Experience: Redesigning Radio Test Specifications via Field Performance Correlations

Narayana Sastry* , Balvinder Singh† , Sujoy Roychowdhury‡, Sourav Mazumdar§
Ericsson
Email: *venkata.dharma.surya.narayana.sastry.rachakonda@ericsson.com, †balvinder.a.singh@ericsson.com,
‡sujoy.roychowdhury@ericsson.com, §sourav.mazumdar@ericsson.com

***Abstract*—Products manufactured in Telecommunications companies need to go through a large number of test cases, often running into thousands. These test cases have test specifications and test limits decided by designers from their understanding of the components and experience. Test limits used are usually binary thresholds so that if a manufactured unit is on the correct side of the test limit then the corresponding test is said to be passed. The impact of variation in test performance on field performance has not been extensively studied. To study this, we consider the field performance of a large number of Remote Radio Units (RRU) and try to correlate the field performance with the test performance. We describe here a system which can integrate data from production tests and field performance and run large number of statistical experiments automatically. These statistical experiments allow us to narrow down on a very small number of production tests which are clearly correlated to system performance. The process makes it feasible to bring down the number of production test cases which need further analysis and identify possible changes in thresholds. In particular, we look at downlink performance degradation and linearization faults as one set of field performance results and channel quality indicators as another. Initial results on our system show orders of magnitude difference in the number of test cases directly impacting field performance results to the actual number of tests performed. We believe that incorporation of such systems and checks would lead to better telecommunications equipment performance across the industry and therefore better end user experience.**

I. Background

Telecommunications manufacturing is a complicated process because very high precision equipment needs to be built and tested. In particular, Remote Radio Units (RRU) which are the focus of our study, are manufactured globally by various companies. Every manufactured Radio unit contains a large number of components often sourced from different vendors. Components could themselves be manufactured units like power amplifiers or filter units and they can also be individual components like capacitors which then go into a printed circuit board (PCB). Every unit is then subjected to a very detailed testing regimen often going into thousands of tests. These tests include basic health checks, tests which indicate the unit's stated performance as well as tests required from regulatory purposes. Examples of health check tests include testing of idle current and voltages, testing of temperatures of equipment at various load levels etc. Calibration of different components like filter and Voltage Standing Wave Radio (VSWR) as well as testing of gain accuracy and noise figure go into the set of performance tests mentioned earlier. In addition, some performance tests maybe required for regulatory requirements like those of Electromagnetic field (EMF) and Out of Band Unwanted Emission (OBUE). Most of these tests are done at the manufacturing factory itself. Almost all test cases measure numerical values. Most manufacturing tests (except some calibration tests) have one or more limits set by the design community. The limits are almost exclusively decided based on the stated performance of the unit in consideration and regulatory requirements, if applicable. Test cases typically have an order in which they must be executed. Tests which have one limit are typically ones where the measured value needs to be either larger than or smaller than the stated test limits, whereas for tests which have two limits the measured value needs to be between the two limits. A test case is said to be passed if the measured value is in line with the defined limits as explained above. Calibration test cases often do not have a pass vs. fail criterion. Instructions for the order, limits and re-run are available in test design documents. A unit must pass all test cases to be considered for further verification and shipping to customers which are in most cases telecom operators around the world.

§All authors contributed equally in this work

Once shipped to the telecom operators these equipment are deployed on field and monitored at a regular basis based on different performance parameters. Radio units need to maintain some Service Level Agreements (SLA) in the field. Many of these SLAs are mandated by standards bodies like Third Generation Partnership Project (3GPP), European Telecomunications Standards Institute (ETSI) etc. Others may be enforced by national regulatory bodies like Federal Communications Commission (FCC) , Telecom Regulatory Authority of India (TRAI) etc. Other performance parameters may be chosen by the telecom operator to give the level of promised customer service to the end user.

