



# **NIST Technical Note NIST TN 2231**

## **Stabilograms: Testbed Stability in a Pilot Experiment for Cellular Communication Anomaly Detection**

Michael Frey  
Mary Gregg  
Jacob D. Rezac  
Jason B. Coder  
Aziz Kord  
Anna Otterstetter  
Jeanne Quimby  
Alec Weiss

This publication is available free of charge from:  
<https://doi.org/10.6028/NIST.TN.2231>

**NIST Technical Note  
NIST TN 2231**

**Stabilograms: Testbed Stability in a  
Pilot Experiment for Cellular  
Communication Anomaly Detection**

Michael Frey

Mary Gregg

*Statistical Engineering Division  
Information Technology Laboratory*

Jacob D. Rezac

*Radio Frequency Division*

*Communications Technology Laboratory*

Jason B. Coder

Aziz Kord

Anna Otterstetter

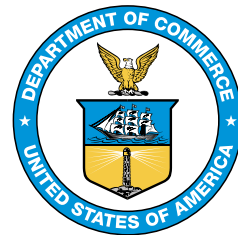
Jeanne Quimby

Alec Weiss

*Spectrum Technology and Research Division  
Communications Technology Laboratory*

This publication is available free of charge from:  
<https://doi.org/10.6028/NIST.TN.2231>

October 2022



U.S. Department of Commerce  
*Gina M. Raimondo, Secretary*

National Institute of Standards and Technology  
*Laurie E. Locascio, NIST Director and Under Secretary of Commerce for Standards and Technology*

Certain commercial entities, equipment, or materials may be identified in this document in order to describe an experimental procedure or concept adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the entities, materials, or equipment are necessarily the best available for the purpose.

#### **NIST Technical Series Policies**

[Copyright, Fair Use, and Licensing Statements](#)

[NIST Technical Series Publication Identifier Syntax](#)

#### **Publication History**

Approved by the NIST Editorial Review Board on 2022-07-26

#### **How to cite this NIST Technical Series Publication:**

Michael Frey, Mary Gregg, Jacob D. Rezac, Jason B. Coder, Aziz Kord, Anna Otterstetter, Jeanne Quimby, Alec Weiss (2022) Stabilograms: Testbed Stability in a Pilot Experiment for Cellular Communication Anomaly Detection. (National Institute of Standards and Technology, Gaithersburg, MD), NIST TN 2231. <https://doi.org/10.6028/NIST.TN.2231>

#### **NIST Author ORCID iDs**

Michael Frey: 0000-0003-2368-9106

Mary Gregg: 0000-0003-2991-6939

Jacob D. Rezac: 0000-0003-2992-1738

Jason B. Coder: 0000-0003-0408-4400

Jeanne Quimby: 0000-0001-8320-0056

Alec Weiss: 0000-0001-5642-8872

#### **Contact Information**

[jacob.rezac@nist.gov](mailto:jacob.rezac@nist.gov)

## **Abstract**

The National Institute of Standards and Technology's Communications Technology Laboratory conducted a cellular communication system pilot experiment in the summer of 2021. This pilot experiment was done to prepare for a broad-scope experiment to demonstrate anomaly detection from radiated emissions, possibly due to cyberattack. A goal of the pilot experiment was to check the experiment testbed for instabilities. A testbed is stable if measurements made over time with identical experiment settings exhibit no systematic variation. The data collected in the pilot experiment evidenced strong instabilities in the testbed's operation. These instabilities are seen in sequence plots of the experiment responses and, especially, in stabilograms devised expressly for this purpose in this work. Stabilograms are an intuitive and easy way to assess testbed instability. The instabilities reflected in the stabilograms are large enough to foreclose further analysis of the pilot data. The stabilograms from this study, though, suggest that if the testbed's instabilities can be effectively addressed, anomaly detection might be accomplished in the scenario under consideration.

## **Keywords**

Anomaly detection; ANOVA model; cellular communication; pilot experiment; testbed stability.

## Table of Contents

|                                                            |    |
|------------------------------------------------------------|----|
| 1. Introduction . . . . .                                  | 1  |
| 2. Stability Analysis . . . . .                            | 2  |
| 3. Stabilogram ANOVA Model . . . . .                       | 11 |
| 4. Summary and Final Remarks . . . . .                     | 14 |
| Appendix A: Derivations of Expected Mean-Squares . . . . . | 15 |

## 1. Introduction

The Communication Technology Laboratory (CTL) of the National Institute of Standards and Technology (NIST) conducted a cellular communication system pilot experiment in the summer of 2021. This pilot experiment, conducted with statistical support from the NIST Statistical Engineering Division, was done in preparation for a broad-scope experiment to demonstrate anomaly detection from radiated emissions, possibly due to cyberattack. Ciphering, ON or OFF, served as the anomaly to be detected in this experiment<sup>1</sup>. The pilot experiment had five purposes:

1. Gain facility with automated operation of the experiment testbed, and demonstrate an ability to collect data from the testbed in a stable manner.
2. Develop procedures for processing data collected from the testbed, to prepare the data for analysis.
3. Make determinations about which measurands to retain in the main, follow-on experiment.
4. Gain analysis experience with extraction of salient features for anomaly detection, and demonstrate the construction of anomaly detectors and associated performance (receiver operating characteristic, ROC) curves.
5. Use the pilot data to inform the design of the follow-on broad-scope anomaly detection experiment.

The pilot experiment was organized into  $N_S = 10$  sessions, each with  $N_B = 8$  blocks. The pilot experiment's sole factor, ciphering  $C$ , was the state of encryption between the wireless user equipment (UE) and the cellular base station (eNB). Ciphering  $C$  was operated at  $N_C = 2$  levels, ON and OFF. In each block, ciphering was ON for  $N_P = 9$  runs and then OFF for the same number  $N_P = 9$  of runs, amounting to  $N_C N_P N_B N_S = 1440$  runs for the experiment. Data were collected on the same set of 44 measurands in each run. Two runs in the collected data were identified as outliers, yielding, then,  $N = 1438$  runs for analysis. Many of the 44 measurands take the form of continuous distributions, which for analysis were represented by their centiles. Other measurands have discrete distributions which were represented by their proportions. Considering each centile/proportion as a response, and after excluding trivial responses<sup>2</sup>, each run in the pilot experiment yielded observations of 129 responses, representing in total the 44 measurands.

An experiment testbed is stable if measurements made over time with identical experiment (factor) settings exhibit no systematic, or patterned, variation. The statistical analysis of the pilot experiment response data found significant instabilities in the operation of the

---

<sup>1</sup>Ciphering refers here specifically to encryption of user plane traffic between the user equipment and the base station in a Long-Term Evolution (LTE) system.

<sup>2</sup>A run's location can be identified by its session  $S$ , its block  $B$  within the session, and its position  $P$  within the block. A trivial response is a response that changed very little or not at all as a function of  $C$ ,  $P$ ,  $B$ , or  $S$ .

testbed. These instabilities are detailed in this report through the use of sequence plots and stabilograms. The stabilogram is a novel graphical presentation of instability developed here expressly for this analysis. Given the testbed's instabilities, no further analysis of the data was attempted, either to build anomaly detectors or to proceed directly on to design the full-scope experiment. Rather, subsequent small, targeted experiments are planned to identify, understand, and minimize the source of the testbed's instabilities. The stabilograms used in this study suggest that, after these instabilities are removed, the research goal of detecting ciphering state might be achievable for the scenario represented by the pilot experiment's design.

The remainder of this report is organized as follows. Section 2 details the responses in the pilot experiment. For a representative set of five of these responses, this section also illustrates the testbed instabilities that were found. Section 3 presents the underlying four-way ANOVA (analysis of variance) model for the stabilograms presented in this report. Section 4 closes with a summary and some related final remarks.

## 2. Stability Analysis

Table 1 lists the 44 measurands in the pilot experiment, along with their column numbers and column labels in the associated data file [3]. Each experiment response is a centile or proportion derived from the sample distributions of these measurands in each experiment run. Each measurand in Table 1 is accompanied by a description of the measurand's type, whether discrete, continuous, binary, or categorical and a brief indication of its physical meaning. Table 2 spells out the acronyms used in Table 1.

