Measurement Procedures for Design and Enforcement of Harm Claim Thresholds

Janne Riihijärvi and Petri Mähönen
Institute for Networked Systems, RWTH Aachen University
Kackertstrasse 9, D-52072 Aachen, Germany
email: {jar, pma}@inets.rwth-aachen.de

J. Pierre de Vries
Silicon Flatirons Center
University of Colorado, Boulder
email: pierre.devries@colorado.edu

Abstract—Harm claim thresholds (HCTs) are a promising approach for regulators to specify interference limits in a technology-neutral fashion, and a useful parameter spectrum access systems can use to manage the aggregate interference caused by transmitters they control. However, existing literature provides very little guidance how HCTs should be set and enforced. In this paper we propose a detailed regulatory framework for gathering and processing of measurement data for enforcing and setting harm claim thresholds. We introduce the central concepts of stratification and weighting of measurement data, and show their importance in ensuring representativeness of measurements and enabling robust estimation of statistical confidence on results. For deriving HCT thresholds from measurements, we propose additional representativeness criteria that a regulator should apply to avoid underestimation of field strength levels related to existing wireless services. We demonstrate application of our proposed framework using an extensive drive test data set, and show that the chosen HCT percentile is critical in determining how much data needs to be gathered for enforcement. We also discuss the various design choices and parameters needed by our framework, and show through examples how they can be derived in cooperation with the different stakeholders. The developed framework has several applications beyond HCT enforcement, some of which are briefly described in the paper.

I. INTRODUCTION

Harmful interference can be caused by unwanted transmissions and/or poor receiver performance. However, regulatory mandates of receiver performance are at best difficult to formulate and at worst inefficient. They are very rare, even though they are much talked about since receivers that cannot reject interfering signals can preclude or constrain new spectrum allocations. Harm claim thresholds (HCTs) were introduced as a way for regulators to specify the interference environment in which an affected radio systems should expect to operate, without specifying receiver performance [1].

A HCT defines the in-band and out-of-band interfering signals that must be exceeded before a system can claim that it is experiencing harmful interference. It is deemed to be exceeded if the measured or modeled signal strength exceeds a stated value for more than some percentage of locations and times, for a prescribed confidence level. In the example we use throughout this paper, the HCT is 50 dB(μV/m) per MHz to be exceeded at no more than 5% of locations, at the 95% confidence level.1 This is equivalent to imposing the threshold on the 95th percentile of field strength.

HCTs are particularly useful in managing dynamic spectrum access since they allow a spectrum access system to ignore the locations and (more importantly) the characteristics of individual receivers when calculating the allowed transmit power of the radios it controls. As long as the aggregate resulting signal strength statistics meet the HCT, then by definition harmful interference has been avoided. This facilitates the implementation of spectrum access systems and policies.

The first HCT papers provided little guidance for specifying the regions over which threshold measurements should be made. Reference [2] suggests measuring every 250 m over a 5 km × 5 km area, while [3] suggests an area of 1 km × 1 km with measurement points on a 50 m grid; neither justify their protocols beyond the need for a sufficiently large data set to provide statistical reliability (n = 400 in both cases). In [4] we first studied the statistics of drive test measurements that could be used to make harm claims. We showed that assuming independent measurements some hundreds of measurements were needed for sufficient confidence for the 95th percentile of field strength, with much higher measurement counts needed for the 99th percentile. However, these results do not yield information on the needed measurement area, and the independence of measurements is a strong assumption that is useful for initial analysis but (as we shall illustrate in the following section) is untenable for actual claim determination.

The objective of this paper is to propose a concrete regulatory framework for measurements to be conducted in relation to a claim for harm under HCT policy. We will also explain how a regulator can formulate HCTs in a way that can be readily used by licensees making or defending claims of harmful interference, or wishing to ensure that their deployments are in compliance. It will also use a much larger data set than the previous study [4] to show how a regulator could derive HCTs for a cellular deployment.

