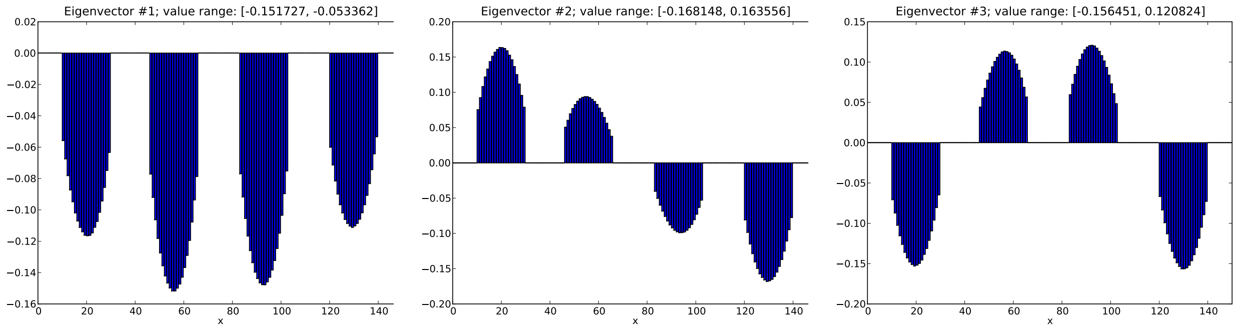

(a) Elements of the first eigenvector. (b) Elements of the second eigenvector. (c) Elements of the third eigenvector.

Fig. 8: Elements of the 3 leading eigenvectors of $\mathbf{A}^{\mathrm{T}}\mathbf{A}$ for the sensor placement of Figure 6.

(a) Actual sensor placement (ground truth). The single malfunctioning sensor is indicated by an arrow.

(b) Four measurement clusters (green, yellow, blue and red) determined by the sign-based spectral clustering. Sensors are shown at locations computed by MDSR from ranging measurements. Note the poorly localized bad sensor on the far left.

Fig. 10: Depiction of a 1D sensor placement scenario with 4 clusters.

Let us use the computed sensor location as the index. Given the computed sensor locations and measurement clusters shown in Figure 10b, we now apply conditions C1 and C2 of Section II to classify bad sensors as follows. For condition C1, we deem a sensor bad if it resides in a standalone cluster smaller than a certain threshold size S_{min}. For condition C2, we classify a sensor as bad if it is one of the τ % sensors furthest away from the centroid of the measurement cluster which contains the sensor, where τ is some small threshold. As we can see in Figure 10b, the bad sensor indicated by the arrow is at the edge of its cluster, so it can be correctly identified even with a minimal τ. Note that the distance metric used here is that of the index space as described in Section III. In this example, the distance for a pair of sensors is the absolute difference of their locations computed by MDSR.

This specific way of detecting bad sensors based on condition C2 relies on the following two assumptions. First, measurement clusters are stable under a few bad sensors. Second, caused by measurement errors the computed location of a bad sensor is normally pushed to the edge of its measurement cluster in the index space. We will have further discussion on this in Section V.

C. Simulation Results on Large Systems

In this section, we present simulation results on the model application with 400 sensors. We assume that the sensors are evenly partitioned into four groups of 100. We further assume that some randomly selected sensors are bad sensors, in the sense that all their measurements of other sensors have a random measurement error $\alpha \sim \mathcal{N}(\mu_{bad}, \sigma_{bad})$. That is, if the true measurement for a sensor-target pair is w, then the observed one is αw. We used the error parameters $\mu_{bad} = 10$ and $\sigma_{bad} = 3.2$. For all other links where both transmitting and receiving sensors are good, we omitted the error term; that is, $\mu_{good} = 0$ and $\sigma_{good} = 0$. The purpose of the simulation is to assess the effectiveness of our spectral clustering method in identifying these bad sensors. In the simulation, there are 50 bad sensors.