Figures 1–5 show sequence plots and stabilograms for the subset of measurands  $X_1$ ,  $X_{13}$ ,  $X_{19}$ ,  $X_{41}$ , and  $X_{43}$ . While not exhaustive, this collection of measurands is sufficient 1) to support the conclusion that the pilot data evidence significant instabilities in the testbed's operation and 2) showcase the use and interpretation of stabilograms. This investigation of testbed instability found no significant dependence of instability on ciphering  $C$ . Therefore, the sequence plots and stabilograms in this report show results for ciphering ON and OFF combined. The upper left panel in each array of sequence plots shows the measurand as a function of run number, while the remaining three panels show the measurand as a function of session  $S$ , block  $B$ , and position  $P$ . The black bars plotted in the session, block, and position panels of these figures denote averages. For the discrete measurand  $X_1$  and  $X_9$  and the categorical measurand  $X_{19}$ , the plotted values are distribution proportions. The value whose proportion is displayed is denoted at the top of these figures. The values plotted for the continuous measurands  $X_{41}$  and  $X_{43}$  are the medians of their distributions.

Stabilograms are included in Figs. 1–5 for the representative subset of measurands  $X_3$ ,  $X_{13}$ ,  $X_{19}$ ,  $X_{41}$ , and  $X_{43}$  examined in this report. The stabilogram is a diagrammatic device developed expressly for this investigation of testbed instability. Each square in a response stabilogram corresponds to a component source of relative variability in the response; the

| Measurand  | Column # | Column label | Type | Description                        |
|------------|----------|--------------|------|------------------------------------|
| $X_0$      | 7-8      | G-H          | d    | Mean bits per symbol               |
| $X_1$      | 10-13    | J-M          | d    | Mean RB allocation                 |
| $X_2$      | 15-26    | O-Z          | q    | Signaled MCS value                 |
| $X_3$      | 28-31    | AB-AE        | d    | Retry count                        |
| $X_4$      | 33-36    | AG-AJ        | d    | Real RSRP                          |
| $X_5$      | 38-39    | AL-AM        | d    | Real RSRQ                          |
| $X_6$      | 41-43    | AO-AQ        | c    | Aggregate PCC DL-SCH throughput    |
| $X_7$      | 46-48    | AT-AV        | c    | Mean PCC DL-SCH throughput         |
| $X_8$      | 51-53    | AY-BA        | c    | Mean UEs signaled on PCC/TTI       |
| $X_9$      | 56-57    | BD-BE        | d    | Minimum DL RB usage                |
| $X_{10}$   | 59-61    | BG-BI        | c    | Transport block rate               |
| $X_{11}$   | 64-66    | BL-BN        | c    | Mean HARQ TB size                  |
| $X_{12}$   | 69-70    | BQ-BR        | d    | Mean bits per symbol               |
| $X_{13}$   | 72-74    | BT-BV        | d    | Average retransmission count       |
| $X_{14}$   | 76-78    | BX-BZ        | c    | Number of active TTIs              |
| $X_{15}$   | 81-85    | CC-CG        | c    | DL-SCH throughput                  |
| $X_{16}$   | 88-91    | CJ-CM        | d    | Number of uses of 16QAM            |
| $X_{17}$   | 93-95    | CO-CQ        | c    | Retransmission count               |
| $X_{18}$   | 98       | CT           | c    | Transport block size               |
| $X_{19}$   | 101-108  | CW-DD        | q    | HARQ ID                            |
| $X_{20}$   | 110-113  | DF-DI        | d    | Retry count                        |
| $X_{21}$   | 115-116  | DK-DL        | b    | CRC for the DL decode              |
| $X_{22}^*$ | 118-120  | DN-DP        | -    | Radio network temporary identifier |
| $X_{23}$   | 123-125  | DS-DU        | c    | UL-SCH throughput                  |
| $X_{24}$   | 128-132  | DX-EB        | c    | DL-SCH throughput                  |
| $X_{25}$   | 135-143  | EE-EM        | d    | Average UL-SCH throughput per UE   |
| $X_{26}$   | 145      | EO           | c    | Average DL-SCH throughput per UE   |
| $X_{27}$   | 148-149  | ER-ES        | d    | Number of MAC control blocks       |
| $X_{28}$   | 151-152  | EU-EV        | d    | Padding size                       |
| $X_{29}$   | 154-157  | EX-FA        | q    | Signaled MCS value                 |
| $X_{30}$   | 159-161  | FC-CE        | c    | Number of PDUs                     |
| $X_{31}$   | 164-174  | FH-FR        | c    | Mean PDU size                      |
| $X_{32}$   | 177-178  | FU-FV        | d    | Padding size                       |
| $X_{33}$   | 180-186  | FX-GD        | d    | Aggregate SDU throughput           |
| $X_{34}$   | 188-194  | GF-GL        | d    | Aggregate SDU throughput           |
| $X_{35}$   | 196-198  | GN-GP        | c    | PDU receive rate                   |
| $X_{36}$   | 201-216  | GS-HH        | d    | PDU count                          |
| $X_{37}$   | 218-220  | HJ-HL        | d    | PDU ACK count                      |
| $X_{38}$   | 222-223  | HN-HM        | d    | Status PDU count                   |
| $X_{39}$   | 225-227  | HQ-HS        | c    | DL-SCH throughput                  |
| $X_{40}$   | 230      | HV           | c    | PDCCP layer DL throughput          |
| $X_{41}$   | 233-235  | HY-IA        | c    | Average DL UEs per TTI             |
| $X_{42}$   | 238-239  | ID-IE        | d    | Average SNR                        |
| $X_{43}$   | 241-251  | IG-IQ        | c    | VSA power                          |

**Table 1.** List of 44 measurands in the pilot experiment with their column numbers and labels in the associated data file [3]. Measurand types: b, binary; c, continuous; d, discrete; q, categorical. Acronyms and abbreviations used in this table and elsewhere can be found in Table 2. \*There was an error in the data processing of  $X_{22}$ , and it should be excluded from any analysis.

|       |                                   |
|-------|-----------------------------------|
| ANOVA | analysis of variance              |
| CRC   | cyclic redundancy check           |
| DL    | downlink                          |
| HARQ  | hybrid automatic repeat request   |
| ID    | identification                    |
| MAC   | medium access control             |
| MCS   | modulation coding state           |
| MS    | mean square                       |
| RB    | resource block                    |
| PCC   | primary component carrier         |
| PDU   | protocol data unit                |
| QAM   | quadrature amplitude modulation   |
| ROC   | receiver operating characteristic |
| RSRP  | reference signal received power   |
| RSRQ  | reference signal received quality |
| SCH   | scheduled                         |
| SDU   | service data unit                 |
| TTI   | transmit time interval            |
| UE    | user equipment                    |
| UL    | uplink                            |

**Table 2.** Abbreviations and acronyms.

area and side length of each square represent, respectively, mean-square (MS) variation and root-mean-square (RMS) variation. The squares in each stabilogram are scaled so that the variation associated with random measurement error  $E$  (black square) is represented by a unit square with sides of unit length. Then, the stabilogram squares for each of the other sources of variation represent relative MS and RMS variation.

The stabilograms in Figures 1–5 are color-coded: the variation due to change in ciphering  $C$  in blue, random error  $E$  in black, and the systematic errors associated with  $P$ ,  $B$ , and  $S$  in red for testbed instabilities reflected by position, block, and session. A testbed instability can be considered to be present if the red  $P$ ,  $B$ , and  $S$  squares are about the same size or larger than the black unit square for  $E$ . These instabilities (red squares) are separately of practical concern for detection if they are also large relative to the ciphering state (blue square) that is sought to be determined. These stabilograms can be understood as depicting strength and power signal-to-noise ratios for  $C$  relative to the noise  $E$ , and interference-to-noise ratios for  $P$ ,  $B$ , and  $S$ . From this signal detection perspective, the interfering sources  $P$ ,  $B$ , and  $S$ , are indeed extant if their strength and power is on the order of or greater than those of the ambient noise  $E$ ; they become an issue for signal (anomaly) detection if they dominate the signal  $C$ .