1We emphasize that the value of 50 dB(μV/m) per MHz is chosen for illustrative purposes only. In particular, we have chosen a value that is sufficiently low to enable us to demonstrate how a HCT violation would be determined from measurements, using realistic data obtained from an actual operational network. We discuss in the latter half of the paper how the regulator would actually select the threshold reliably so as not to place existing network deployments in violation.
This paper will use cumulative case study examples to motivate and illustrate the regulatory specification which is laid out in Section VI. After giving the regulatory design goals in Section II, we illustrate the limitations of naïve measurement specifications in Section III. In Section IV we work through the key elements of measurement procedure design. Section V discusses confidence intervals, and Section VI explains how a regulator would go about choosing a threshold level, stratification distance and confidence interval in the first place. Section VI also summarizes the parameters a regulator would specify, and the key steps a claimant would have to take to show a HCT has been exceeded.

II. REGULATORY GOALS

The premise of the HCT approach is that a regulator would define thresholds at a level of generality comparable to the way it specifies emission limits. Operators would then be able to use the values in such rules to ensure that their operations would not trigger a harm claim, or conversely, to prove a harm claim against another party. Since HCTs are a regulatory tool, they need to be easy to understand and use by all concerned. We believe that goal can be achieved if the following design objectives are met:

- Straightforward to specify at a high level in rules, e.g. a small number of technology- and service-neutral parameters; implementation details can be promulgated in ancillary publications, e.g. bulletins in the FCC OET Knowledge Database.
- Relatively easy to accommodate new technologies, e.g. by updating regulatory bulletins not changing rules.
- Easy to understand and apply, and in particular should not require sophisticated knowledge of statistics.
- Contain as few parameters as possible.
- Based on ex ante stratification distances (one of the key parameters in our method, introduced in detail in the following two sections) rather than estimates derived in the course of a continuous drive test.
- Enable simple estimation and planning of measurements.

Two key questions immediately arise:

1) What is the simplest way the regulator can specify a HCT and still meet its regulatory goals? This includes choosing both the parameters to be used, and their values.

2) Given the HCT parameters and values, how would an operator go about proving that a HCT is (or is not) met?

When specifying a HCT, the regulator needs to determine the statistic of interest — we select a point on a cumulative distribution function (CDF) of field strength — and the confidence level with which the statistic needs to be measured.

III. WHY SPECIFIC PROCEDURES FOR GATHERING AND PROCESSING MEASUREMENT DATA ARE NEEDED

Let us begin by considering the data set shown in Figure 1, which corresponds to a 10 km x 10 km subset within a large drive test campaign conducted in a major metropolitan area in the US. The full data set covered an area of approximately 120 km x 170 km and consisted of 4 million measurements in total. The gathered measurements are all for the 2 GHz downlink band for a major cellular operator. We will first use this data set to illustrate the pitfalls of naïve statistical analysis when applied to drive test data, and will also use it later on to motivate our suggested approach for data gathering and processing.

The simplest approach for testing whether our data set indicates a violation of the HCT would be to ignore any dependencies in the data, and treat each of the 7266 measurements as an equally valuable source of information. We can compute the 95th percentile from the data directly, or read it from the empirical complimentary cumulative distribution function (CCDF) as shown in the Figure 2 and compare it against the set HCT. As can be seen from the figure, the estimated value of 46.7 dB(µV/m) per MHz for the 95th percentile is well below the harm claim threshold, indicating that a violation is unlikely. Further, computing the confidence interval for this percentile estimate using the measurement count n = 7266 yields “error bars” of well less than one dB, indicating high confidence on the obtained estimate.

But how good is the obtained estimate really? In this case, since we have used a small fraction of a much larger data set, the best comparison is against the 95th percentile value obtained from the entire data set. If our 10 km x 10 km sample would be representative, and our data processing method sound, these two estimates should agree relatively closely. Unfortunately the 95th percentile for the entire data set turns out to be 53 dB(µV/m) per MHz which actually exceeds the HCT, showing that the tentative conclusions drawn from the data set of Figure 1 are incorrect. The discrepancy becomes even larger when we consider the field strength measured.

Fig. 1. Example drive test data set.

2We emphasise again that even though we use throughout this paper the value of 50 dB(µV/m) per MHz as the putative HCT, not to be exceeded more than five percent of locations, it is not intended to suggest an actual HCT value. We discuss in later sections the procedures a regulator should actually use in order to derive suitable HCT values.

