

**LOKAD**

**TESTING AND BENCHMARKING  
OUT-OF-SHELF MONITORING SYSTEMS**



[www.lokad.com](http://www.lokad.com) • [contact@lokad.com](mailto:contact@lokad.com)

## Contents

---

Introduction.....	2
Test/Benchmark objectives .....	2
Performance criteria .....	3
Methodologies for making OOS systems comparable.....	4
Features .....	5
Quantification of system profitability .....	6
Project phases.....	10
Responsibilities .....	11
Timeline.....	11
Summary .....	12
Authors .....	13
Lokad.....	13

# Testing and Benchmarking Out-of-Shelf Monitoring Systems

## Introduction

---

On-shelf availability has been a concern in grocery retail for many years. Many retail organizations today seek greater visibility into out-of-shelf situations within their network and are considering evaluating monitoring technology that detects out-of-shelf situations and sends alerts to store management.

The testing and benchmarking of competing OOS monitoring systems is an important step in quantifying the value for the organization of such systems, and in identifying the most suitable vendor. This document outlines key considerations for a trial project and provides suggestions on how to set up a benchmark for competing OOS monitoring solutions.

## Test/Benchmark objectives

---

The objectives of a test should include the following criteria:

1. System performance
  - a. Precision of alerts
  - b. Coverage
  - c. Latency
  - d. Response time
2. Operational integration
  - a. Process/workflow integration
  - b. Impact on store staff
3. Scalability of the solution
4. System agility
  - a. Speed of setup/integration
  - b. Ease of calibration
5. Features
6. Vendor partnership

## Performance criteria

---

An OOS monitoring system's performance is defined via several interrelated criteria which are outlined below.

**Sensitivity:** System sensitivity denotes the percentage of in-store OOS situations that are captured.

**Precision:** Precision is the ratio of true OOS situations with respect to all OOS events reported by the software. This is also often called accuracy.

**Coverage:** The share of the product portfolio for which the system will perform within the required specifications. The ability of an indirect measurement system to detect an OOS situation is directly correlated with the sales frequency (or rotation) of the product. An indirect system will have a minimum rotation threshold below which the required sensitivity and precision cannot be provided.

**Latency:** The latency of an indirect measurement system describes how much time elapses from the start of an OOS situation until it is detected with the desired precision. Latency is also directly linked to product rotation. For a given time period, detection latency decreases with the number of units sold during this time period.

**Response-time,** or 'technical turn-around' time describes how quickly alerts are produced by the system once a data update has been received. Limiting factors are typically the degree of automation and the scalability of the system from an IT standpoint.



## Methodologies for making OOS systems comparable

---

How to compare competing OOS systems is not immediately obvious. The problem is best demonstrated with an example: System A delivers 50 alerts/day with 95% precision and System B produces 500 alerts/day with 85% precision. Which system has better performance? System A has higher precision, while system B has higher sensitivity. The core of the problem is that there is a trade-off between precision and sensitivity. If system B had only delivered the 50 most precise alerts per day (provided that its features include quantification of alert precision) – what would have been its precision? It is impossible to compare the performance of the two systems on a 1:1 basis. We propose two methods for solving this problem:

### Method 1: Scoring rule

The scoring rule provides a simple and efficient way to make two systems comparable, and it is a very 'expressive' method, in that it captures both sensitivity and precision. Both systems are scored according to a simple rule:

- Each true alert counts as +1.
- Each false alert counts as -3.

The system with the highest number of points has the best performance.

This approach **completely solves the sensibility/precision trade-off**. The number -3 is chosen so that a system with a precision of 75% scores *zero*, however this value can be adjusted by choosing a different value for false alerts (e.g. -2 or -4). We believe 75% to be a fair threshold, with simple systems achieving roughly a precision of 70% for products sold at a rate of more than 6 units / day. Obviously, sophisticated systems should significantly outperform this level and achieve a strongly positive score.

### Methods 2: Freezing the number of alerts

While less 'expressive', and therefore less desirable in a test scenario, we describe the freezing approach here, as in some cases it is more appropriate (for example in cases where only a set number of alerts can be processed per day, e.g. for operational reasons).

The interrelation issue between precision and sensitivity can be solved by 'freezing' one criterion to allow objective assessment and benchmarking of the second. For practical reasons we propose to freeze sensitivity by defining the number of alerts to be issued per

time period, and to assess the precision of the alerts by verifying in store the correctness of each alert. The number of alerts should be limited to a level that is manageable from a store perspective, and sensible from a sensitivity point of view.