The relative MS variations shown in a stabilogram derive from a standard ANOVA and assume a four-way additive effects ANOVA model of response with fixed effects  $C$ ,  $P$ ,  $B$ , and  $S$ . This statistical model is presented in detail in the following section of this report. The remainder of this section interprets the sequence plots and stabilograms in Figs. 1–5.

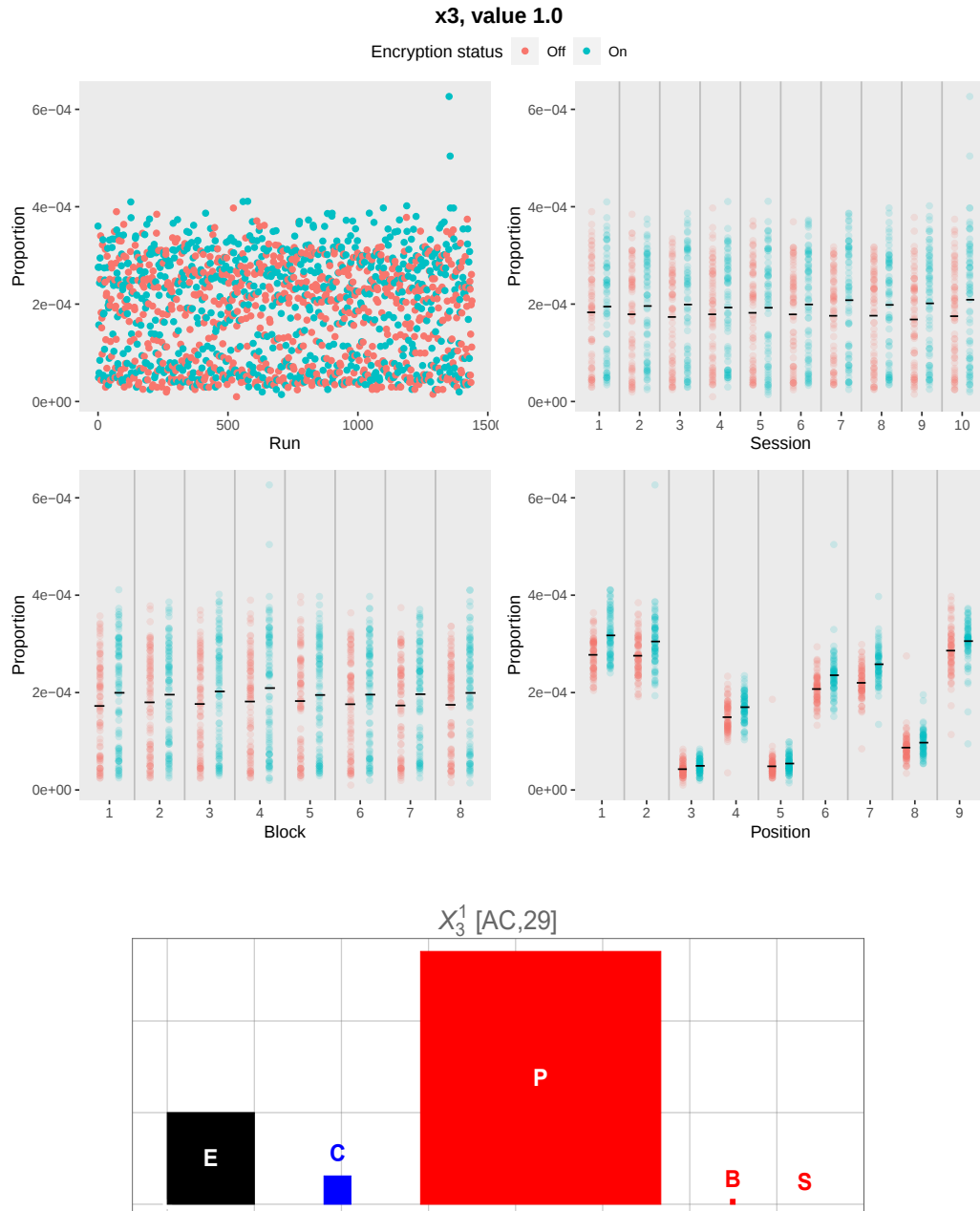
**Measurand  $X_3$ :** Figure 1 shows the proportion  $X_3^{1.0}$  of times the measurand  $X_3$  took the value 1.0 for each run. The session and block sequence plots show no significant patterned variation in  $X_3^{1.0}$  over the course of the experiment. The position sequence plot, though, shows that  $X_3^{1.0}$  changed systematically as a function of  $P$ . The stabilogram summarizes the same effects; the red square for  $P$  is large relative to the black unit square for random error  $E$ , while the red squares for  $B$  and  $S$  are small relative to the black square. The sequence plots show that the response  $X_3^{1.0}$  tends to be slightly greater when ciphering is *ON*. The stabilogram describes the size of this effect by the blue square’s size relative to that of the black square.

**Measurand  $X_{13}$ :** Figure 2 shows for each run the proportion  $X_{13}^1$  of times the measurand  $X_{13}$  took the value 0.01. A large interference effect due to position  $P$  can be seen in the position sequence plot. In particular, the second measurement within each block tended to exhibit a much higher proportion for the value 0.01 for measurand  $X_{13}$  compared to other measurement positions. The stabilogram confirms the out-sized position effect. As with the response  $X_3^{1.0}$  in Figure 1, the sequence plots and the stabilogram for response  $X_{13}^{0.01}$  agree on the presence of a small, but visible effect due to ciphering.

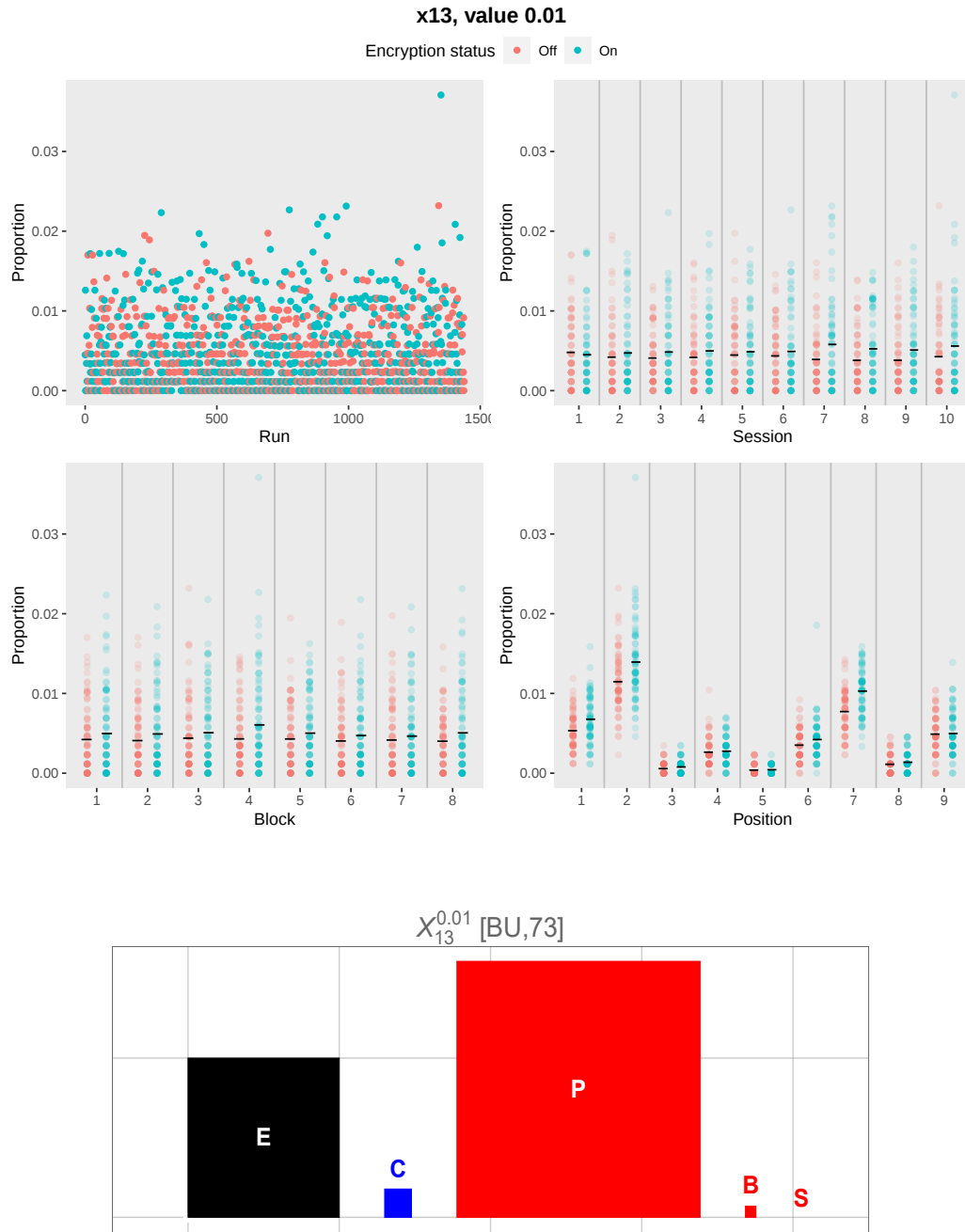
**Measurand  $X_{19}$ :** Figure 3 shows the proportion  $X_{19}^0$  of times measurand  $X_{19}$  took the value 0 for each run. The stabilogram for  $X_{19}^0$  show no instability (vanishingly small red squares); however, it also shows that  $X_{19}^0$  is useless for anomaly detection (small blue square). In the session plot in this figure, the mean proportion is fairly constant over  $S$ , but some heteroscedasticity seems to be present; less random variation seems present during session 6 in the session plot compared to other sessions. The run plot shows this, also.