3We will discuss the exact procedure for this in Section V.
by a typical user which turns out to be approximately 55.8 dB(μV/m) per MHz. This strongly indicates that such a naïve statistical treatment of the drive test data can result in highly misleading results. In the following section we will discuss the origins of the observed discrepancy in more detail, and also formulate procedures for dealing with drive test data that corrects the shortcomings of this naïve treatment.

IV. MEASUREMENT PROCEDURE DESIGN

We begin by arguing that the discrepancy observed in the previous section arose for two distinct reasons. First, we treated each measurement as being independent of every other. But by looking at Figure 1 we see only a few regions of higher field strength, suggesting that only a few transmitters have been deployed in the region. Measurements from any two nearby locations are therefore heavily correlated, and this must be taken into account when estimating the effective measurement data count. Second, much of the drive covers highways or other major road segments, making it unclear how representative these measurements are for field strength experienced by a typical user. We can get insight into this by studying the population density at the different measurement locations which, as discussed in footnote 4, we use as a proxy for user location.

In order to remedy these two shortcomings of the naïve measurement protocol, we argue that the following two key concepts need to be incorporated into the processing of drive test data:

1) We introduce a standard process called stratification to reduce the measurement data set to a much smaller subset in which measurements are approximately independent of each other. This can be done, for example, by imposing a minimal distance for any two measurement points that the reduced data set must satisfy. Stratification will be discussed in detail in Section IV.B.

2) Further, in order to ensure that the statistics are representative, we further introduce weighting of measurement values, in which each measurement carries a weight that is proportional to the density of the foreseen user population at the measurement location. Weighting will be further discussed in Section IV.C.

The purpose of these additions is to fairly estimate the number of measurements for confidence interval computation, and to help ensure representativeness.

Figure 3 illustrates the outcome of applying these two steps to our first example measurement data set. Stratification has drastically reduced the number of measurement points to consider (from n = 7260 down to n = 67), and the population-weighting also shown in the figure further reveals that only a small fraction of measurement points actually contribute to estimating interference on the expected user population. Thus, it is not really surprising that the naïve statistical approach resulted in poor results. In fact, when computing the confidence interval using the new effective measurement count, it turns out to have a formally infinite length, strongly indicating that no statistically reliable conclusions about the 95th percentile of field strength can be drawn from the original data set.

A. Example Application to More Extensive Data Set

An obvious question to ask is whether the concepts of stratification and weighting can be practically applied to drive test data at all given the negative results obtained above? By way of example, let us consider another subset from the same large-scale metropolitan measurement campaign as our original example. This alternative data set is illustrated in Figure 4 both in terms of samples remaining after stratification as well as the corresponding population-weighting. Clearly despite the same measurement region size, the new data set is much more dense even after stratification (with n = 260 samples remaining), and also much more representative as evidenced by the homogeneity of population weights.

When the stratified data set is combined with the population weights, we obtain the weighted CCDF shown in Figure 5, now indicating a possible violation. Note also that the new point estimate of 56.5 dB(μV/m) per MHz is quite close to the “ground truth” of 55.8 dB(μV/m) per MHz obtained by considering the entire data set. Note that these estimates are still missing the needed confidence intervals, the addition of which we will discuss in Section V. However, before proceeding there we discuss the roles of stratification and population-weighting in slightly more detail. In particular, our objective is to convince the reader that these are not ad hoc measures, but are instead based on sound statistical reasoning.

Footnote 4: By “as measured by typical user” we mean the distribution of field strength a randomly selected member of the user population would be expected to measure, as opposed to simply considering the distribution of field strength over space. These two distributions can differ substantially, since users are not uniformly spread around any larger region. In the following we always assume that the user locations follow the overall population density, estimates of which we obtained from the ORNL LandScan global high-resolution population database [5].
the effects of both the traffic lights as well as overall speed of measurement locations for our second example data set, where and complex routes that may cover individual roads more than drives, where traffic lights, congestion, varying speed limits, is particularly important for processing data from evidence interval estimation. Second, it also helps to reduce inhomogeneity in the data, ensuring that the measurements cover the intended region roughly uniformly. The latter function is particularly important for processing data from urban test drives, where traffic lights, congestion, varying speed limits, and complex routes that may cover individual roads more than once, all lead to highly uneven measurement densities over space. Figure 6 illustrates this by showing the local density of measurement locations for our second example data set, where the effects of both the traffic lights as well as overall speed of the test vehicle are clearly visible.