Let B be the ground truth number of bad sensors input to the simulator, T the total number of sensors deemed as bad sensors by our clustering method, and B^* the number of sensors that the clustering method correctly identified as bad. Hence, $B^* \leq B$ and $B^* \leq T$. The higher the detection rate B^*/B is, the better the method is. Note that $T - B^*$ is the number of false positives. We are interested in raising the *detection rate* B^*/B, without raising the false positive rate $(T - B*)/T$ significantly. But usually there is a tradeoff between the two factors.

We study performance under varying numbers k of leading eigenvectors used in the spectral clustering method. In the simulator we assume that the minimum cluster size S_{min} from Section IV-B is 2. Meanwhile, we explore a range of values for the parameter τ, or the fraction of sensors furthest from a cluster center which we will declare bad. Increasing τ leads to identifying more bad nodes, but usually at the expense of more false positives.

The following observation supported by simulation results in the rest of this section (see Figure 13) and in

Section V, is worth mentioning. Before a certain threshold (e.g., 6) on the number k of leading eigenvectors used in spectral clustering is reached, when we increase k, we can catch more bad sensors while keeping the τ value constant. This means that by increasing k, we can increase the bad sensor detection rate, without increasing the false positive rate. This result follows from the fact, noted earlier in Section III, that use of more eigenvectors (before reaching some threshold on k) will give finer-grained measurement clusters, which in turn will allow us to identify bad sensors which have smaller errors in their computed locations.

Figures 11 and 12 show results on two performance metrics: the bad sensor detection rate and false positive rate, respectively, with respect to the τ threshold. Additionally, Figure 13 shows the effect of the number k of leading eigenvectors used on detection and false positive rates. We note the following from the figures:

1) When k increases, the detection performance improves for each value of τ. When k is at its highest value of 14, the best performance is achieved (see Figure 11).
2) For certain values of k, the false positive rates as a function of τ exhibit a minimum; in Figure 12 these minimums are reached around $\tau = 10\%$ to 15%.
3) For a fixed τ, there are values of k which give minimal false positive rates; for example, for $\tau = 10\%$, the lowest false positive rate in Figure 13 is reached for $k = 6$.
4) Increasing k past that giving the minimum false positive rate only provides a minor increase in detection rate. Given that the false positive rate increases rapidly, the best value for k is close to the number of major sensor clusters, or, equivalently, the number of leading eigenvalues significantly larger than the rest.

V. Application 2: 2D Localization with 2D Sensor Indexing

After having described in the previous section the method of detecting bad sensors using 1D indexing, we now have enough context and terminology to describe Application 2, where we extend the method to 2D indexing and show its utility in improving the performance of a 2D localization application.

We again assume that there are N ranging sensors which take measurements of received RF signal strength at each other, but are now placed in a 2D space. As in Application 1, we will form the measurement matrix and use spectral clustering methods to identify bad sensors whose measurements are inconsistent with those of peer sensors.

Since received signal strength reflects geographical locations of sensors, it would be natural to define the 2D index of a sensor to be its geographical 2D location. But we cannot assume that we know sensors' 2D locations, for otherwise we would not need to do the localization application in the first place. Also, if we know the 2D locations of the sensors, we could have simply used this location information to check if signal strength measurements are compatible with transmission distances and, accordingly, validated or invalidated the associated sensors.

Instead, like the 1D indexing in Section IV, we will define the 2D index of a sensor to be its *computed* 2D location, obtained from a localization method (e.g., MDSR) using the given measurement matrix as input. Consider a bad sensor, X, which belongs to a measurement

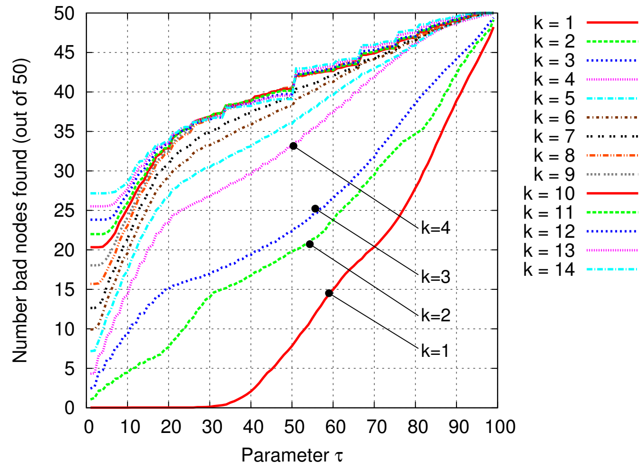

Fig. 11: Number of bad sensors, out of 50, identified for increasing values of parameter τ and a range of values for the number of leading eigenvectors k used for spectral clustering.