**Measurand  $X_{41}$ :** Figure 4 shows the median  $X_{41}^{50\text{th}}$  of measurand  $X_{41}$ . This response is another illustration of extreme position instability. The median of  $X_{41}$  varies systematically across measurement position; the first measurement within a block exhibited consistently lower values than other positions, while the highest values were seen at measurement positions 5–7. This position effect is strongly reflected in the stabilogram.

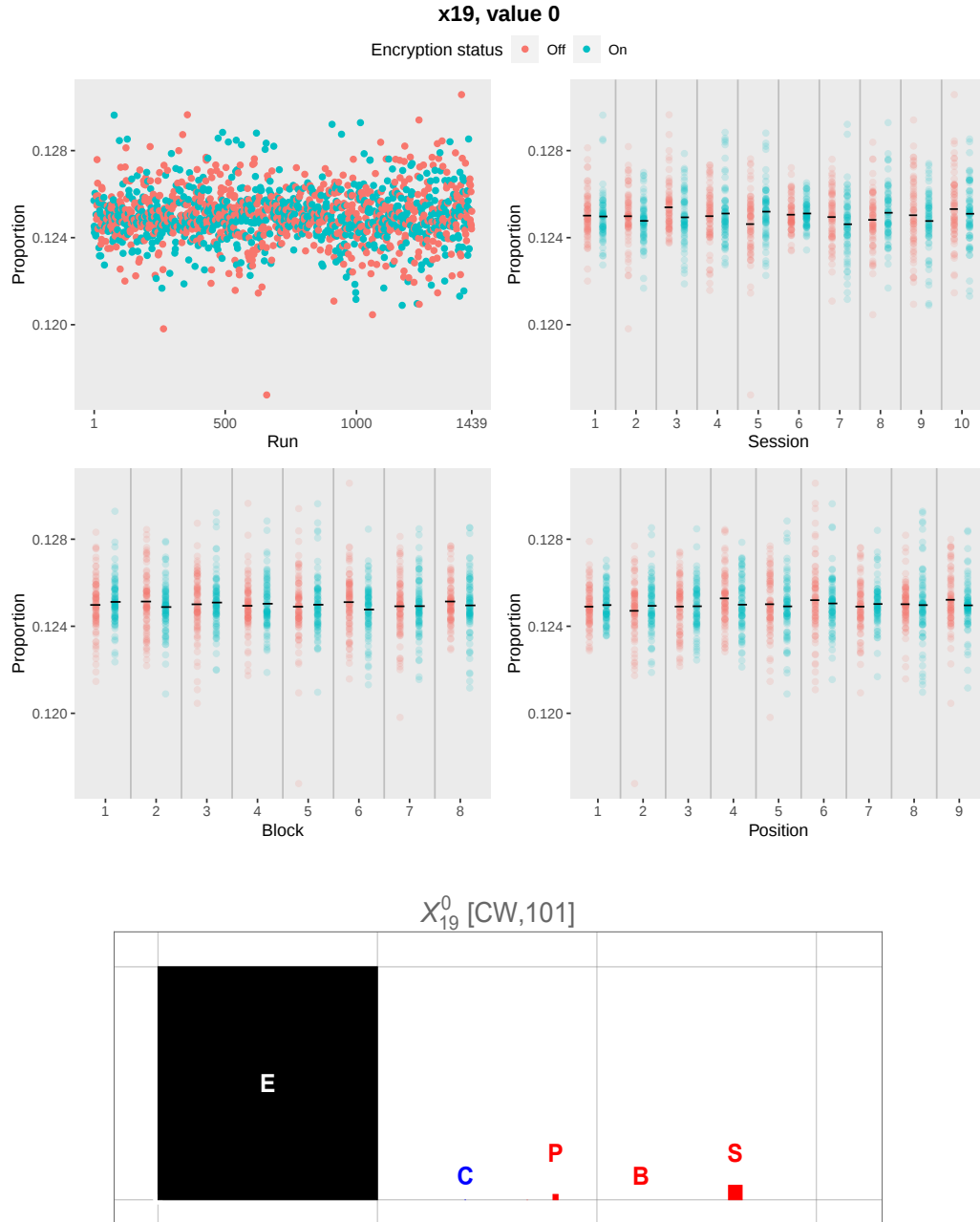
**Measurand  $X_{43}$ :** Figure 5 shows the median  $X_{43}^{50\text{th}}$  of measurand  $X_{43}$ . The run plot shows the median following a general rise over the time of the pilot experiment; this shows up, too, in the session plot. Testbed instability was also expressed through position  $P$ , with measurements occurring at later positions tending to exhibit larger values. The stabilogram shows that these interference effects,  $P$  and  $S$ , obscure what might otherwise be, in coordination with other responses, a small, but potentially detectable ciphering effect  $C$  (blue square).