Stratification can be implemented in several ways, each with different trade-offs between efficiency in terms of the number of retained measurement locations, and computational burden imposed. One possibility would be to consider the selection of retained measurement locations as an optimization problem, where we would seek to maximize the number of retained points under the constraint that no two locations would be closer apart than a set stratification distance\(^5\) \(d_S\). Unfortunately this is not computationally feasible as the number of subsets of locations to consider increases exponentially with overall measurement data set size. A slightly less efficient but

\(^5\)We will explain how a suitable stratification distance can be determined through measurements or simulations in Section VI-B
also computationally lighter method would be analogous to how carrier sense multiple access (CSMA) medium access control protocols operate. We would assign to each measurement location a random number (“arrival time”) uniformly from the interval [0, 1], and retain a location if and only if it has the smallest arrival time of all the locations within distance $d_S$. The most complex part here is the computation of distances between the points, requiring $O(n^2)$ operations.

An even more light-weight approach is illustrated in Figure 7, where the measurement region is divided into squares with side length $d_S$, and only one measurement from each square is retained. In our example we have chosen the location that is farthest away from the edge of the surrounding square, but other similar choices could be made as well. This is the approach we have used on the preceding data sets when constructing the stratified equivalents. Its downside compared to the previous ones is that it does not guarantee that the distances between nearby measurement locations are strictly larger than $d_S$, but this can be mitigated by a slightly more conservative choice of the stratification distance. Since only comparisons of coordinates and distances to a fixed number of squares need to be computed, only $O(n)$ operations are needed.

Notice that the three algorithms described above differ in other fundamental ways besides just their computational costs. For instance, the arrival time algorithm results in a random stratified location set, whereas the grid-based algorithm is strictly deterministic once the division of the measurement area into squares is defined. Our examples in Figures 3 and 4 use grid-based stratification.

**C. Interpretation of Weighting when Estimating Percentiles**

As discussed above, the purpose of weighting is to ensure that the percentiles estimated are representative of what the population of interfered users would measure, as opposed to raw spatial estimates. No weighting would be needed if the interfered-with population itself would conduct the measurements, or if the measurement locations would be carefully selected to follow the corresponding spatial density after stratification. While theoretically possible, we believe such an approach to be overly complex. Instead, we propose to first obtain a spatially uniform sample of sufficient size (by conducting a conventional drive test followed by stratification), and then weight that sample with the estimated interfered-with population density when computing the HCT percentile for the field strength.

Weighted percentiles have an appealing geometric interpretation illustrated in Figure 8. In the figure we have applied to each (stratified) measurement location the threshold of

\[ \text{Field strength \ dB(uV/m)/MHz} \]

Fig. 5. The population weighted and stratified CCDF from the data set of Figure 4.

Fig. 6. Illustration of the inhomogeneity of the density of measurement locations arising from changes in velocity during a typical drive test, in particular induced by traffic lights showing up as small dark spots in the plot.

Fig. 7. Example of a grid-based method for implementing stratification.

6Crowdsourcing is of course a possibility here, but conducting extensive spatial field strength measurements through user terminals with highly varying receiver qualities and configurations in a reliable fashion is still very much an open research problem.
50 dB(μV/m) per MHz, with black circles corresponding to locations that were measured below this value, and red circles exceeding. For testing whether the unweighted percentile estimated from the data would exceed the 95th, it is sufficient to check whether at least five percent of the dots are red. If yes, the measured 95th percentile has to exceed the threshold. When weights are applied (represented in the figure by the area of the individual points), we compare the total areas of the points of the two colors instead of simply the counts.

V. ADDING CONFIDENCE INTERVALS

In the previous sections our focus has been on how to obtain reasonable estimates for the percentile of field strength given in the HCT specification, together with the effective measurement count \( n \) arising from our stratification procedure. In this section we discuss how to further construct the confidence intervals around these estimates; they can be used to quantify the reliability of our decision on whether a given set of measurements indeed can be determined to indicate HCT violation or not. We begin by briefly discussing the precise meaning of confidence intervals as they are often misinterpreted in the literature. We then outline how confidence intervals can be estimated first for unweighted measurement data, and then for weighted data as well. Finally, we give examples of the computations involved using our two example data sets.