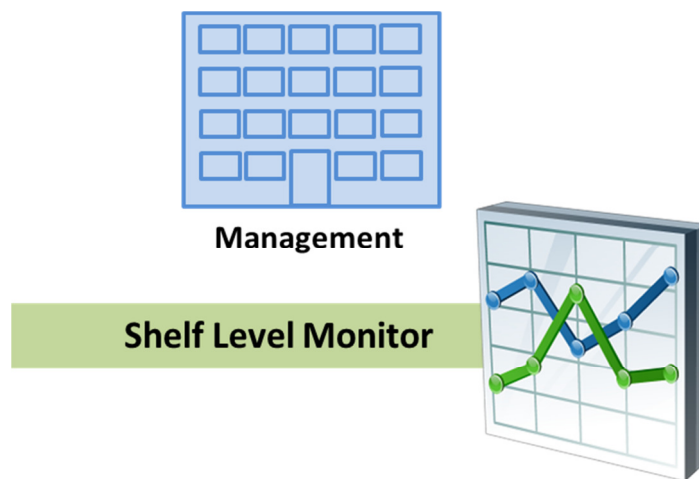
The disadvantage of this method is **reduced ‘expressiveness’** as it does not provide the ability to capture potential strength in terms of system sensitivity.

## Features

In addition to performance quantification, systems should be assessed for the features they provide. The following list is non-exhaustive and provides a starting point for a feature check-list.

Feature	Comment
Level of automation	Crucial for scalability and low TCO
Rule based vs. self-learning system	Self-learning reduce the complexity for the operating retailer (no defining rules)
Alert prioritization	Avoid ‘flooding’ of store staff
Quantification of alert ‘profitability’	Measure and manage system profitability
Network wide benchmarking	Use as a management information system
System cost	
Implementation time	
Development agility	How long do upgrades take, and how short are feedback loops on system changes?

**Table 1:** Feature catalogue for OOS monitoring technology





## Quantification of system profitability

---

Shelf availability is considered to have an important mid- to long-term impact on customer loyalty<sup>1</sup>, and investments in shelf availability can therefore be considered to be strategic. However, out-of-shelf monitoring systems should be, and often are, profitable on a day-to-day basis.

Profitability and amortization time (if investments are necessary) of an out of shelf system should therefore be quantified as part of a system trial.

The profitability analysis of an OOS system breaks down into three main components:

1. The gain associated with the resolution of an OOS situation
2. The cost of resolving an OOS situation (primarily labor)
3. The cost of implementing and running an OOS monitoring system

### Quantifying profitability from OOS resolution

In practice, the vast majority of OOS situations will be resolved after a certain amount of time through regular re-stocking of shelves. What therefore needs to be quantified is not the gain due to the resolution of the OOS situation, but rather the *gain due to the shortening of an OOS situation as the result of an OOS alert*.

### Quantifying the financial gain from an OOS resolution

As part of the analysis, an assumption needs to be made as to how long it will take to resolve an OOS situation, once detected. A typical assumption, for example, is that any OOS situation that is detected on day  $x$  will be resolved the following day  $x + 1$ . This

---

1

- *Improving On-Shelf Availability for Retail Supply Chains Requires the Balance of Process and Technology*, Gartner, 26 May 2011
- *Optimal Shelf Availability, Increasing shopper satisfaction at the moment of truth*, Roland Berger Consultants, 2003
- *Retail Out of Stocks - A Worldwide Examination of Extent, Causes, and Consumer Responses*, Thomas W. Gruen (University of Colorado), Daniel S. Corsten (University of St. Gallen), Sundar Bahradwaj (Emory University), 19 May 2002

assumption is invalid in some cases such as an out-of-stock situation at the supplier's or in the warehouse. In this case, detection of the OOS situation will not result in faster resolution. On the other hand, in some organizations a resolution might be faster once the OOS situation is detected. In short, the assumption should be adjusted to the specific business, with x+1 being the most common situation.

The variables in the analysis of the financial effect of shorter latency in resolving OOS situations are therefore:

- V:** The expected volume of sales during the OS situation
- U:** The unit price of the SKU
- M:** The gross margin of the SKU
- C:** The internal substitution rate
- D:** The original duration of an OOS situation (in the absence of monitoring)
- L:** The detection latency of the OOS system

We define **the gain resulting from resolution** of an OOS situation as follows:

$$Gain_{Resolution} = V \times U \times M \times (1 - C) \times (D - L)$$

Concretely, the formula quantifies the additional gross margin obtained by shortening the OOS situation.