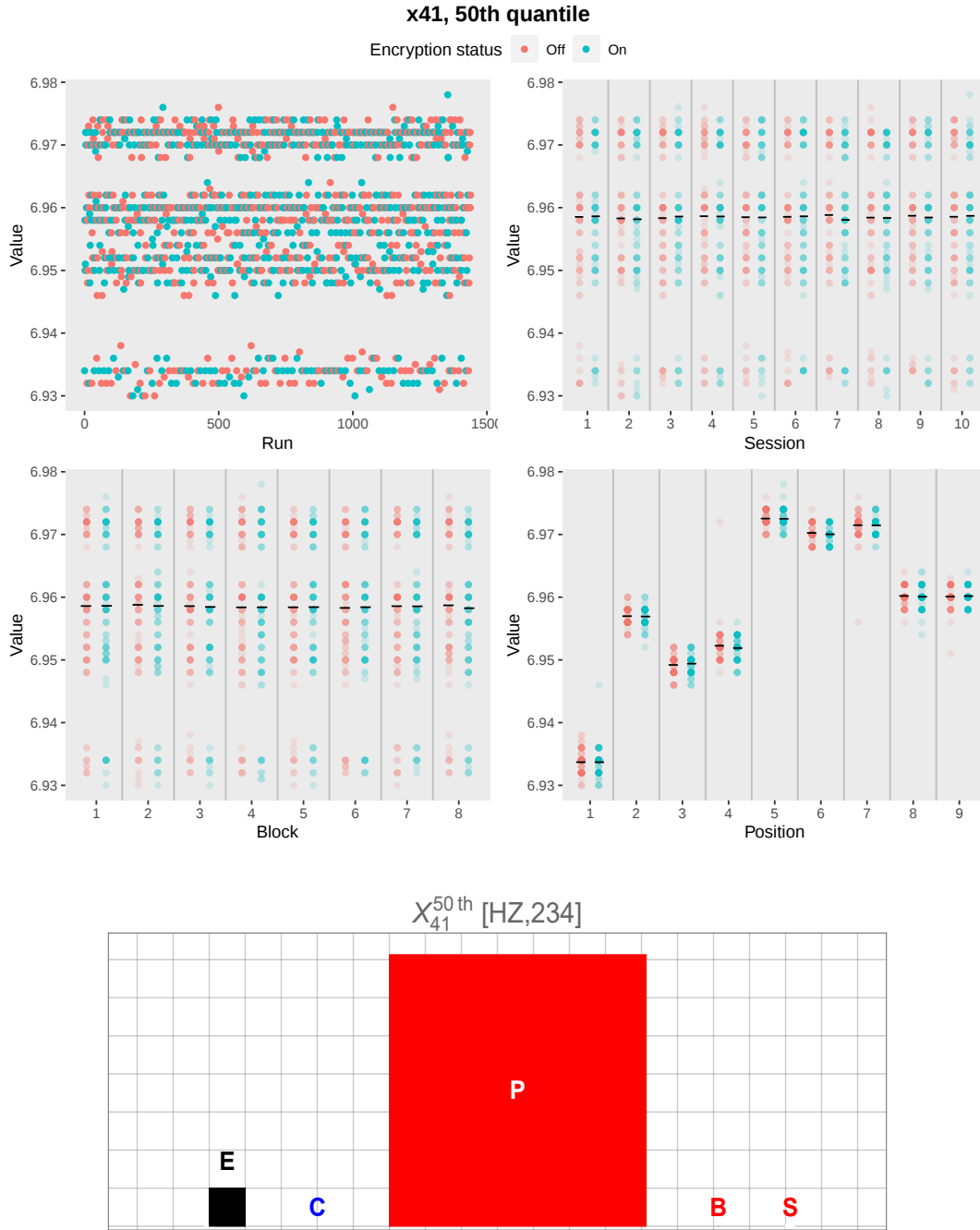
**Fig. 1.** Sequence plots (top) and stablogram (bottom) for the response  $X_3^1$ . The testbed appears in the  $X_3^1$  position plot to change state in a systematic fashion as a function of block position  $P$ —the signature of testbed instability. This is confirmed by the large position effect (red square) in the stablogram. The sequence plots also show a small ciphering effect; the mean responses of  $X_3^1$  with respect to  $P$ ,  $B$ , and  $S$  are consistently higher with  $C = \text{ON}$ . This accords with the small, but present, ciphering effect  $C$  (blue square) in the stablogram. The effect  $S$  is so small relative to the noise  $E$  that the square associated with  $S$  in stablogram is not visible.



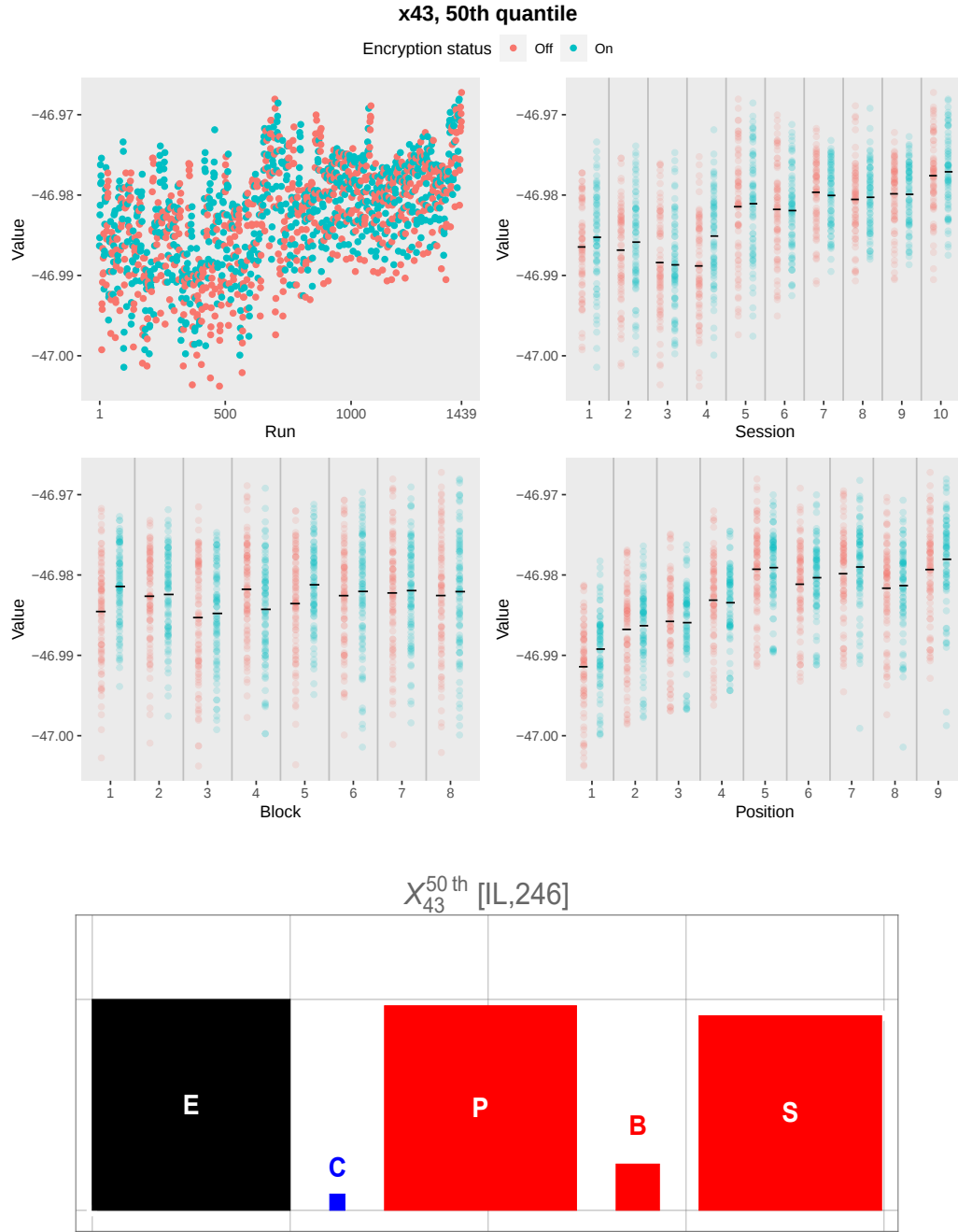
**Fig. 2.** Sequence plots (top) and stabilogram (bottom) for the response  $X_{13}^{0.01}$ . With respect to  $X_{13}^{0.01}$ , the testbed appears to consistently depart from its baseline operation at position 2 in each block of each session. The ciphering effect  $C$  (blue square) in  $X_{13}^{0.01}$  is small relative to both the random variation (black square) and the testbed's interfering position effect  $P$ . The session effect  $S$  is so small relative to the noise  $E$  that the square associated with  $S$  in stabilogram is not visible.



**Fig. 3.** Sequence plots (top) and stablogram (bottom) for the response  $X_{19}^0$ . The interference effects  $P$ ,  $B$ , and  $S$  due to testbed instability are vanishingly small compared to the random variation (black square) present in  $X_{19}^0$ . The run and session plots suggest an unusual testbed instability in the form of heteroscedasticity—the variability in the response seems somewhat smaller in session 6 than in other sessions. The ciphering and block effects  $C$  and  $B$  are so small relative to the noise  $E$  that the squares associated with them in stablogram are not visible.



**Fig. 4.** Sequence plots (top) and stabilogram (bottom) for the response  $X_{41}^{50th}$ . The interference effect  $P$  due to testbed instability dominates all other effects in  $X_{41}^{50th}$ . The ciphering, block, and session effects  $C$ ,  $B$ , and  $S$  on  $X_{41}^{50th}$  are so small relative to the noise  $E$  that the squares associated with them in the stabilogram are not visible.