A. Meaning of Confidence Intervals

Confidence intervals are an example of a so-called frequentist statistical tool, designed to have rigorously defined properties as a statistical method applied a very large number of times. In the case of a confidence interval for the \( p \)th percentile, if we repeat taking a random sample from a given large set of data, the true \( p \)th percentile of the underlying data would fall within the \( 100(1 - \alpha)/\% \) confidence interval with probability \( 1 - \alpha \). This guarantees that even though for a single measurement run the true percentile can fall far outside the computed confidence interval in rare cases, the fraction of such errors is guaranteed to converge to \( \alpha \) which we can make as small as desired (at the cost of more measurements being needed in order to obtain tight enough confidence intervals).

Figure 9 illustrates this statistical meaning of confidence intervals over 30 samples (with \( n = 260 \) for each sample) from the overall measurement data set, and the associated point estimates and confidence intervals at the \( \alpha = 0.05 \) confidence level (all of these were computed using the procedures introduced in the following section). Approximately once every 20 measurement runs we would expect the true value (indicated by the dashed horizontal line) to fall outside the estimated confidence interval. This indeed corresponds well with the results shown in the figure, with one run out of 30 (shown in red) resulting in a significant overestimation of the percentile along with its confidence interval.

B. Confidence Intervals for Unweighted Data

How can we in practice find “good” confidence intervals? We give in the following one of the standard statistical procedures, but note that several alternatives exist [6]. Before stating the computational procedures, we need to introduce a small amount of mathematical notation and some auxiliary concepts. Let \( X_1, \ldots, X_n \) denote our \( n \) measurements, i.e. field strength values measured along a drive path. The order statistics of these are the measurement values sorted in ascending order, denoted by \( X_{(1)} \leq \cdots \leq X_{(n)} \). Thus \( X_{(1)} = \min\{X_1, \ldots, X_n\} \), \( X_{(n)} = \max\{X_1, \ldots, X_n\} \), with the rest of the order statistics carrying information about the percentiles \( P_{X}^{-1}(p) \) of \( X \), where \( P_X \) denotes the distribution function of \( X \). For large enough sample \( X_{(p \times (n+1))} \) serves as an estimate for the \( p \)th percentile of field strength.

The confidence interval for the \( p \)th percentile can be constructed from \( n \) measurements in the unweighted case as follows. Letting \( Z_{p} \) denote the \( p \)th percentile of the standard
normal distribution $N(0, 1)$, the upper limit of the $100(1−\alpha)\%$ confidence interval is given by $X(u)$, where

$$u = p(n + 1) + \frac{Z_{1−\alpha/2}}{\sqrt{np(1−p)}},$$  

and the lower limit by $X(l)$, where

$$l = p(n + 1) - \frac{Z_{1−\alpha/2}}{\sqrt{np(1−p)}}. \quad (2)$$

If $u$ or $l$ turns out to be non-integer, the corresponding endpoint of the confidence interval can be determined by interpolating between the nearest adjacent order statistics. The confidence intervals shown in the example of Figure 9 were computed using these procedures.

We will need the one-sided confidence intervals, giving only the upper or lower bound for the error, rather than these two-sided confidence intervals since the HCT is given as a value not-to-be-exceeded. These are obtained simply by replacing $Z_{1−\alpha/2}$ by $Z_{1−\alpha}$ in the corresponding formula above, with other end point being $\pm\infty$.

C. Confidence Intervals for Population-Weighted Data

For weighted data the above procedure can be used with minor modification. Notice that the $u$ and $l$ values obtained as the interval endpoints have a direct interpretation as percentiles of the measurement data. In order to compute the corresponding confidence intervals for weighted data, we simply need to find the corresponding weighted percentiles as discussed in Section IV-C, and use these as confidence interval endpoints instead.

VI. MEASUREMENT-RELATED CONSIDERATIONS FOR SETTING HCT POLICIES

In this section we focus on the measurement-related considerations and decisions the regulator should be mindful of when setting HCT policies. We first discuss how the outlined measurement procedures can be applied by the regulator in the first place to arrive at a suitable field strength threshold, and also how the expected measurement burden influences the percentiles that can be reasonably used in relation to HCT policy. We also discuss how the stratification distance $d_s$ should be set, and what are the trade-offs involved.