While most equipment perform as per SLA, few of them might display varying/degrading performance over their lifetime. Such variation in performance can be attributed to range of factors, including operational network performance parameters and external factors like environment etc. Telecom operators continuously collect and record network performance data. Original equipment manufacturer (OEMs) can leverage this information and incorporate it as a feedback to enhance their existing manufacturing process esp. redesigning or fine tuning the existing test limits. In many cases, OEMs lack access to

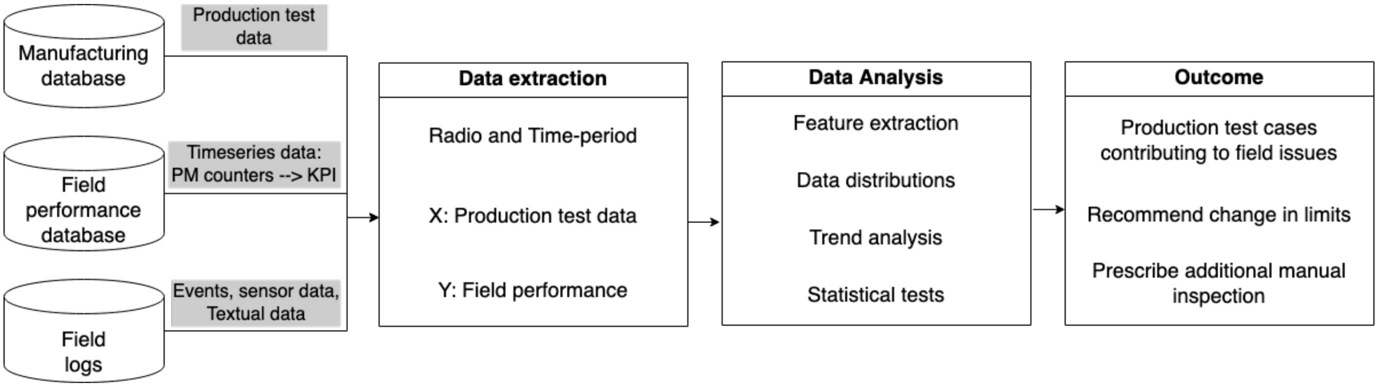


Fig. 1: Architecture diagram showing overall flow of the project.

the equipment operational data , unless these are returned by operators to OEMs for trouble shooting, repair or maintenance purpose. The concept of correlating field performance with product design and testing improvements represents a mutually beneficial scenario for both operators and OEMs.

In this study, we consider a case where we have access to both manufacturing test results and field performance metrics. Using this relatively unique opportunity to look at data across the entire life cycle of radio units, we describe here a system to integrate end to end information across them. In particular, we describe a system which can run systematic statistical experiments to understand the correspondences between field performance and manufacturing test results. Our objective is to build a scientifically robust system of identifying if certain test limits or certain test results need more careful inspection. This may involve identification of rules for specialized re-inspection as well as suggested changes for test limits. As the number of test cases often run into thousands, such specialized re-inspection rules cannot be setup in any feasible way on the full test set. However, our statistical system identifies a handful of such test cases, which can therefore be subjected to further scrutiny.

Other than the system description, we describe here two types of field performance. First, we consider a case of linearization faults which is available as an indicator on a specific radio and a specific port. Second, we consider a case where we look at field performance KPIs such as Channel Quality Indicator (CQI), Signal-to-Inference-plus-Noise Radio (SINR), Block Error Rate (BLER) etc. These are not available as binary performance flags and we look at them from the perspective of their measured values. It is not surprising therefore that our results for the latter are less prescriptive than those for the former. However, in both cases our system provides a scientifically robust process to narrow down from a large volume of test cases to a manageable subset, which can then be handled by expert teams judiciously.

We would like to highlight two aspects of our study at the outset. Firstly, our system only provides insights backed up with statistical evidence for radio product engineers consideration. The goal of the data driven approach is focused towards assisting radio engineers in designing test criteria's. Our results are therefore purely prescriptive and consultative and any decisions on changes in actual process or test limits remain the sole discretion of the radio product design and engineering teams. Secondly, our system is now setup for analysis of radio equipment only. However this framework can be extended to include any equipment within the telecommunication domain, provided relevant data is available. We also believe that certain aspects of our approach could hold relevance for other industries as well and hence we would encourage colleagues in other industries to explore its applicability within their respective domains.

II. Caveat

We digress from the relatively standardized sections in academic literature to point out some caveats in this study. First, we would need to keep the details of the radios in question, specifics on test cases, test limits and KPIs anonymous for confidentiality purposes. We would not be able, now or in the future, to respond to any questions on specifics regarding the above nor share even small samples of data or code to the community.