**Fig. 5.** Sequence plots (top) and stabilogram (bottom) for the median  $X_{43}^{50th}$  of measurand  $X_{43}$ . The run and session plots show an unstable testbed drifting over time. Testbed instability is also clearly reflected in position within block. Some instability may also exist across block, though if it does exist, it is small and seems to diminish in the later blocks of each session. The stabilogram is in accord with these observations, with relatively large red squares for  $P$  and  $S$  and a small square for  $B$ .

### 3. Stabilogram ANOVA Model

The stabilograms presented in this report are based on the additive, fixed-effects, four-way ANOVA model [1, 2]

$$X_{cpbs} = \mu + \alpha_c + \beta_p + \gamma_b + \delta_s + \varepsilon_{cpbs}, \quad (1)$$

where  $X_{cpbs}$  is an experiment response measured with the ciphering set to  $C$  at sequence position  $P$  in block  $B$  of session  $S$ , with  $c = \text{ON, OFF}$ ,  $p = 1, \dots, N_P$ ,  $b = 1, \dots, N_B$ , and  $s = 1, \dots, N_S$ , where  $N_P = 9$ ,  $N_B = 8$ , and  $N_S = 10$ . The ANOVA model represents the overall response average by  $\mu$  and allows for additive deviations  $\alpha_c$ ,  $\beta_p$ ,  $\gamma_b$ , and  $\delta_s$  from  $\mu$  that depend on the ciphering setting  $C$  and the position  $P$ , block  $B$ , and session  $S$  in which the response  $X_{cpbs}$  was observed. The deviations  $\alpha_c$ ,  $\beta_p$ ,  $\gamma_b$ , and  $\delta_s$  satisfy the condition

$$\sum_{c=\text{ON}}^{\text{OFF}} \alpha_c = \sum_{p=1}^{N_P} \beta_p = \sum_{b=1}^{N_B} \gamma_b = \sum_{s=1}^{N_S} \delta_s = 0, \quad (2)$$

so that all the parameters in (1) are estimable. The model assumes that, beyond the additive fixed effects due to  $C$ ,  $P$ ,  $B$ , and  $S$ , all variation remaining in  $X_{cpbs}$  is random, represented by the random error  $\varepsilon_{cpbs}$ . These random errors  $\varepsilon_{cpbs}$  are assumed to be uncorrelated with zero mean and common variance  $\sigma^2$ . (Beyond these, no assumptions about the model errors  $\varepsilon_{cpbs}$  are required for the development of the stabilogram that follows.) ANOVA model (1) has only a single replicate per treatment (that is, per given  $C$ ,  $P$ ,  $B$ ,  $S$  setting combination) and does not allow for possible interactions among  $C$ ,  $P$ ,  $B$ , and  $S$ . Excluding interactions is appropriate for the present exploratory analysis; the model's limited purpose is to assess the individual effects of  $\beta_p$ ,  $\gamma_b$ , and  $\delta_s$  relative to  $\alpha_c$  and  $\sigma$ . While mathematically some instability is present if and only if some of the  $\beta_p$ ,  $\gamma_b$ , and  $\delta_s$  are non-zero; more practically, for purposes of testbed evaluation, instability should be considered to possibly be present if any of the position, block, or session effects are larger than that of the random variation.

ANOVA response model (1) has 31 parameters,  $\mu$  and  $\sigma$  plus the sets of fixed effects parameters  $\alpha_c$ ,  $\beta_p$ ,  $\gamma_b$ , and  $\delta_s$ . These parameters are all readily estimated from the measured responses  $X_{cpbs}$ . Unbiased estimators<sup>3</sup> for these parameters are<sup>4</sup>

$$\begin{aligned} \hat{\mu} &= \bar{X}_{\bullet\bullet\bullet\bullet}, & \hat{\sigma}^2 &= \frac{1}{D_E} \text{SS}_E, \\ \hat{\alpha}_c &= \bar{X}_{c\bullet\bullet\bullet} - \bar{X}_{\bullet\bullet\bullet\bullet}, & \hat{\beta}_p &= \bar{X}_{\bullet p \bullet\bullet} - \bar{X}_{\bullet\bullet\bullet\bullet}, \\ \hat{\gamma}_b &= \bar{X}_{\bullet\bullet b \bullet} - \bar{X}_{\bullet\bullet\bullet\bullet}, & \hat{\delta}_s &= \bar{X}_{\bullet\bullet\bullet s} - \bar{X}_{\bullet\bullet\bullet\bullet}, \end{aligned} \quad (3)$$

<sup>3</sup>An estimator is any function of the data  $X_{cpbs}$ . An estimator  $\hat{\kappa}$  of a parameter  $\kappa$  is unbiased if on average it neither over- nor underestimates the parameter; that is  $E[\hat{\kappa}] = \kappa$ .

<sup>4</sup>Standard "dot" notation for sums and averages is used here. For example,  $X_{\bullet\bullet\bullet\bullet}$  is the responses  $X_{cpbs}$  summed over all four of their indices  $c, p, b, s$ . Similarly,  $\bar{X}_{\bullet\bullet\bullet\bullet}$  is the responses  $X_{cpbs}$  averaged over all four indices;  $\bar{X}_{c\bullet\bullet\bullet}$  is the responses  $X_{cpbs}$ , for each of value of  $c = \text{ON, OFF}$ , averaged over its other indices  $p, b, s$ , etc.

where  $SS_E$  is the ANOVA model's error sum-of-squares

$$SS_E = \sum_{c,p,b,s} e_{cpbs}^2 \quad (4)$$

with residuals

$$e_{cpbs} = X_{cpbs} - (\mu + \hat{\alpha}_c + \hat{\beta}_p + \hat{\gamma}_b + \hat{\delta}_s). \quad (5)$$

The error degrees-of-freedom in (3) is

$$D_E = N - 1 - (N_C - 1) - (N_P - 1) - (N_B - 1) - (N_S - 1), \quad (6)$$

where  $N = N_C N_P N_B N_S$ . The other sums-of-squares associated with the fixed effects in model (1) are

$$\begin{aligned} SS_C &= \sum_{c,p,b,s} \hat{\alpha}_c^2, & SS_P &= \sum_{c,p,b,s} \hat{\beta}_p^2, \\ SS_B &= \sum_{c,p,b,s} \hat{\gamma}_b^2, & SS_S &= \sum_{c,p,b,s} \hat{\delta}_s^2. \end{aligned} \quad (7)$$

The total sum-of-squares

$$SS_T = \sum_{c,p,b,s} (X_{cpbs} - \hat{\mu})^2 \quad (8)$$

is a measure of all the (squared) variation in the  $X_{cpbs}$ . The algebraic identity

$$SS_T = SS_C + SS_P + SS_B + SS_S + SS_E, \quad (9)$$

called an ANOVA decomposition, partitions the total variation  $SS_T$  into parts attributable to each of  $C$ ,  $P$ ,  $B$ ,  $S$ , and  $E$ . Associated with the ANOVA sums-of-squares in (4) and (7) are the ANOVA mean-squares presented with their expectations in Table 3. These mean-squares, suitably adjusted, yield the unbiased estimators of the squared effect sizes given in Table 4.

For the adjusted mean-squares in Table 4, two observations can be made. First, because the adjusted mean-squares are unbiased estimators of the squares of the corresponding effects in the ANOVA model, they are meaningful measures of the variability due to those effects. Second, because of the ANOVA decomposition (9) from which the mean-squares in Table 4 derive, they account for *all* the variability present in the response  $X_{cpbs}$ .

