Document Title: Drawing Conclusions from Monitoring Data

Author(s): George Lindsey

Date(s): April 1996

Document Type/Physical Description: Type-written, 11pgs (includes 5 graphs)

Fonds/Collection Name: George Lindsey Fonds

Series: NACD Verification

File/Box Number: 3/1

Original Archival Reference: N/A

Item Description: This document is a typed short paper that Lindsey wrote as a Senior Research Fellow at the Canadian Institute of Strategic Studies.

Keywords: verification; arms control agreements; Gulf War; games theory; statistical sampling; surveillance; nuclear weapons; monitoring problems

**When citing material that has been digitized from other archives, the LMH Archive encourages users to cite the original reference information provided in the above field.*

DRAWING CONCLUSIONS
FROM
MONITORING DATA

by

GEORGE LINDSEY

CANADIAN INSTITUTE OF STRATEGIC STUDIES

OTTAWA

APRIL 1996

DRAWING CONCLUSIONS FROM MONITORING DATA

by George Lindsey

THE CHANGING REQUIREMENTS FOR VERIFICATION

During the last few decades nations have been attempting to improve their security through arms control agreements, freely undertaken by states which still harboured deep suspicions regarding the future actions of some of the other parties to the treaty. To build trust in future mutual compliance with its provisions, it has become necessary to include in the agreement provisions for verification. To an increasing degree, parties with a genuine desire for effective arms control have been willing to agree to measures of verification requiring cooperation and involving a considerable degree of intrusiveness.

New challenges to security are now presenting new types of situation which differ from the ones described above. One concerns unilateral arms control, in which a strong party (or, as in the case of the Gulf War, a victorious coalition) coerces a weaker one into undertaking measures of disarmament, in the verification of which the latter may or may not elect to cooperate

Another, quite different type of challenge can be presented not by armaments, but by nature, or by dangers caused by human activities which, while not intended to influence opponents or cause harm, do in fact present a threat to the well-being of the global community.

In order to deal with either of these two "non-traditional" types of challenge it is necessary to ascertain just what is occurring, without having to rely on the cooperation of the opponent. Apart from the collection of collateral intelligence, and the observations made during inspections, this usually requires monitoring by remote sensing techniques, and assembly of the information collected by the sensors. Finally, all of the available data needs to be combined and analyzed, in order to arrive at a logical conclusion as to the nature of the activity believed to constitute a threat. The purpose of this brief paper is to describe some mathematical techniques which can serve to make the best use of all of the data that has been collected.

STATISTICAL SAMPLING TECHNIQUES

There will be few practical situations in which it will be possible to monitor 100% of the relevant objects or activity on a continuous basis, whether by remote sensing or close inspection. It will be necessary to depend on incomplete sampling, although it will usually be possible to direct the monitoring effort toward selected targets, and sometimes at selected times, but never all of the targets all of the time. Two important strategic decisions that have to be addressed are the planning of the sampling operations and the conclusions that should be inferred from analysis of the resulting data.

THE THEORY OF GAMES

If parties who nurture suspicions regarding the ultimate intentions of eachother negotiate an arms control agreement, with provisions for verification, the process of verification can be likened to a two-person game. Each should assume that the other party is an intelligent and malevolent opponent, who will act in such a manner as to obtain maximum advantage for himself. The theory of games is designed to discover the strategy most likely to produce the best possible results in spite of the opponent's most intelligent actions.

In the case of dangers caused by nature, which is not an intelligent and malevolent opponent, the situation can be described as a one-person game. There is only one player who has to make decisions and take actions, but he must make his decisions without full knowledge of the situation. In this case, the statistical techniques that have been designed for quality control can indicate the strategic options open to the monitoring party.

The example of environmental pollution caused by human activity can be a one- or a two-person game, since the monitoring party may wish to identify the source as well as the existence of the pollution, while the source may take steps to avoid identification.

VERIFICATION OF ARMS CONTROL AGREEMENTS

The application of game theory and statistical analysis to the problem of concluding whether the observations indicate that the opponent is complying with his undertakings, or not, begins with the assumption that the objectives of verification are:

-if the opponent <u>is complying</u> with the agreement,
(A)-always to conclude that he is complying
(B)-never to conclude that he is violating

-if the opponent is violating the agreement,
(C)-always to conclude that he is violating
(D)-never to conclude that he is complying.

Incomplete samples (less than 100%) do not permit error-free conclusions (with 100% confidence that they are correct), but, assuming that the actual number of limited objects does not change over the period in which monitoring is conducted, increasing the size of the sample increases the probability of achieving all four of the objectives (although at the cost of a longer program of sampling).