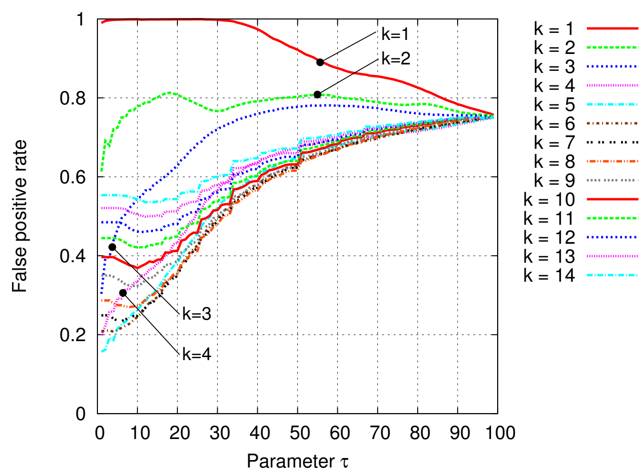

Fig. 12: False positive rate of the spectral clustering method under increasing threshold τ, for a range of number k of leading eigenvectors used for spectral clustering.

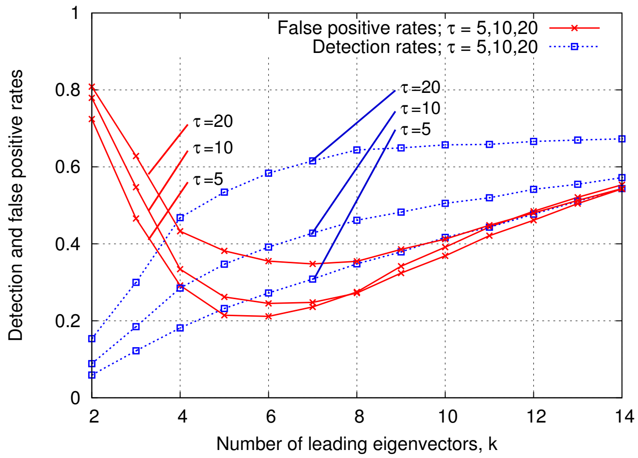

Fig. 13: Detection rate and false positive rate plotted against number k of leading eigenvectors used for spectral clustering for three representative values of τ.

cluster of sensors. Suppose that most of the other sensors in the measurement cluster are good. Being in the same measurement cluster, these good sensors will have similar measurements with respect to sensors outside the cluster. Thus we can expect that a localization method will place these good sensors close to each other, resulting in a peer cluster in the index space. But for X, being a bad sensor, the computed location tends to be pushed away from those of the good nodes. Thus in the index space, the bad sensor X is likely to be on the edge of the measurement cluster containing X.

By this reasoning, our method of identifying bad sensors for the 2D localization application works as follows:

1) Based on the measurement matrix, localize sensors using some standard localization method.
2) Each sensor is given a 2D index which is its computed location obtained from Step 1.
3) Based on the measurement matrix, compute measurement clusters of sensors using a spectral clustering method as described in Section IV-A.
4) Identify bad sensors by examining how measurement clusters are distributed in the 2D index space of sensors. More specifically, check conditions C1 and C2 of Section II; for this purpose, we again employ the techniques from Section IV-B using thresholds S_{min} and τ to catch bad sensors.

It is important to note that for the purpose of providing sensor indexing, the localization method for step 1 is only required to preserve measurement clusters for good nodes. Other metrics, such as absolute accuracy of computed locations are not important. Thus computed measurement clusters could be displaced, rotated or flexed from the their ground-truth locations. Well-known localization methods, such as MDS [1], normally perform well under this relaxed criterion, and thus can be successful in providing the 2D sensor indexing.