One may therefore ask the rationale behind making this study public. We would like to inform the broader telecommunications community the value of building data driven manufacturing test cases as from our experience this is an opportunity in many telecommunications products and manufacturers. As each one of us are finally consumers of telecommunication equipment in various ways in today's digitally connected world, a more scientific system leading to better manufacturing processes would lead to a better experience for everybody. We would also like to hear from our colleagues in the industry, including our competitors, to understand their best practices in manufacturing and testing. Finally, we would like to highlight the value of combining data driven methods with domain expertise - in a world where statistical techniques, machine learning (ML), artificial intelligence (AI) and generative AI (GenAI) is going to be dominant, this combination will be more and more important and we would like to use to highlight how a data driven and a domain expertise approach complement each other in one use case.

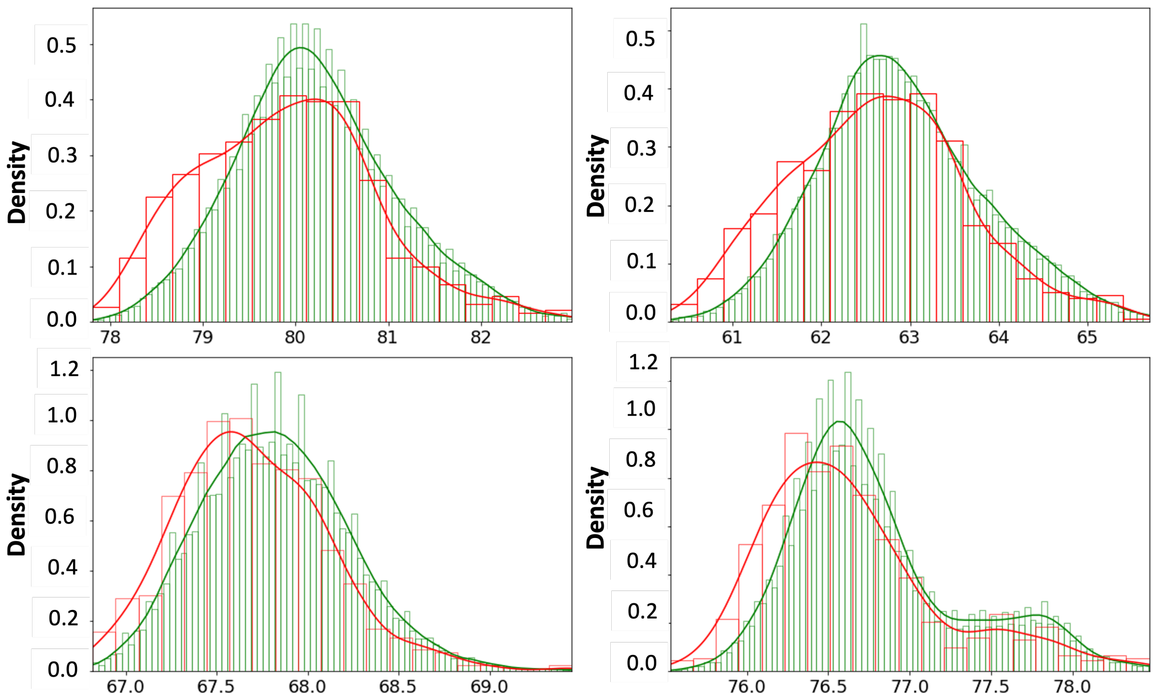


Fig. 2: Density plots for some test cases which have statistically significant differences between ports with linearization fault in red ("bad port") and those without in green ("good port"). These tests measure different aspects related to Out of Band Unwanted Emissions (OBUE). Statistical test is KS test with $p < 0.05$

III. Literature Review

There is very little work on the correspondence between field tests and network performance. Zeng at al [1] look at the correlation between Error Vector Magnitude (EVM) and various Key Performance Indicators (KPI) like SINR, average cell spectrum efficiency (SSE), cell edge cell spectrum efficiency (ESE) and Modulation Coding Scheme (MCS). Hasan et al look at grouping of cells based on EVM to improve channel state information(CSI) performance and in a separate research [2] try to maximise spectral efficiency by optimizing the EVM scheme.

IV. System Description

In this section, we describe the source data, the types of problems investigated and the statistical tests which are relevant for the study.