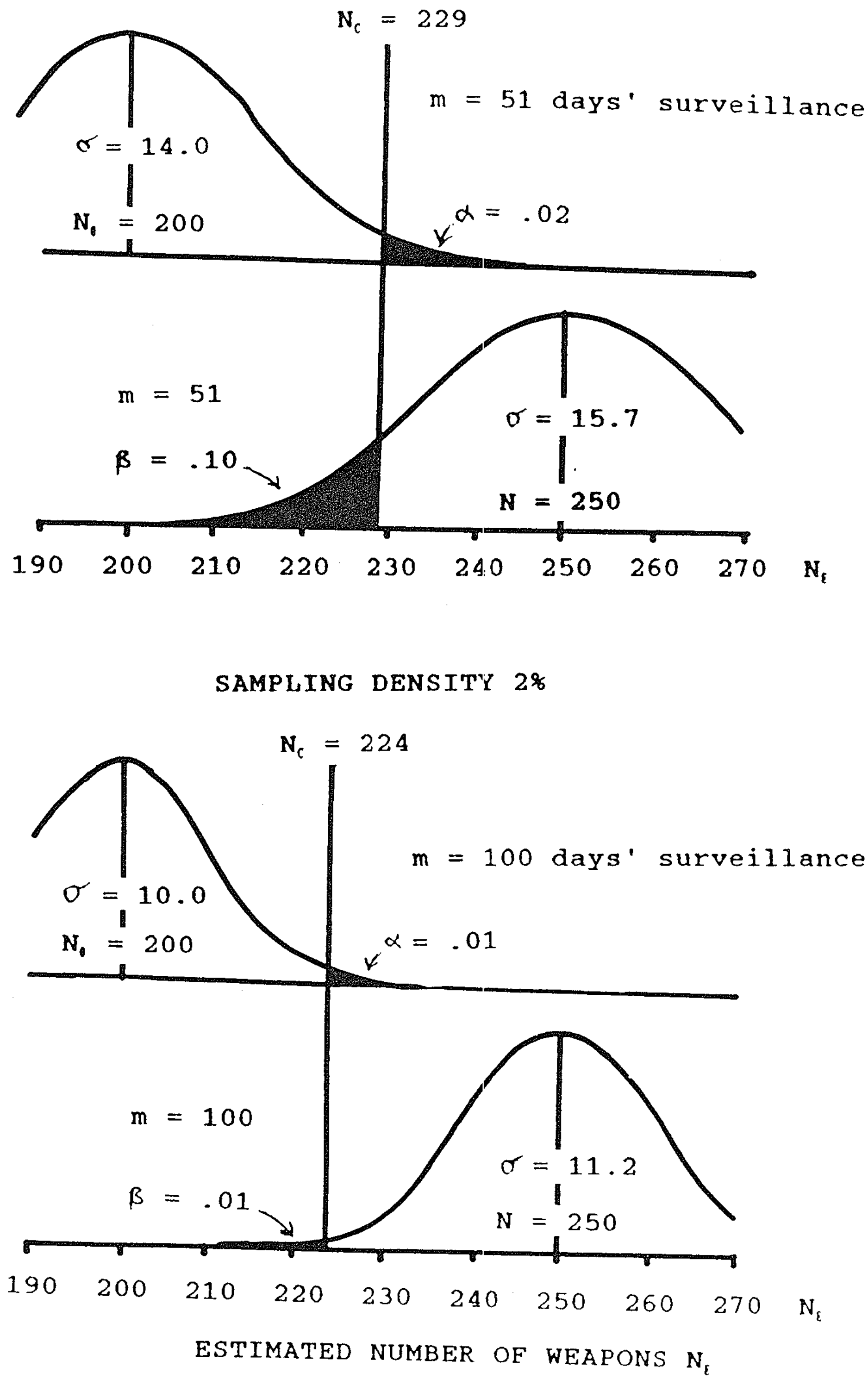
Failures to Detect Violations and False Alarms

Inference (D), the failure to detect an actual violation, is clearly undesirable to the monitoring party. The probability of its occurrence can be reduced by altering (making less severe) the criterion for concluding that a violation has occurred. However, this will increase the probability of arriving at Inference (B), the false accusation of cheating, which is also undesirable. The proper selection of the criterion needs to be made with a decision as to the relative penalties incurred by the two undesirable mistakes. This, however, is a strategic and political question which cannot be answered by statistical analysis.