A. Choosing the Threshold

Arguably the most critical part of a HCT policy is the choice of the threshold field strength that is not to be exceeded more often than the chosen percentage. Values that are too low could outlaw existing deployments, or greatly diminish the value of spectrum by making the operation of a viable number of concurrent transmitters impossible. Too high a value, on the other hand, offers insufficient protection, making the regulation ineffective.

When new services are introduced, simulation studies based on realistic system deployment and propagation models might be needed. However, when dealing with existing services measurement-based methods as proposed above could also be used by the regulator to arrive at a reasonable field strength threshold.

The only modification that seems unavoidable is the introduction by the regulator of additional controls on the representativeness of the measurements conducted or used. Even with population weighting, test drives in regions with sparse wireless infrastructure tend to result in underestimation of the field strengths compared to estimates from an entire metropolitan region. This is not a problem if the measurement region is selected by a plaintiff, since the introduced bias would in general be towards underestimating the field strength level. However, if such sparsely built regions would be measured by the regulator and used directly to set the HCT, the results could be disastrous.

A simple solution to this problem is to consider only measurements from a region that has at least average user density. In our framework this translates to only utilising data that comes from the areas of the drive where the stratified measurements carry a large enough total weight. Figure 10 illustrates how this simple principle works by showing a scatterplot of total weights and the measured 95th percentile of field strength for all the distinct 10 km x 10 km regions present in our data. We see that for areas with low total weights the measurements are highly varying (with the downward bias discussed above being clearly visible), but in the regions with higher aggregate weights the measured values agree well with each other. Thus a practical approach would be to use measurement from such a high-weight region, and impose an additional safety margin on top of the measured value. The dashed lines in Figure 10 correspond to a 10 dB safety margin being imposed on top of the “ground truth” 95th percentile value. Using out data set, this would lead to the regulator setting a HCT of 65.8 dB(μV/m) per MHz (or suitably upwards rounded number), rather than the 50 dB(μV/m) per MHz we have been using for illustrative purposes.

B. Setting the Stratification Distance

The stratification distance $d_s$ entails a trade-off between the degree of independence of the stratified measurements, and the effort needed to construct the test drive. Large values of $d_s$ ensure the most reliable percentile estimates and confidence intervals for a given measurement count $n$, but will also require very large regions to be covered in order to harvest the $n$ stratified measurements. As we saw in our initial case study (Figure 1), very small stratification distances risk on the other hand spurious conclusions to be drawn from the data. In this section we outline a statistical method enabling the derivation of $d_s$ based on “first principles”, including the formalization of the trade-off discussed above.

For our proposed method we first need to choose a similarity measure, which given a distance $r$ between two measurements $X_i$ and $X_j$ yields a quantitative measure of how similar $X_i$ and $X_j$ are expected to be. The preferred choice for such a measure in the spatial statistics community—whose research focuses on applications of statistical methods to spatially dependent data as is definitely case in our domain—is the semivariogram \cite{7}

$$\gamma(r) = \frac{1}{2} \text{Var}\{X_i - X_j\}, \quad (3)$$
smaller measurement data sets, as discussed in the text. However, the semivariogram enables additional diagnostics to be applied for small measurement count and covered region the two are almost equivalent statistically. X

resulting in small differences between them and correspondingly small values of $\gamma(r)$. For large $r$ the values of the semivariogram become larger and larger, until saturating at the overall variance of the data.

We can estimate the value of $\gamma(r)$ simply by finding all pairs $(X_i, X_j)$ of measurements where the locations are separated by distance $r \pm \Delta$, where $\Delta$ essentially defines a width of a histogram bin, followed by the computation of the variance of $X_i - X_j$ over all such pairs. An example of such a computation is shown in Figure 11, illustrating both the histogram nature of this estimator (bars) together with the estimates of $\gamma(r)$ assigned in the centres of the histogram bins (red dots). As can be seen from the figure, the obtained estimates are quite noisy, making them unsuitable for direct use in estimating a suitable stratification distance. Instead, we proceed by fitting a parametric model to the data. In the literature the exponential form $a + b \exp(-r/c)$ is commonly used [7], [8], and works well also for our example data set as shown in the figure (black line).

