1961, Cumulative sum control charts. Technometrics 3, 1–9.

Petitgas P., 2009, The CUSUM out-of-control table to monitor changes in fish stock status using many indicators. Aquat. Living Resour. 22, 201–206.

R Development Core Team, 2005, R: A language and environment for statistical Computing. Vienna, Austria, R Foundation for Statistical Computing. URL http://cran.r-project.org/

Reynolds M.R., Stoumbos Z.G., 2004, Control charts and the efficient allocation of sampling resources. Technometrics 47, 409–424.

Scandol J.P., 2003, Use of cumulative sum (CUSUM) control charts of landed catch in the management of fisheries. Fish. Res. 64, 19–36.

Scandol J.P., 2005, Use of quality control methods to monitor the status of fish stocks. In: Kruse G.H., Gallucci V.F., Hay D.E., Perry R.I., Peterman R.M., Shirley T.C., Spencer P.D., Wilson B., Woodby D. (eds.), Fisheries Assessment and Management in Data-Limited Situations. Alaska Sea Grant College Program, University of Alaska Fairbanks, pp. 213–233.

Starks T.H., Flatman G.T., 1991, RCRA ground-water monitoring decision procedures viewed as quality control schemes. Environ. Monitoring Assess. 16, 19–37.

Whetherill G.B., Brown D.W., 1991, Statistical process control: theory and practice. London, Chapman & Hall.

Woodall W.H., Adams B.M., 1993, The statistical design of Cusum charts. Qual. Eng. 5, 559–570.