**Average sales volume V:** this is essentially an estimation of the sales lost during an OOS situation. In practice, since a benchmark will entail a large portfolio of products with different volumes of sales, we suggest using the average daily sales volume of the SKU for that specific retail location.

**Unit price U:** we recommend using actual retail prices or, if not available, an average unit price.

**Gross margin M:** again, it is preferable to use the actual gross margin; otherwise an average value can be applied.

**Internal substitution rate C:** the available literature on OOS situations suggests that only part of the OOS situation will lead to an actual loss in sales due to the fact that in some cases consumers will switch to substitutes (e.g. a different brand) or delay their purchases. As a rule of thumb, a loss of sales is often expected in 30% of OOS situations, while in 70% of cases an internal substitute is purchased.



**The average duration D** of an OOS situation is the average time it typically takes for an OOS situation to be resolved via the store's normal replenishment operations.

**The detection latency L** is the average time it takes to detect an OOS situation. For a monitoring system to make sense, this latency must clearly be shorter than the average duration of OOS situations. The sooner an OOS situation is detected, the bigger the financial gain in terms of gross margin 'recovered' by the system.

### Quantifying the financial cost of an OOS resolution

The cost of resolving an OOS situation is largely staff time, which includes manual verification and resolution. The calculation must also take into account the time spent verifying false positives (alerts that are not correct).

This cost can be modeled

**S:** Hourly labor cost of store staff

**F:** Hourly frequency of checks, i.e. the number of references that can be checked in 1 hour

**P:** Precision, i.e. the percentage of alerts that are correct

The cost of the resolution of an OOS situation can be calculated as:

$$Cost_{Resolution} = \frac{S}{P \times F}$$

**The labor cost S** should be fully loaded

**The checking frequency F** is the number of checks that can be performed within one hour and should be quantified based on the individual business. A typical value, for example, is 60 checks per hour, provided that a listing of alerts by department is available so that the employee does not have to crisscross the store to validate alerts.

**The precision P** is the percentage of true alerts, which needs to be factored into the total cost in staff-hours. For example, a 50% precision (which is very low) would double the total cost to resolve an OOS situation, as the employee would need to check two alerts in order to identify and resolve one OOS situation.

### Summary of the economic equation

In combining the two equations given above, we can establish the global economic equation for the system. To do this, however, we need to introduce one more variable:

**A:** The number of alerts per day for the store.

The final equation for the quantification of the *daily* profit of the system is therefore:

$$Profit = (Gain_{Resolution} - Cost_{Resolution}) \times P \times A$$

It is worth noting that the compromise between precision and sensitivity is reflected in this equation:

- If the precision P increases, the first two fractions increase in value, while the third fraction (number of alerts), decreases
- If the number of alerts A, increases, the inverse happens and the first two fractions decrease.

### **Total system profitability and amortization**

By subtracting the operating cost of the system (e.g. license fees, IT costs, etc.) the overall monthly profitability can be calculated, which will determine the amortization time for the entirety of setup costs and investments.

## Project phases

---

We recommend separating system performance testing from system scalability testing. To establish system performance, it is sufficient to work with a small number of representative stores; scalability is then ideally demonstrated in a subsequent step. This approach aims at minimizing resource investment and risks for the retailer.

We suggest a three-phase process for fully assessing the solution, with each phase followed by a review/debrief prior to moving on to the next phase.

1. **Initialization:** system testing and tuning prior to going live in store
2. **Store alerts:** assessment of the analytic performance in a small number of representative stores
3. **Scaling:** assessment of scalability

### Initialization

Prior to starting with in-store alerts, vendors should perform an initial analysis of the data, verify the assumptions (e.g. sales velocities of products in scope), test and compute the expected performance of the system and verify data integration. In this phase, vendors should make use of system capabilities such as self-calibration, self-diagnosis and historical analysis of the store's OOS history.

Furthermore, the data feed for daily updates to sales data should already be in place, as it will allow testing and verification of faultless data transfer.

### Store alerts phase

During the store alerts phase, vendors will produce alerts for the pilot store(s) within the agreed time-periods (e.g. daily or intra-day).. On 'going live' with the test, it is helpful to establish a direct feedback loop. The goal here is to have a qualitative feedback on the alerts and to make sure the alerts are received, can be read (correct codes etc. are used) and are generally working within the organizational constraints of the store.