Once the parametric model for the semivariogram $\gamma(r)$ is obtained, we can choose a stratification distance $d_S$ based on how close the semivariance $\gamma(d_S)$ is to the asymptotic saturation value (overall variance of the data set). A very conservative choice would be to demand that at least 95% of the asymptotic level is reached, but as can be seen from our example this results in very large stratification distances, and thereby impractically large drive test data sets being needed for any HCT related claim. In our opinion a judicious tradeoff is to choose a somewhat lower fraction, such as having $\gamma(d_S)$ equal half of the saturation value, which results in slightly optimistic estimates of the stratified measurement

7The most commonly suggested alternative is the autocorrelation function $C(r)$, defined as the correlation coefficient between measurements $X_i$ and $X_j$ separated by distance $r$. For large enough data sets in terms of both measurement count and covered region the two are almost equivalent statistically. However, the semivariogram enables additional diagnostics to be applied for smaller measurement data sets, as discussed in the text. [7]

8It also might occur that the estimated semivariogram does not flatten out to an asymptotic value. This is a strong indication that the measurements are taken from a region that is too small, resulting in a trend in field strength values across the measurement region, which in turn presents itself as a constantly growing semivariogram values as $r$ is increased.
counts and thereby slightly tighter confidence intervals than can be expected in ideal conditions, but requiring at least an order of magnitude less data in order for the estimates to be obtained. For our example data set this prescription results in a stratification distance of $d_S = 305 \, \text{m}$, as shown in Figure 11.

The obtained stratification distance will depend strongly on the technology involved, deployment patterns, and the propagation environment. The obtained value of 305 m is typical for small cell and microcellular deployments in urban areas, while for rural macrocellular systems and broadcasting technologies much larger values would be expected.

Finally, we emphasise that while the above presentation is formulated in terms of measurement procedures, the definitions can directly be applied to simulation data and analytical models as well. While exact field strength levels obtained from such alternatives procedures tend to be unreliable, in our experience the correlation structure, whether expressed in terms of the semivariogram or autocorrelation function, is more reliable. Thus especially system level simulations with appropriate propagation modeling tools might be valuable in helping to find initial stratification distance even prior to large-scale deployments of wireless communication systems involved.

C. Specifying the Confidence Interval and Percentile

It is important to understand the relationships between the number of measurement points $n$, chosen HCT percentile, and the arising confidence interval lengths, as they are directly related to the expected size and complexity of the drive needed to gather the data. Figure 12 illustrates the relationship between $n$ and obtained confidence interval for the overall data set, again for 95th percentile with one-sided confidence intervals computed at the $\alpha = 0.05$ level. For each $n$ we generated 100 samples of $n$ measurements each, and computed the associated confidence interval lengths from the point estimate of the percentile to the lowest edge of the one-sided confidence interval, the distributions of which are shown as the box and whiskers plot. We see strongly diminishing returns from collecting more data, with the first 200–300 measurements being sufficient to reduce the confidence interval length to approximately 3 dB. We believe this to be sufficient accuracy given the vagaries of RF field measurements. This indicates that for the 95th percentile, having some 200–300 measurements after stratification forms a natural design target when planning a test drive. However, if the estimated field strength percentile is expected to be very close to the HCT one, arbitrarily large measurement counts might be needed to confirm or rule out HCT violation.

Another major factor in the number of data needed to achieve a particular confidence interval length is the underlying percentile for which the confidence interval is being estimated. Figure 13 illustrates the distribution of the achieved confidence interval lengths for different percentiles, in each case assuming $n = 260$ measurement locations as for our second example data set. We see that the confidence interval lengths increase rapidly as higher percentiles are estimated. For the 99th percentile $n = 260$ measurements no longer yield a finite confidence interval length, and for the 98th percentile individual confidence intervals can already exceed 10 dB. This strongly indicates that unless a significantly different measurement data processing method is adopted, imposing HCTs with higher than 95th percentile specifications on field strengths will require significantly more measurements (for both HCT determination and enforcement) than the few hundred we have found to be sufficient for the 95th percentile case. This will likely impose an unreasonable burden both on the regulator and on parties disputing a HCT claim.