NUMERICAL EXAMPLE

An imaginary example to illustrate the applications of mathematical techniques can be given by the use of a photographic satellite to monitor a field of mobile objects, of which the number, N_0, that is allowed to be deployed in a given area, has been limited by an arms control agreement. Suppose that the satellite is only able to observe a small sample of the relevant area every day, but when the visibility conditions are met it will detect all of the missiles within its field of view. After a few days of observation it will be possible to give a very imprecise estimate, N_E, of the number of objects actually deployed, and (if the actual number N does not change) the confidence level in the accuracy of the estimate N_E will slowly improve with each subsequent observation. If the actual number of objects, N, does suddenly increase, there will be a delay before the estimated number N_E rises to a value high enough that it can be inferred with high confidence that there has indeed been an increase in the true number N.

Figure I has been calculated for a case where $N_0 = 200$ mobile missiles are permitted anywhere in a specified area, and



THE SHRINKING OF α AND β AS MORE OBSERVATIONS ACCUMULATE

FIGURE I

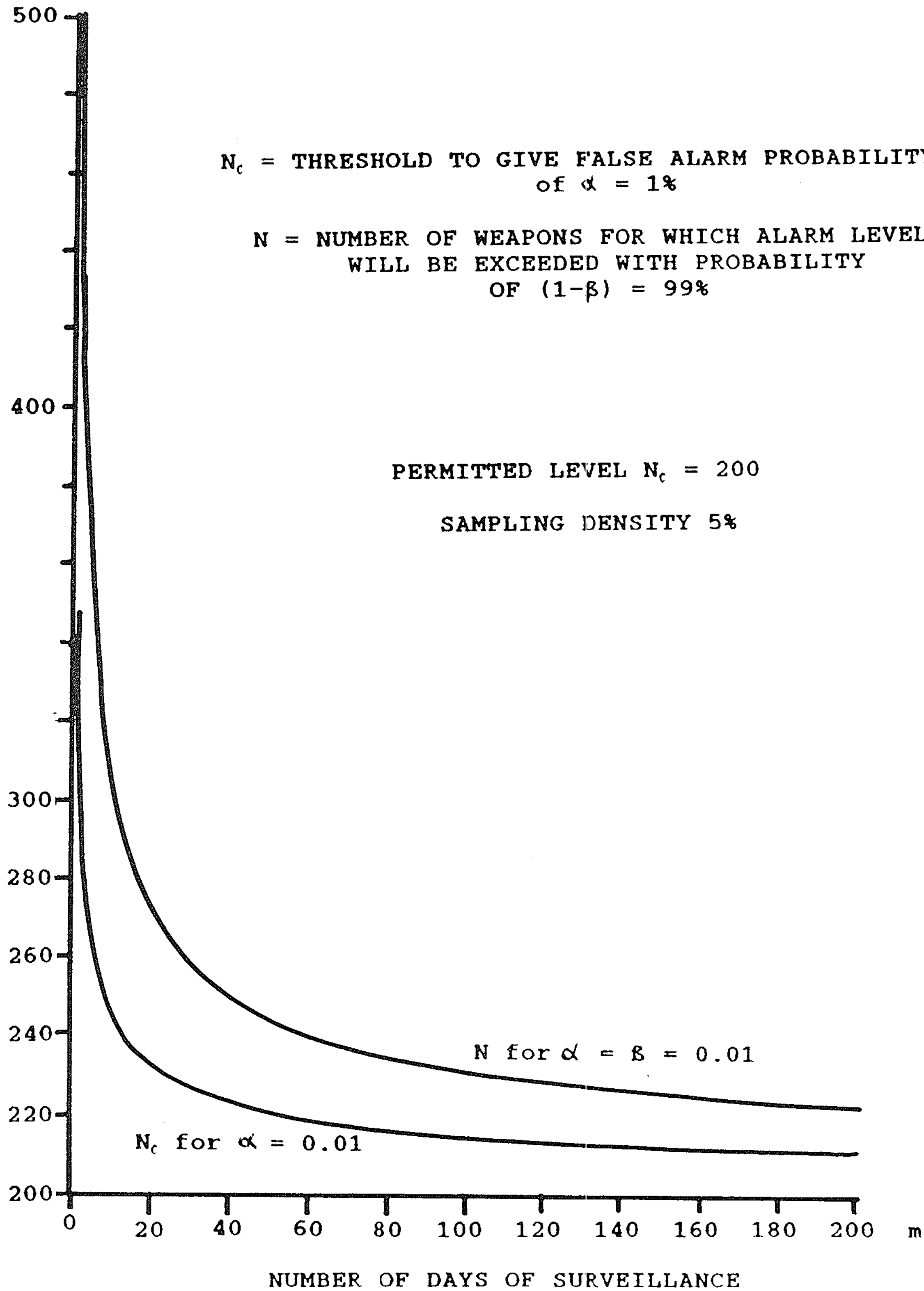

Nc = THRESHOLD TO GIVE FALSE ALARM PROBABILITY of α = 1%
N = NUMBER OF WEAPONS FOR WHICH ALARM LEVEL Nc WILL BE EXCEEDED WITH PROBABILITY OF (1-β) = 99%
PERMITTED LEVEL Nc = 200
SAMPLING DENSITY 5%
N for α = β = 0.01
Nc for α = 0.01
500
400
300
280
260
240
220
200
0
20
40
60
80
100
120
140
160
180
200
m
NUMBER OF DAYS OF SURVEILLANCE

FIGURE II

degrees of violation of an agreement to keep the number of missiles to $N_0 = 200$. For this example the sampling density is taken to be 5%. The lower curve of Figure II shows the alarm level N_C, calculated to give probability $\alpha = 1\%$ of being exceeded if $N = 200$. As the number of days of surveillance m increases N_C is reduced, falling to 233 after 20 days and 215 after 100 days. But the true number N could be as high as 233 without the probability that the estimate N_E would exceed $N_C = 233$ being as high as 50%, and for the probability of (correctly) inferring a violation to be as high as 99% it would be necessary for the real number N to be 271. The upper curve indicates the value of N corresponding to this 99% probability for different durations m of surveillance. Surveillance extending over 100 days, for which $N_C = 215$, would give a 99% probability of signalling a violation if $N = 231$. The shape of the curves shows that the capability of the system improves rapidly during the first 50 days, but comparatively slowly thereafter. To cite other values, 5 days would be insufficient to detect (with 99% confidence) levels of N below about 400; it would take 10 days before a deployment of 304 weapons would cause a violation to be inferred, and 30 days for 258.