A. *Simulation Results on Using Measurement Clusters of Sensors in Identifying Bad Sensors*

We performed a simulation study on the effectiveness of using measurement clusters of sensors in identifying bad nodes. We created four groups of 100 nodes in the corners of a square region. Within each group the 100 sensors had locations chosen uniformly at random, resulting in node locations such as those depicted in Figure 14.

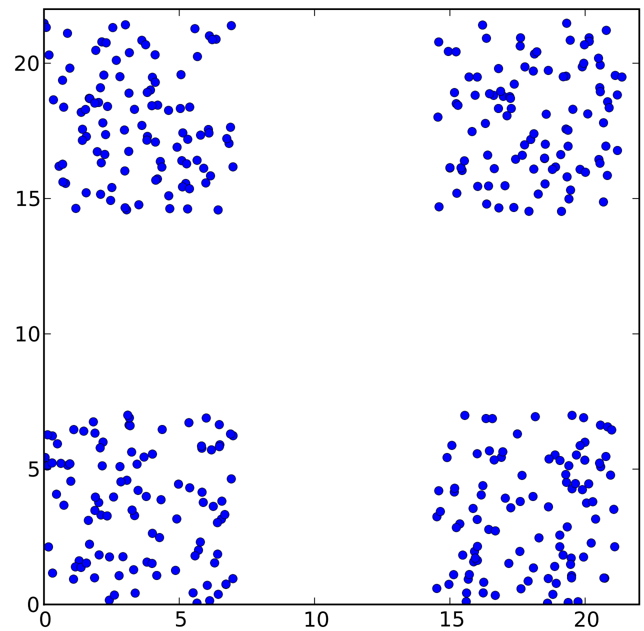

Fig. 14: 100 nodes each are placed uniformly at random into 7×7 regions in each corner.

From these ground truth locations we computed synthetic RF ranging measurements using an empirical two-ray propagation model trained on real-world data [2]. Then, we chose a random sample of $B = 50$ sensors to be bad nodes. The error behavior of bad nodes was identical to that in Section IV, that is, measurements in both directions involving a bad node would be attenuated by a random variable $\alpha \sim \mathcal{N}(10, 3.2)$. Such errors represent some persistent effect such as the attenuation due to lossy cables or antenna connectors, shadowing, etc. The ground truth measurements combined with bad node errors constituted the measurement matrix $\mathbf{A}$.

We performed the 4 steps of identifying bad nodes listed earlier in this section as follows.

1) Based on the measurement matrix $\mathbf{A}$, we computed relative sensor locations using the MDSR [1]

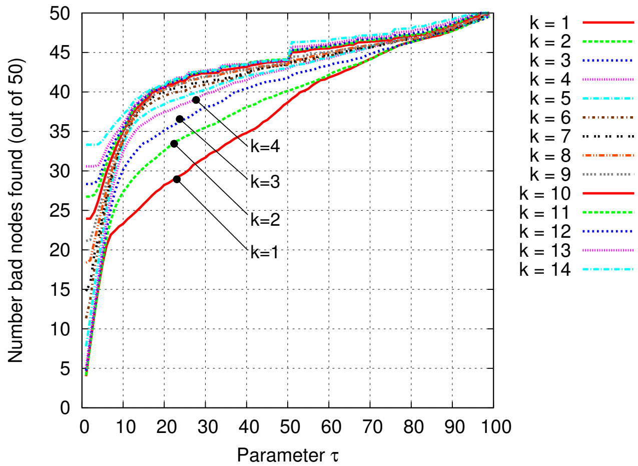

Fig. 15: Number of bad sensors, out of 50, identified for increasing values of parameter τ and a range of values for the number k of leading eigenvectors used for spectral clustering.