The adjusted mean-squares in Table 4 can be compared to assess the relative effect sizes of  $P$ ,  $B$ ,  $S$ ,  $E$ , and  $C$ . A stablogram shows the relative effect sizes in Table 4 as red squares for the systematic instability effects, a black square for the relative random measurement effect, and a blue unit square for the ciphering effect that is the target for detection. The areas of the stablogram squares are the estimated relative mean-square effect sizes, while the lengths of the squares' edges are the relative root-mean-square (RMS) effect sizes. The areas in a stablogram are determined as in the fourth column of Table 4.

| ANOVA<br>sources of<br>variation | Mean-squares                  | Expected<br>mean-squares                              | ANOVA<br>$F$ statistics   |
|----------------------------------|-------------------------------|-------------------------------------------------------|---------------------------|
| Ciphering $C$                    | $MS_C = \frac{SS_C}{N_C - 1}$ | $\frac{N}{N_C(N_C - 1)} \sum_c \alpha_c^2 + \sigma^2$ | $F_C = \frac{MS_C}{MS_E}$ |
| Position $P$                     | $MS_P = \frac{SS_P}{N_P - 1}$ | $\frac{N}{N_P(N_P - 1)} \sum_p \beta_p^2 + \sigma^2$  | $F_P = \frac{MS_P}{MS_E}$ |
| Block $B$                        | $MS_B = \frac{SS_B}{N_B - 1}$ | $\frac{N}{N_B(N_B - 1)} \sum_b \gamma_b^2 + \sigma^2$ | $F_B = \frac{MS_B}{MS_E}$ |
| Session $S$                      | $MS_S = \frac{SS_S}{N_S - 1}$ | $\frac{N}{N_S(N_S - 1)} \sum_s \alpha_s^2 + \sigma^2$ | $F_S = \frac{MS_S}{MS_E}$ |
| Error $E$                        | $MS_E = \frac{SS_E}{D_E}$     | $\sigma^2$                                            |                           |

**Table 3.** Mean-squares and their expectations for ANOVA model (1). The expected mean-squares lead algebraically to the unbiased estimators presented in Table 4.

| Effect        | Average<br>squared<br>effect size | Unbiased<br>estimator             | Estimated relative<br>average squared<br>effect size |
|---------------|-----------------------------------|-----------------------------------|------------------------------------------------------|
| Ciphering $C$ | $\frac{1}{N_C} \sum_c \alpha_c^2$ | $\frac{N_C - 1}{N} (MS_C - MS_E)$ | $\frac{N_C - 1}{N} r(F_C - 1)$                       |
| Position $P$  | $\frac{1}{N_P} \sum_p \beta_p^2$  | $\frac{N_P - 1}{N} (MS_P - MS_E)$ | $\frac{N_P - 1}{N} r(F_P - 1)$                       |
| Block $B$     | $\frac{1}{N_B} \sum_b \gamma_b^2$ | $\frac{N_B - 1}{N} (MS_B - MS_E)$ | $\frac{N_B - 1}{N} r(F_B - 1)$                       |
| Session $S$   | $\frac{1}{N_S} \sum_s \delta_s^2$ | $\frac{N_S - 1}{N} (MS_S - MS_E)$ | $\frac{N_S - 1}{N} r(F_S - 1)$                       |
| Error $E$     | $\sigma^2$                        | $MS_E$                            | 1                                                    |

**Table 4.** Average squared effect sizes for model (1), their unbiased estimators, and their estimated average squared effect sizes relative to the measurement error  $E$ . These last quantities provide the areas for the squares in the stablologram. The ramp function  $r(x) = \max(x, 0)$  in the rightmost column addresses cases where the effect is not significant and the  $F$ -statistic is less than unity.

#### 4. Summary and Final Remarks

A goal of the pilot study was to practice testbed operation and to detect instabilities that might be present in that operation. Standard sequence plots were made to examine the testbed data for instabilities. Also, a new exploratory graphical device called a stabilogram was devised to aid this investigation. Some representative stabilograms are presented in this report. These stabilograms show 1) that significant instabilities were present and 2) that, if these instabilities were removed, the research goal of detecting ciphering state in the scenario represented by the pilot experiment's design might be accomplished.

The pilot experiment was structured into sessions and blocks, consistent with the way the testbed equipment was operated (i.e., with power cycling and ciphering state changes). This structure allowed the testbed stabilograms to be constructed with three meaningful instability components, the stabilograms' red *P*, *B*, and *S* squares. This structure allowed the stabilograms to both detect the presence of instabilities and suggest their potential origin(s). In particular, instability reflected in position *P* within each block is of first concern.

Testbed stability should ideally be monitored or subsequently assessed as part of the execution and analysis of any experiment. The stabilogram is an easy exploratory visual check for potential instability in an experiment testbed's operation. The squares in a stabilogram represent the different sources of variability in the response under consideration: the squares' areas represent variance (signal, noise, or interference power) and the squares' edges are standard deviations (signal, noise, or interference strength). The stabilogram is scaled so that the black square has unit size, with then the other squares representing strengths and powers relative to those of the (black) measurement error. The stabilogram's interpretation is quick and intuitive:

- The red squares represent testbed instabilities, while the black square represents random variation. Compare the sizes of the red squares to that of the black square. If any of these red squares are of the same order or larger than the black square, instability is said to be present for purposes of testbed investigation. The greater the size of a red square relative to the black square, the greater is that testbed instability.
- The stabilogram's blue unit square represents the experiment's factor — e.g., the ciphering *C* in the pilot experiment. Compare the sizes of the red squares (representing instabilities) to that of the blue square. If a red square is large relative to the blue factor effect, this renders the factor difficult to detect with the response under consideration. The larger a red square is relative to the blue factor square, the more difficult the detection problem is using the response in question.
- In the absence of any large red instability square, the problem of detecting the factor represented by the blue square comes down to the blue square's size relative to that of the black random variation (noise) square. In this case the smaller the black (noise) square is relative to the blue (factor) square, the easier detection will be with the response in question.

- A pilot experiment with an instrumented testbed often captures many experiment responses in parallel. Provided these response stabilograms evidence little testbed instability, the stabilograms can be compared to visually rank the different responses for their relative potential as features for an anomaly detector.

The stabilogram is applicable to experiments of the present type with one two-level factor. By simple modification of the ANOVA model in (1), stabilograms can be prepared equally well for experiments with more than one factor and/or to factors with more than two levels and, in particular, to factor screening experiments with many factors. For a stabilogram to be applicable, though, replicate runs are needed, a requirement that can usually be satisfied in the setting of a pilot experiment.

## Appendix A: Derivations of Expected Mean-Squares

This appendix presents derivations for the expected mean-squares  $E[\text{MS}_c]$  and  $E[\text{MS}_E]$  in Table 3. These quantities are the basis for the Table 4 entries that are used to construct stabilograms. The derivations presented here, while original, are typical of those associated with linear statistical models. Derivations for models similar to (1) can be found in [1, 2].

Derivation of  $E[\text{MS}_C]$ : The ANOVA sum-of-squares  $\text{SS}_C$  in (8) can be written

$$\begin{aligned}\text{SS}_C &= \sum_{c,p,b,s} \hat{\alpha}_c^2 \\ &= N_P N_B N_S \sum_c (\bar{X}_{c\dots} - \bar{X}_{\dots})^2 \\ &= \frac{N_C}{N} \left( \sum_c X_{c\dots}^2 - \frac{X_{\dots}^2}{N_C} \right),\end{aligned}\tag{10}$$

using  $N = N_C N_P N_B N_S$ . Then the expectation of  $\text{SS}_C$  in (10) is

$$\begin{aligned}E[\text{SS}_C] &= \frac{N_C}{N} \sum_c E[X_{c\dots}^2] - \frac{1}{N} E[X_{\dots}^2] \\ &= \frac{N_C}{N} \sum_c E \left[ \sum_{p_1, p_2} \sum_{b_1, b_2} \sum_{s_1, s_2} X_{cp_1 b_1 s_1} X_{cp_2 b_2 s_2} \right] - \frac{1}{N} E[X_{\dots}^2] \\ &= \frac{N_C}{N} \sum_c \sum_{p_1, p_2} \sum_{b_1, b_2} \sum_{s_1, s_2} E[X_{cp_1 b_1 s_1} X_{cp_2 b_2 s_2}] - \frac{1}{N} E[X_{\dots}^2].\end{aligned}\tag{11}$$