A decision as to what level of violation constituted a serious threat and required some form of counteraction, takes us beyond the realm of mathematics.

COMBINING STATISTICAL DATA WITH OTHER INFORMATION

Classical statistics allows us to obtain estimates of numbers, based entirely on measurements, together with an indication of the accuracy of the estimates. Usually these measurements are all of the same type, although statistical techniques allow for cases in which some measurements differ from others in their accuracy.

In the practical conduct of verification of an arms control agreement, a situation may arise in which auxiliary information, possibly not based on measurements, perhaps not even quantitative at all, is available. But we may wish to make use of this information, probably incorporating it into the eventual conclusion as to what the real numbers are.

There is a theory involving subjective probabilities and Bayesian statistics that can be applied to problems of this type. The application of the method can be illustrated starting with the example already described.

The first step is to calculate the probabilities prior to using the measurements that the measurements will result in an estimated number N_E. The top three distribution curves on Figure III show these probabilities, using the assumed probabilities for the three hypotheses A, B, and C. The top diagram shows the three bell-shaped curves of the normal distribution, centred on the estimated numbers 200, 250, and 300, with areas proportional to the prior probabilities 0.6, 0.1, and 0.3. It can be seen that only hypothesis A is at all likely to produce an estimate N_E less than 220, only B an estimate between 235 and 260, and only C an estimate above 280. But either hypothesis A or B could produce a value of N_E between 220 and 235, and either B or C could produce an N_E n the range $260 < N_E < 280$.

The bottom three distribution curves show the posterior probabilities that hypotheses A, B, or C is true, after using the information that the measurement has produced its estimate N_E. For example, the curve labelled $Pr\ (A \mid N_E)$, the notation indicating the conditional probability than A is true given that measurement estimates that $N = N_E$, indicates that for $N_E < 215$ the probability that hypothesis A is true is virtually a certainty, but in the range $215 < N_E < 240$ it drops to nearly 0. At $N_E = 228$ the posterior probability that hypothesis B is true exceeds that for A. But when $N_E = 272$ it is more probable that C, rather than B, is true.

It can be seen that, if $0 < N_E < 225$, one can infer that $N = 200$; if $230 < N_E < 265$, N is probably 250 (almost certainly 250 if $240 < N_E < 260$) ; while if $N_E > 275$, then N can be inferred to be 300. If N_E is in the range 225-230 it is not possible to choose decisively between hypotheses A and B, while if $265 < N_E < 275$, B or C may be true.