A. Source Data Description

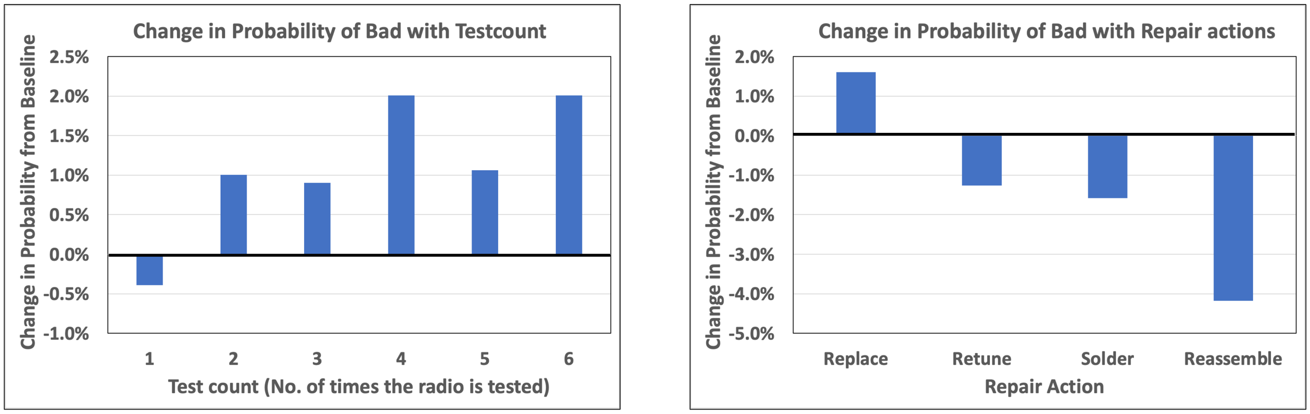
There are two main data sources used for this study. As described above, every radio goes through large number of test cases, often running into thousands. This data is stored in a data warehouse and accessed via a pyspark system [3]. This data is at the grain of every test case run for every radio - test cases which fail for a certain radio or need to be re-run because of dependencies will have multiple records. Also many test cases are run separately for every port of a radio - this in turn sets the grain of the data for such test cases at the level of the port.

The second data source consists of field performance records. This in turn has two types - log file based and KPI based. Multiple radio log files are collected and these captures events, alarms, faults and regular measurements. Of particular interest is linearization fault happenning due to power amplifiers working in the non-linear operating zone [4]. Such logs are event trigerred and therefore irregularly spaced in time. KPI performance measures are measurements of KPIs like throughput, SINR, CQI etc. These are computed from low level telecom events at 15 minute intervals and are therefore regularly spaced. Also unlike faults and alarms, these events are collected for radios in both good and bad operating conditions. Also, these KPIs are collected at a cell level - one radio typically has more than one cell.

B. Types of Problems investigated

While our system is suitable for processing a wide variety of problems, we looked at two important types of problems.

1) Linearization fault: Radio units have a Power Amplifier (PA) which amplifies the input signal for better coverage in a Over-The-Air (OTA) transmission. However, for maximum efficiency power amplifiers need to operate in non-linear operating zone (Type B, Type AB, Type C etc.) [5]. However, the increased efficiency comes at the cost of linearity which in turn introduces effects of loss of gain and intermodulation effects [6]. To overcome this problem, power amplifiers are typically preceded by a Digital Pre-Distortion (DPD) unit [7] which pre-distorts the signal so that the net effect post the power amplifier is a linear gain. Despite this, sometimes a radio unit works in



(a) (b)

Fig. 3: Change in conditional probability w.r.t baseline shown as 0 for linearization fault under various testing conditions for one type of Radio. All ratios more than the baseline are statistically significant ($p < 0.05$) by a one-sided t-test. (a) Change in conditional probability of a linearization fault for various different numbers of re-tests. (b) Change in conditional probability of a linearization fault for various different repair actions

the non-linear operating zone and this causes what is known as a linearization fault. This typically happens on one port of a radio and is identified as a hardware fault on the radio. Whilst operating conditions like power settings and load do impact the linearization issues, we would like to understand if there is a correspondence on manufacturing issues for the same. In particular, we would like to understand if some test cases have a statistically significant correspondence with hardware failures for linearization fault.