The first expectation in (11) is

$$\begin{aligned}E[X_{cp_1 b_1 s_1} X_{cp_2 b_2 s_2}] &= E[(\mu + \alpha_c + \beta_{p_1} + \gamma_{b_1} + \delta_{s_1} + \varepsilon_{cp_1 b_1 s_1})(\mu + \alpha_c + \beta_{p_2} + \gamma_{b_2} + \delta_{s_2} + \varepsilon_{cp_2 b_2 s_2})] \\ &= (\mu + \alpha_c + \beta_{p_1} + \gamma_{b_1} + \delta_{s_1})(\mu + \alpha_c + \beta_{p_2} + \gamma_{b_2} + \delta_{s_2}) + \delta_{p_1 p_2} \delta_{b_1 b_2} \delta_{s_1 s_2} \sigma^2,\end{aligned}$$

where, for example,  $\delta_{s_1 s_2}$  is the Kronecker delta function with  $\delta_{s_1 s_2} = 1$  if  $s_1 = s_2$  and  $\delta_{s_1 s_2} = 0$  otherwise. Then, using the constraints in (2), the multiple sum in (11) is

$$\begin{aligned}
\sum_c \sum_{p_1, p_2} \sum_{b_1, b_2} \sum_{s_1, s_2} E[X_{cp_1 b_1 s_1} X_{cp_2 b_2 s_2}] &= \sum_c \sum_{p_1 p_2} \sum_{b_1 b_2} \sum_{s_1 s_2} (\mu + \alpha_c)^2 \\
&\quad + N_P N_B N_S \sum_{p_1 b_1 s_1} (\beta_{p_1} + \gamma_{b_1} + \delta_{s_1}) \sum_c (\mu + \alpha_c) \\
&\quad + N_P N_B N_S \sum_c (\mu + \alpha_c) \sum_{p_2 b_2 s_2} (\beta_{p_2} + \gamma_{b_2} + \delta_{s_2}) \\
&\quad + N_C \sum_{p_1 b_1 s_1} (\beta_{p_1} + \gamma_{b_1} + \delta_{s_1}) \sum_{p_2 b_2 s_2} (\beta_{p_2} + \gamma_{b_2} + \delta_{s_2}) \\
&= \frac{N^2}{N_C^2} \sum_c (\mu + \alpha_c)^2 + N \sigma^2 \\
&= \frac{N^2}{N_C^2} \left( N_C \mu^2 + \sum_c \alpha_c^2 \right) + N \sigma^2. \tag{12}
\end{aligned}$$

The second expectation in (11) is

$$\begin{aligned}
E[X_{\bullet\bullet\bullet\bullet}^2] &= V[X_{\bullet\bullet\bullet\bullet}] + E^2[X_{\bullet\bullet\bullet\bullet}] \\
&= \sum_{c, p, b, s} V[X_{cpbs}] + \left( \sum_{c, p, b, s} E[X_{cpbs}] \right)^2 \\
&= \sum_{c, p, b, s} V[\epsilon_{cpbs}] + \left( \sum_{c, p, b, s} (\mu + \alpha_c + \beta_p + \gamma_b + \delta_s) \right)^2 \\
&= N \sigma^2 + N^2 \mu^2, \tag{13}
\end{aligned}$$

using  $N = N_C N_P N_B N_S$  to arrive at (13). Returning results (12) and (13) to (11) yields

$$\begin{aligned}
E[SS_C] &= \frac{(N^2/N_C^2) (N_C \mu^2 + \sum_c \alpha_c^2) + N \sigma^2}{N/N_C} - \frac{N \sigma^2 + N^2 \mu^2}{N} \\
&= \frac{N}{N_C} \sum_c \alpha_c^2 + (N_C - 1) \sigma^2. \tag{14}
\end{aligned}$$

We therefore conclude that

$$E[MS_C] = \frac{N}{N_C(N_C - 1)} \sum_c \alpha_c^2 + \sigma^2,$$

as given in Table 3. The expectations in Table 3 for the ANOVA mean-squares  $MS_P$ ,  $MS_B$ , and  $MS_S$  have analogous derivations.

Derivation of  $E[\text{MS}_E]$ : The ANOVA total sum-of-squares  $\text{SS}_T$  in (8) can be written

$$\begin{aligned}\text{SS}_T &= \sum_{c,p,b,s} (X_{cpbs} - \bar{X}_{\bullet\bullet\bullet\bullet})^2 \\ &= \sum_{c,p,b,s} X_{cpbs}^2 - \frac{X_{\bullet\bullet\bullet\bullet}^2}{N}.\end{aligned}\quad (15)$$

Using  $E[\varepsilon_{cpbs}] = 0$  and the constraints in (2), the expectation of the first term in (15) becomes

$$\begin{aligned}\sum_{c,p,b,s} E[X_{cpbs}^2] &= \sum_{c,p,b,s} E[(\mu + \alpha_c + \beta_p + \gamma_b + \delta_s + \varepsilon_{cpbs})^2] \\ &= N\mu^2 + \sum_{c,p,b,s} (\alpha_c + \beta_p + \gamma_b + \delta_s)^2 + N\sigma^2 \\ &= N\mu^2 + \frac{N}{N_C} \sum_c \alpha_c^2 + \frac{N}{N_P} \sum_p \beta_p^2 + \frac{N}{N_B} \sum_b \gamma_b^2 + \frac{N}{N_S} \sum_s \delta_s^2 + N\sigma^2.\end{aligned}\quad (16)$$

The expectation of the squared total  $X_{\bullet\bullet\bullet\bullet}^2$  in (15) is derived in (13). Returning the results from (13) and (16) to the total sum-of-squares  $\text{SS}_T$  in (15) yields the expected value

$$\begin{aligned}E[\text{SS}_T] &= \left( N\mu^2 + \frac{N}{N_C} \sum_c \alpha_c^2 + \frac{N}{N_P} \sum_p \beta_p^2 + \frac{N}{N_B} \sum_b \gamma_b^2 + \frac{N}{N_S} \sum_s \delta_s^2 + N\sigma^2 \right) - \frac{N\sigma^2 + N^2\mu^2}{N} \\ &= \frac{N}{N_C} \sum_c \alpha_c^2 + \frac{N}{N_P} \sum_p \beta_p^2 + \frac{N}{N_B} \sum_b \gamma_b^2 + \frac{N}{N_S} \sum_s \delta_s^2 + (N-1)\sigma^2.\end{aligned}\quad (17)$$

The expected error sum-of-squares  $E[\text{SS}_E]$  can be obtained from the ANOVA decomposition in (9), using the expected sums-of-squares in (14) and (17); we have

$$\begin{aligned}E[\text{SS}_E] &= E[\text{SS}_T] - E[\text{SS}_C] - E[\text{SS}_P] - E[\text{SS}_B] - E[\text{SS}_S] \\ &= \frac{N}{N_C} \sum_c \alpha_c^2 + \frac{N}{N_P} \sum_p \beta_p^2 + \frac{N}{N_B} \sum_b \gamma_b^2 + \frac{N}{N_S} \sum_s \delta_s^2 + (N-1)\sigma^2 \\ &\quad - \left( \frac{N}{N_C} \sum_c \alpha_c^2 + (N_C-1)\sigma^2 \right) - \left( \frac{N}{N_P} \sum_p \beta_p^2 + (N_P-1)\sigma^2 \right) \\ &\quad - \left( \frac{N}{N_B} \sum_b \gamma_b^2 + (N_B-1)\sigma^2 \right) - \left( \frac{N}{N_S} \sum_s \delta_s^2 + (N_S-1)\sigma^2 \right) \\ &= D_E \sigma^2,\end{aligned}\quad (18)$$

where  $D_E$  is the error degrees-of-freedom in (6). The expected error mean-square is therefore

$$E[\text{MS}_E] = \frac{E[\text{SS}_E]}{D_E} = \sigma^2.\quad (19)$$

This completes the derivations of the expected mean-squares in Table 3, justifying the quantities in Table 4 used to construct stabilograms.

## References

- [1] Neter, J., Kutner, M. H., Nachtsheim, C. J., & Wasserman, W. (2004). *Applied linear statistical models. 5th ed.* McGraw-Hill/Irwin.
- [2] Montgomery, D. C. (2017). *Design and analysis of experiments.* John Wiley & sons.
- [3] Frey, M., Gregg, M., Rezac, J., Coder, J., Kord, A., Midzor, M., Quimby, J., Weiss, A. (2022), Data Publication: Stabilograms for a Cellular Communication Anomaly Detection Experiment , National Institute of Standards and Technology, <https://doi.org/10.18434/mds2-2679>