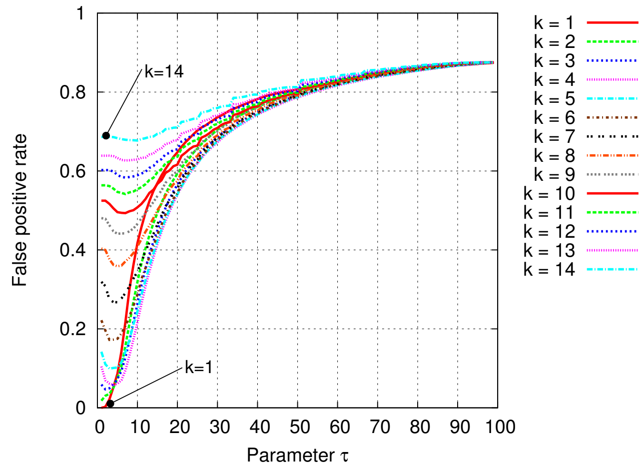

Fig. 16: False positive rate of the spectral clustering method under increasing threshold τ for a range of number k of leading eigenvectors used for spectral clustering.

method. MDSR operates in two steps; first, a classic multi-dimensional scaling (MDS) step [6] computes an approximate solution from the measurement matrix $\mathbf{A}$. This step requires no initial solution. The outcome of MDS then serves as an initial solution for the second step, least squares optimization (LSQ)[1], which produces a more accurate final result. For the purpose of indexing, we only performed the first, MDS step.

2) The computed location of each sensor was used as its 2D index.
3) Using sign-based spectral clustering on matrix $\mathbf{A}^{\mathrm{T}}\mathbf{A}$, we computed measurement clusters of sensors. We varied the number of eigenvectors k used in this process between 1 and 14.
4) We identified bad sensors as those sensors which are either (1) in a measurement cluster of size 1, or (2) within each measurement cluster, in the subset of τ% of sensors furthest from the cluster centroid. We varied the parameter τ% between 0 and 100%.

We measured the performance of each 4-step run using similar metrics as in Section IV-C, that is, the detection and false positive rates. We present performance results in these metrics in Figures 15 and 16 with respect to the τ parameter. Additionally, in Figure 17 we plot results on the same metrics with respect to k.

The most notable property of these results are the large increases in the number of bad sensors detected as τ increases in the region $\tau \leq 20$, especially for the first few smallest values of k (see Figure 15). This indicates that the density of bad nodes is indeed larger on the edge (in the index space) of the measurement clusters. Furthermore, just like in the 1D scenario of Section IV, Figure 17 shows that, at first, increasing k *decreases* the false positive rate; the reason is that the clustering with too few eigenvectors generates too few clusters, and consequently fails to expose some cluster border areas with high densities of bad nodes. Finally, increasing the k parameter starts to increase the false positive rate once the bad node-rich border areas have been exhausted. The minimum false-positive rate is

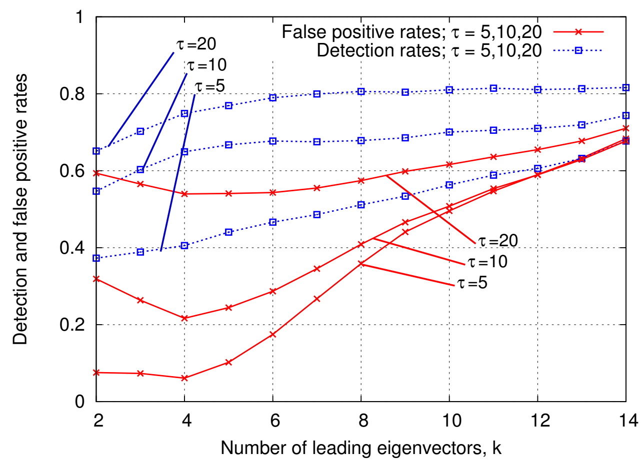

Fig. 17: Detection rate and false positive rate plotted against number k of leading eigenvectors used for spectral clustering for three representative values of τ.

[1] In fact, MDSR performs a weighted least-squares optimization, but for the results reported in Table I, we chose not to use it since it performed worse than unweighted LSQ on our input.

again achieved near the number of significant eigenvectors induced by the clustering structure of the sensor placement.

The bad node classification can be useful to applications; for example, a localization algorithm might weight down those measurements associated with bad nodes to avoid excessive distortion of the results. We tested this process by modifying the LSQ step of MDSR to use weighted least squares optimization, where any distance constraints involving a bad node are weighted down by some factor $\beta < 1$; we used $\beta = 1/100$ in our simulations. We evaluated the performance of unmodified MDSR and the weighted MDSR (denoted WMDSR for short) at the known good nodes only; we report the outcomes in Table I.