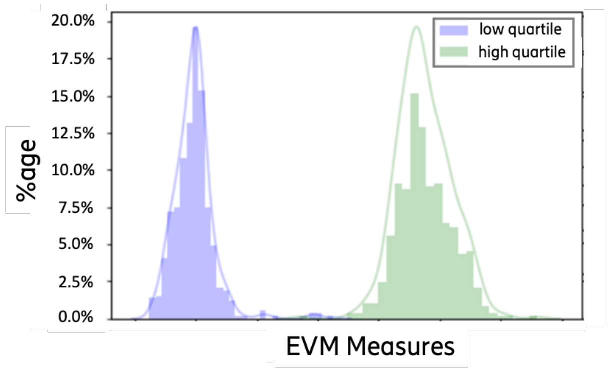
2) CQI and EVM correspondence: The error vector magnitude (EVM) [8] is a key metric of the quality of digital modulation in modern wireless communication systems. In the Long-Term-Evolution (LTE) or 4G system, there is a certain difference in amplitude, phase and frequency between the digital modulated signal received by the user equipment and the ideal signal due to the modulation error of the modulator. The root mean square (RMS) of all error vector magnitudes between the received symbol locations and their closest ideal constellation locations constitute the EVM value of the device. Manufacturing process tests for the EVM and checks that it is below a certain threshold. EVM is tested at a port level. As this is a measure of the modulation scheme, a higher EVM will lead to a poorer modulation and therefore a poorer channel quality. This may be reflected in various KPIs like CQI, SINR, BLER etc. These KPIs can of course be poorer because of location, load and other deployment parameters. However, a larger EVM would contribute towards a generally poorer performance of some or all of these channel quality metrics. Hence we would like to to understand via statistical methods if there is a significant correspondence between EVM at manufacturing test time and field performance of KPIs.

C. Statistical Methods

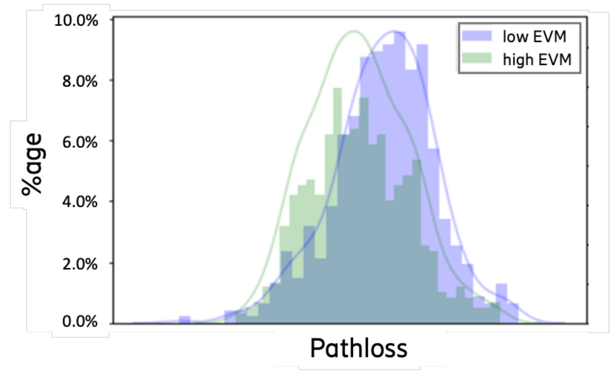
We follow different statistical methods for the two problems above - this is because of the variety of the data and the necessary processing needed for statistical experiments. In this section we would describe the processing and statistical methods we followed for each of them.

1) Linearization Fault: We identified the port of the radio on which linearization fault occurred and considered them as ”bad downlink (DL) performance ports” (”bad ports” in short) and considered all other ports on the radio as ”good DL performance ports” (”good ports” in short). We considered radios without any linearization fault and considered all ports for them as ”good DL performance ports”. We then looked at the test case measurements for these ports from the manufacturing test data. We ran a Kolmogorov–Smirnov (KS) statistical test [9] to compare the performance for good and bad ports and identified tests which have statistically significant differences between good and bad ports. Radios standing out from the normal distribution on these tests can thus be subjected to more careful scrutiny and testing by experts before shipping them to customers for deployment. These are valuable inputs to the test design team to choose better test limits. This is especially true for new radios being designed for which we do not have any field results at all.

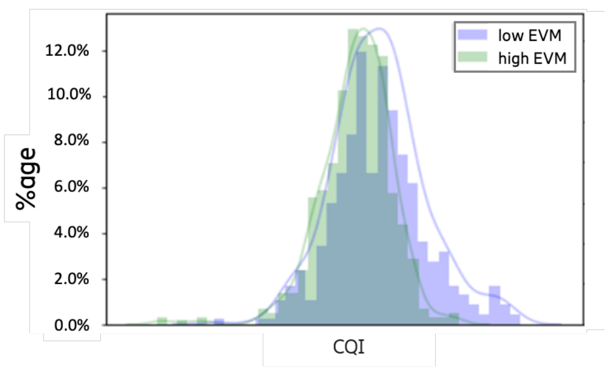
In addition, we looked at the number of re-tests, affected component and repair action. For each of these parameters individually we compared the baseline downlink failure (linearization fault) probability vs. the conditional probability of linearization fault given each of these criteria. For example, for number of re-tests between 1 and 6, we computed $P(linearization\ fault|no\ of\ retests = k)$ where $k \in \{1, 2, 3, 4, 5, 6\}$ and compared this with the baseline probability $P(linearization\ fault)$. Similarly we looked at $P(linearization\ fault|repair\ action)$ where $repair\ action \in \{'Replace','Retune','Solder','Reassamble'\}$ and compared it with the baseline probability. We also did similar experiments for the individual test case identifier and the affected component identifier.