The precision of the alerts will need to be assessed by the store staff. Weekly feedback on performance is ideal and a simple reporting format should be agreed to.

### Scaling phase

In the scaling phase, the scalability of the solution can be tested by running a trial in a large number of stores respectively at full scale.

Furthermore, tighter operational integration can be tested, for example, by excluding warehouse and “voluntary” out-of-stocks.

### Responsibilities

---

**Vendors** should provide an end-to-end OOS-analysis during a trial and seek to adhere to the agreed timeline.

The **retailer** needs to:

- Provide the required data as defined.
- Ensure communication between the vendor and the store during the pilot phase.
- Report precision on a weekly basis according to the agreed format.
- Provide a project manager to be the main point of contact for the vendor
- Provide a project time-line
- Provide short-cycle feedback

The staff-hours required in a retail grocery environment, from a technical perspective, relate to:

- Data extraction and preparation
- Set-up of data stream
- Organization of in-store alert assessment

### Timeline

---

Timing example for a standard test of an OOS system:

1. Data access and ‘data feed’ open – start of trial
2. Initialization: 2 – 4 weeks
3. Store alerts: 4 – 6 weeks
4. Preparation of scaling: 2 – 4 weeks
5. Full scale trial: 6 -12 weeks

Total timeline: 14 – 26 weeks from data access

## Summary

---

Benchmarking OOS systems is fairly straightforward, provided that the test is carefully planned. The good news is that key performance characteristics – precision and sensitivity – can be quantified easily, which leads to objective benchmark results and a fair comparison of system performance.

## Authors

---



### Joannès Vermorel

An internationally recognized expert on cloud computing and statistical learning, Joannès Vermorel holds a Master of Science degree from the 'École Normale Supérieure de Paris' (ENS Ulm). In 2010, Joannès won the worldwide Microsoft Windows Azure Partner of the Year Award in recognition for his pioneering work in the cloud. Joannes also teaches a software engineering course at the École Normale Supérieure in Paris and is the founder of Lokad.



### Matthias Steinberg

CEO of Lokad, Matthias Steinberg was previously Vice President at Summit Partners LTD, a leading global private equity and venture capital firm. His prior experience also includes Airbus Industries and The Boston Consulting Group. He holds a Master of Engineering from RWTH Aachen and an MBA from INSEAD.

## Lokad

---

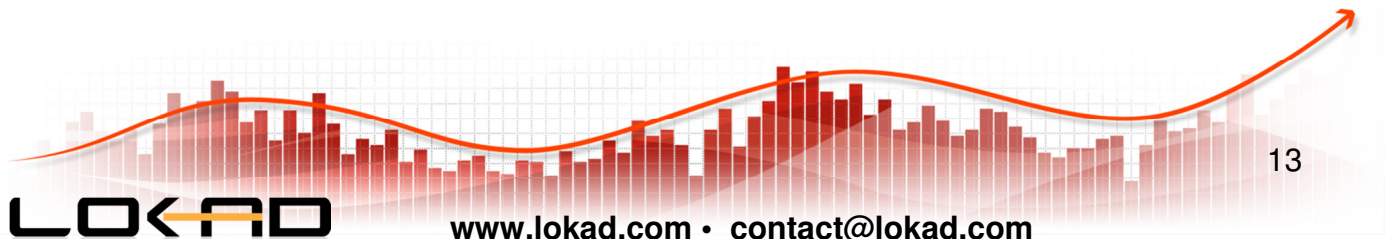
Lokad is a technology company focusing on big data analytics software for retail networks, wholesale and eCommerce. Client solutions include inventory optimization, loyalty card data analysis and out-of-shelf monitoring. The company is the winner of the 2010 Microsoft Worldwide Partner of the Year Award and is recognized as an international leader in cloud computing technology.



### 2010 PARTNER OF THE YEAR

Windows Azure Platform

Winner





## Lokad Shelfcheck Out-of-Shelf Monitoring

---

Shelfcheck On shelf availability monitoring provides prioritized out-of-shelf alerts to store staff and group-wide transparency and benchmarking to management and suppliers.



The solution is fully cloud based and delivered in a software-as-a-service model, which reduces investments and operating cost to a minimum. A full scale field test can be completed in weeks.

For further information please see [www.lokad.com](http://www.lokad.com).