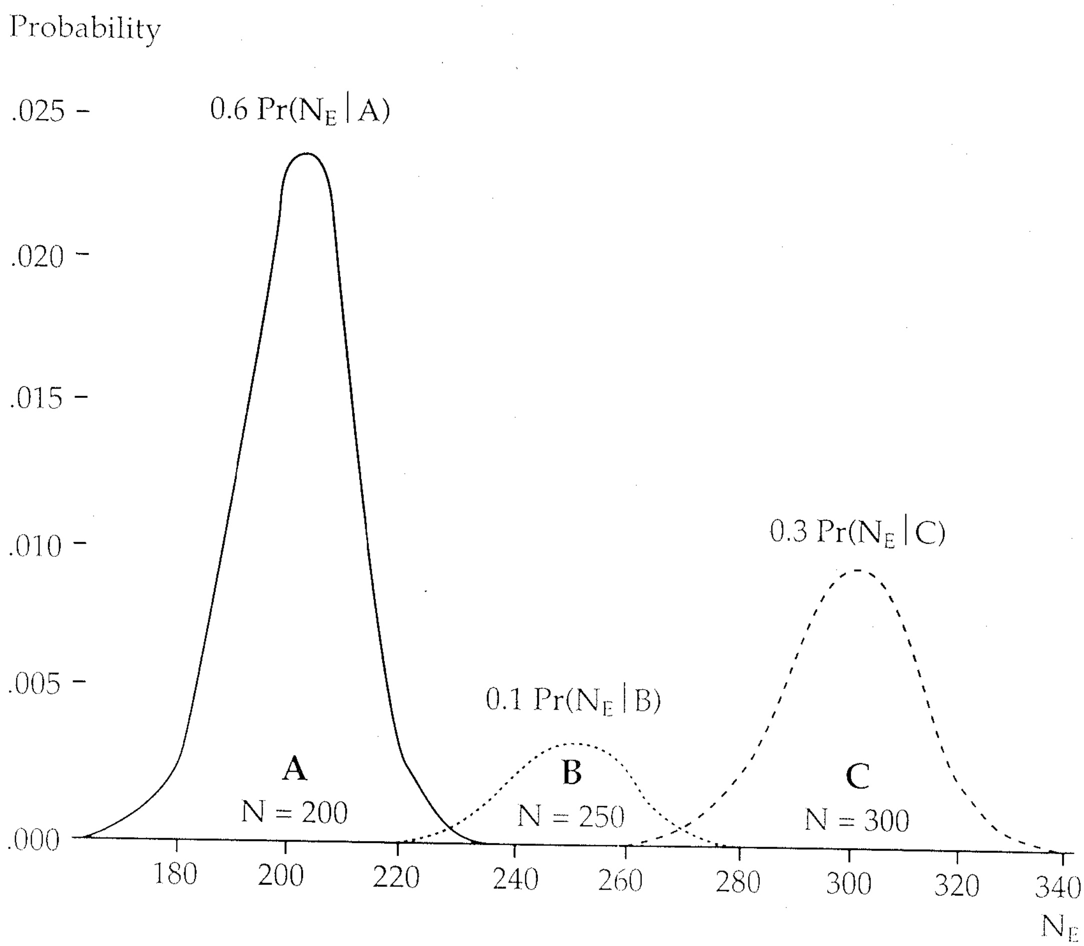
As has been demonstrated earlier, extending the duration of monitoring would narrow the bell curves, which would reduce the ranges of values of N_E for which discrimination between the hypotheses cannot be made with high confidence.

OTHER CATEGORIES OF MONITORING PROBLEMS

Some arms control agreements forbid deployment of any weapons of a specified type. In this case, a single reliable detection proves a violation. The confidence that can be placed in verification will turn on the technical ability of the monitoring system, so that the statistical problems discussed above do not arise. Environmental agreements may stipulate levels of pollution to be permitted, with monitoring seeking to establish whether agreed limits are being exceeded, and perhaps to identify the quantities released by the polluting sources. Or, perhaps more likely in the early stages of attempts to understand and control environmental damage, the objective of monitoring may simply be to estimate the quantity and pattern of pollution.

Figure III

Probability that Observations Will Make Estimate N_E if True Number N is 200, 250 or 300 (Hypothesis A, B and C)



Posterior Probabilities for Hypothesis A, B and C