#Bad Nodes	5	10	15	20	25
% Improvement WMDSR over MDSR	15.3	10.6	15.6	6.8	5.6

TABLE I: The percent decrease in localization error of MDSR of known good nodes achieved by WMDSR, where the nodes identified as bad by sensor clustering are weighted down. The values are averages over 10 runs. Results are reported for the number of bad nodes increasing from 5 to 25.

VI. Conclusion and Future Work

We have presented a spectral clustering approach to validating sensors via their peers. The work is motivated by the fact that it is often difficult to use instruments to validate equipment in the field and, as a result, peer-based validations can be important in these situations. We have modeled the problem as a clustering problem in the sense that sensors in the same environment will be clustered together and will behave similarly as far as the reference targets are concerned, while a bad sensor must be on the edge (in the index space) of the same cluster, or in a different measurement cluster. Thus, we can identify bad sensors through the clustering structure.

A key result of this paper is on the use of leading eigenvectors in forming measurement clusters of sensors. Although spectral-based clustering is well known in the literature [7], our use in validating sensors in the field appears to be novel. More specifically, to the best of our knowledge, our formulation of sensor indexing and our method of identifying bad sensors using conditions C1 and C2 of Section II, are new. Our application of sign-based spectral clustering to finding measurement clusters of sensors also appears to be new. As shown in our simulation experiments involving 1D and 2D indexing, the method can identify bad sensors with high accuracy and low false positive rate. Furthermore, in the 2D indexing example, we have shown that the performance of the corresponding sensor-based application improved when nodes invalidated by our method are weighted down.

The sensor validation methodology proposed in this paper is quite general, and opens up a number of avenues of future work. For instance, there is the question of which specific sensor properties to use for indexing. In addition to the straightforward ones such as those we used, there might be additional ones based on other sensor modalities, or certain transforms of the sensing data such as the Hough transform [8]. Secondly, there could be other potential clustering methods in addition to the spectral clustering we used in this paper. For this purpose we plan to evaluate methods such as k-means clustering, or spectral clustering using k-means to cluster eigenvectors. Finally, there are potentially many other sensor-based applications where the proposed sensor validation approach could be effective; we are exploring candidate applications such as cognitive radio spectrum allocation.

Acknowledgements

This material is based on research sponsored by Air Force Research Laboratory under agreement numbers FA8750-08-1-0220, FA8750-09-2-0180 and FA8750-08-1-0191. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of Air Force Research Laboratory or the U.S. Government.

References

[1] Y. Shang, W. Ruml, Y. Zhang, and M. Fromherz, "Localization from connectivity in sensor networks," *Parallel and Distributed Systems, IEEE Transactions on*, vol. 15, no. 11, pp. 961–974, 2004.

[2] H. T. Kung, C. K. Lin, T. H. Lin, and D. Vlah, "Localization with Snap-Inducing Shaped Residuals (SISR): Coping with Errors in Measurement," in *The 15th Annual International Conference on Mobile Computing and Networking*, Beijing, China, Sep. 2009.

[3] H. T. Kung and D. Vlah, "A Spectral Clustering Approach to Validating Sensors via Their Peers in Distributed Sensor Networks," in *Second International Workshop on Sensor Networks (SN2009)*, San Francisco, CA, Aug. 2009.

[4] H. T. Kung and B. W. Suter, "A hub matrix theory and applications to wireless communications," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, no. 1, pp. 60–60, 2007.

[5] A. Berman and R. J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*. New York, NY: Academic Press, 1979.

[6] T. F. Cox and M. A. A. Cox, *Multidimensional Scaling, 2nd Ed.* New York, NY: Chapman & Hall/CRC, 2001.

[7] A. Y. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *Advances in Neural Information Processing Systems 14*. MIT Press, 2001, pp. 849–856.

[8] R. O. Duda and P. E. Hart, "Use of the Hough Transformation to Detect Lines and Curves in Pictures," *Communications of the ACM*, pp. 11–15, Jan. 1972.