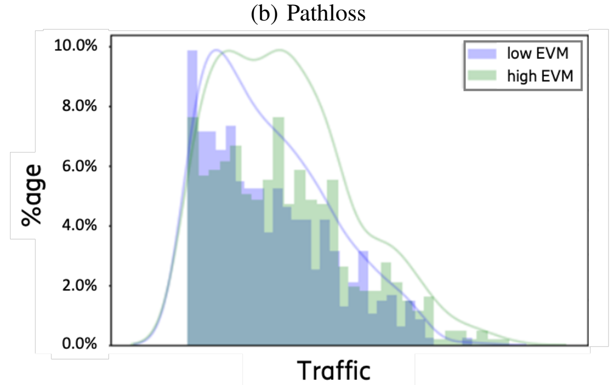
(a) EVM Measurements



(b) Pathloss



(c) CQI



(d) Traffic

Fig. 4: Density plots for different pathloss, CQI and traffic KPIs averaged over 3 months for radios having average EVM in the bottom and top quartiles. Bottom quartile is shown in purple and top quartile is shown in green

2) EVM and CQI: EVM is measured at port level on radio units and KPIs are measured at cell level in the field. The mapping between radio units and cell is known. In order to be able to analyze the EVM with KPIs data, we used average EVM value from all ports on a radio unit. A radio product with large variation in the EVM measures from the testing process was picked for the analysis. Two contrastive sets of radio units were chosen from the EVM distribution - one from the bottom quartile and the other from the top quartile. These two sets were statistically compared for CQI, Traffic and Pathloss. KPI measurements are usually available at 15 minutes interval. KPI data worth one month was analyzed for this study. Average values over time of KPIs are computed per cell for the statistical analysis. Statistical analysis showed that radio units in the top EVM quartile have lower CQI compared to the ones in the bottom EVM quartile. This was the observation considering the influence of traffic and pathloss on the channel quality.

V. Experimental Results

Although the focus of this paper is describing a system to combine data from manufacturing and field performance and how manufacturing test limits can be tweaked using field performance, we present some results on our experiments so far to highlight the value of our approach. As discussed in Section II we would not be able to identify exact radio products etc. for confidentiality purposes. All of these radios are for outdoor transmission in LTE spectrum.

Table I show the counts and proportions of linearization faults for two different radios. As we can see, that the proportion of radios having linearization faults are extremely low. Figure 2 shows the difference in the density plots for some test case measurements between ports with linearization fault (”bad ports”) vs. those without linearization fault (”good ports”). Fig. 3 shows the change in conditional probability of linearization fault with respect to the baseline for various testing conditions. Fig. 3a shows the same for various different test counts and Fig. 3b or various different repair actions.

Our results are able to pin-point test cases (in single digits out of thousands) which can be evaluated by the design team. Similar recommendation is made for radios with more re-tests and needing part replacement as a repair action. Thus our system allows a very efficient way to identify important test cases and test actions of relevance for evaluation by the design team.

For testing relationship between EVM and CQI metrics, we consider one radio and look at the test case measurements on each port. The EVM is measured at manufacturing time and CQI metrics are measured on field.

Considering radios having EVM, averaged across ports, in

	No of units	Prop. units with fault	Number of test Cases	No. of statistically significant test cases
Radio 1	28204	0.5	749	4
Radio 2	16846	1.0	967	4