From (1) and (2) we see that for a fixed data set and percentile $p$, the length of the confidence interval depends further on the percentiles $Z_{1-\alpha/2}$ of the standard normal distribution. This essentially defines the relationship between the length of the final confidence interval and the desired confidence level. However, this influence is decidedly smaller than those of the stratified measurement count and the chosen HCT percentile. For example, going from 90% confidence level ($\alpha = 0.1$) to 99% confidence level ($\alpha = 0.01$) increases...
the confidence interval length just by a factor of two (in terms of values in dB), despite the formally ten-fold reduction in the rate of spurious findings.

D. What a Regulator Needs to Specify

We conclude this section by summarizing what a regulator needs to specify in order to apply the method described here for determining and enforcing harm claim threshold policies. A high level summary of the different design parameters is given in Table I. We have divided these parameters into three categories in order to facilitate their specification to be split between rules yielding the high-level, unchanging requirements, and bulletins for more detailed and dynamic low-level specifications. This approach is similar to that used, for example, in the FCC’s regulation of wireless E911 location systems.

In our proposed split, the HCT policy details would be given in rules, consisting of the parameters comprising the original HCT specification format [1], [2]. Bulletins would then be used for specifying the measurement procedure related to the given HCT policy. Decisions involved here include the stratification procedure to be used (algorithm and choice of the stratification distance $d_S$), weighting to be applied (presumably defaulting to population density), and the various requirements for actually submitting the drive data that would form a basis of claim. The latter should be carefully designed to avoid manipulating the measurement outcomes, for example by submitting only selected subsets of larger measurement data sets. Possible design choices include prior registration of alternative drive paths from which the regulator randomly chooses one, combined with the requirement to submit complete raw log files of the entire measurement drive.

Finally, for setting HCTs for emerging services the regulator might need to specify the extent to which models and simulations can be used to derive these parameters. In particular, we see these tools to be particularly useful in deriving the needed stratification distances $d_S$, as such estimates are not sensitive to small imperfections in propagation models for example.

VII. CONCLUSIONS

In this paper we proposed a regulatory framework for gathering and processing of measurement data for enforcing and setting harm claim thresholds. Our proposal is based on the concepts of stratification and weighting, introduced to ensure representativeness of measurements and enabling robust estimation of statistical confidence on results. We also demonstrated, using data from an extensive test drive for a real wireless network, how our proposed framework could be used to set HCT thresholds and enforce HCT rules. The techniques described here can be transposed to other settings as well, such as setting service level agreements (SLAs) for cellular coverage. In the HCT case, the regulator specifies a high field strength level that cannot be exceeded by a transmitting licensee at more than a few locations without triggering harm claim. In the corresponding SLA case, a cellular operator would specify a low field strength level that can be fallen below only rarely by a tower operator without triggering an SLA default. Similar statistical considerations could be used by the FCC when providing guidelines for testing and verifying the accuracy of wireless E911 location systems.

### REFERENCES


### TABLE I

<table>
<thead>
<tr>
<th>Category</th>
<th>Parameters</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>HCT policy</td>
<td>Frequency band</td>
<td>2 GHz</td>
</tr>
<tr>
<td></td>
<td>Percentile of field strength</td>
<td>95th</td>
</tr>
<tr>
<td></td>
<td>Field strength threshold</td>
<td>50 dB (uV/m) per MHz</td>
</tr>
<tr>
<td></td>
<td>Confidence level</td>
<td>95% ($\alpha = 0.05$)</td>
</tr>
<tr>
<td>Measurement procedure</td>
<td>Stratification procedure</td>
<td>Grid-based</td>
</tr>
<tr>
<td></td>
<td>Weighting method</td>
<td>Population weighting</td>
</tr>
<tr>
<td></td>
<td>Submission of drive data</td>
<td>Complete without gaps</td>
</tr>
<tr>
<td></td>
<td>Responsibility for processing</td>
<td>Claimant</td>
</tr>
<tr>
<td></td>
<td>Requirements on equipment</td>
<td>Standard drive test</td>
</tr>
</tbody>
</table>

**Table I.** SUMMARY OF REGULATORY PARAMETERS WITH EXAMPLE VALUES USED IN THIS PAPER.