TABLE I: Summary statistics for Linearization Fault Experiments - Number of units, approximate proportion (%) of units with linearization fault, number of test cases and no of test cases with statistically significant (p<0.05) differences (KS test) between ”good ports” and ”bad ports”

bottom and top quartiles we look at some of the KPI measures for three months. Our results indicate negative correlation between EVM and CQI considering the influence of network behavior on the KPI. This means that the radio units tested with high EVM measures tend to have lower CQI in the field (Fig. 4). This is expected since EVM is a measure of error whereas CQI metrics measure quality. CQI in the field may also be influenced by other network factors such as traffic, pathloss. In Figure 4 we plot in purple the distribution of radios having average EVM across ports in the bottom quartile and in green those having average EVM across ports in top quartile. We observe that the traffic patterns for the two distributions are not statistically different ($p > 0.05$) as per KS test. However, the pathloss for the two distributions are different ($p < 0.05$) according to KS test. Similar statistical differences are observed for CQI. However, the pathloss for the low EVM is higher which should lead to a lower CQI distribution - pathloss being a measure of loss and CQI being a measure of quality. The observed variation can thus possibly be attributed to the variation of the EVM. To test this, we regress the CQI measurements with the EVM measurements and observe a statistically significant ($p < 0.05$) correlation coefficient of -0.55. This indicates there is a statistical correlation between EVM measures and field performance. This gives the radio engineers insight for choosing the appropriate test limits especially for new radio products.

VI. Discussion

In this section, we discuss our observations and limitations of our study.

A. Observations and Limitations

To the best of our knowledge, there has been no published study around the correspondence of field performance and manufacturing test results for telecommunications products. This approach provides a principled way to design test limits, actions on specific scenarios like re-tests etc. This is especially useful when building new radio products. Learning from existing radios in field can be used onto the new radio roll outs if the new products are using similar technology e.g. similar power amplifiers. These correlations can be used to find out EVM values at which CQI can hit threshold given network conditions. We would like to highlight that all of these equipment are working at high levels of performance - but this approach would allow imporve it further leading to better end user experience.

There are various further enhancements possible with our study to make it more comprehensive. For example, we have looked at statistical correlations of linearization faults with respect to one test case at a time. A multivariate approach is one obvious extension in this regard. In a similar way, we look at re-tests and repair actions individually whereas one can possibly look at them jointly. Similarly, instead of taking average EVM values across ports for a radio and average KPIs over time, one can look at a more fine grained approach. However, this would necessitate more advanced models including use of machine learning models to identify correlations. Such approaches would need a much larger sample of data having various issues and therefore maybe impractical for real world telecom products. In short, we describe the first step towards relating field performance and manufacturing test results.

B. Future Work

Other than handling the limitations above, our approach can be extended in the future to across different components in the telecommunications environment like baseband units, transport layer, core network etc.

VII. Acknowledgments

All authors contributed equally to the work.

The authors would like to thank Rikard Bauer and Richard Franzén for support and guidance during the work as well as detailed review comments of the paper. The authors would like to thank other colleagues and management in Ericsson who were directly involved in related components for this work.

References

[1] Z. Zeng and M. Shao, “Impact of evm on network performance,” in *2018 International Conference on Sensor Networks and Signal Processing (SNSP)*. IEEE, 2018, pp. 13–16.

[2] W. B. Hasan, P. Harris, A. Doufexi, and M. Beach, “Real-time maximum spectral efficiency for massive mimo and its limits,” *IEEE Access*, vol. 6, pp. 46 122–46 133, 2018.

[3] T. Drabas and D. Lee, *Learning PySpark*. Packt Publishing Ltd, 2017.

[4] C.-P. Liang, J.-h. Jong, W. E. Stark, and J. R. East, “Nonlinear amplifier effects in communications systems,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 47, no. 8, pp. 1461–1466, 1999.

[5] N. O. Sokal, “Rf power amplifiers, classes a through s-how they operate, and when to use each,” in *Professional Program Proceedings. Electronic Industries Forum of New England*. IEEE, 1997, pp. 179–252.

[6] J. J. Carr, “Chapter 16 - radio receiver basics,” in *The Technician's EMI Handbook*, J. J. Carr, Ed. Woburn: Newnes, 2000, pp. 163–195. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780750672337500168

[7] A. Rahati Belabad, S. A. Motamedi, and S. Sharifian, “An adaptive digital predistortion for compensating nonlinear distortions in rf power amplifier with memory effects,” *Integration*, vol. 57, pp. 184–191, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167926017300032

[8] E. Acar, "How evm measurement improves system-level performance," Analog Devices, accessed: 12th September 2023. [Online]. Available: https://www.analog.com/en/technical-articles/how-evm-measurement-improves-system-level-performance.html

[9] V. W. Berger and Y. Zhou, "Kolmogorov–smirnov test: Overview," *Wiley statsref: Statistics reference online*, 2014.