Distributed SPS Algorithms for Non-Asymptotic Confidence Region Evaluation

Francesca Bassi, Michel Kieffer, Vincenzo Zambianchi, Davide Dardari,
Gianni Pasolini

Distributed SPS Algorithms
for Non-Asymptotic Confidence Region Evaluation

Vincenzo Zambianchi, Michel Kieffer, Francesca Bassi, Gianni Pasolini, and Davide Dardari

Abstract—In this paper, the distributed computation of confidence regions for parameter estimation is considered. Some information diffusion strategies are proposed and compared in terms of the required number of data exchanges to get the corresponding region. The effects of algorithms truncation is also addressed. As support for the theoretical part, numerical results are presented.

I. INTRODUCTION

A wireless sensor network (WSN) is composed of power-constrained sensing devices, cooperating in processing local measurements to complete a common task [1]. The task can be simple environmental monitoring (*e.g.,* temperature, pollution levels) or the evaluation of a complex function of the measurements (*e.g.* localization of a source of signal [2]). The reference scenario considered here is the estimation of a parameter vector $\mathbf{p}^*$ from the set of local and noisy measurements at the nodes. To provide coherent estimates at all nodes, robustness against equipment failure, and scalability versus network size, a distributed protocol working in absence of a central processing unit is desirable.

In [3], [4] this is accomplished via recursive weighted least square estimation, where a consensus-based algorithm allows to incorporate information from neighbouring nodes in the local estimate. A similar approach is taken for the Bayesian framework in [5], where consensus-based distributed Kalman filtering is proposed. The algorithms in [3]–[5] converge asymptotically in time, but it is not possible to predict the quality of the local estimates in the non-asymptotic scenario without statistical knowledge of noise distribution.

In some applications (*e.g.,* in source localization) the characterization of the confidence region associated to the estimate can be more valuable than the evaluation of the point estimate itself. In this work we thus address the problem of the evaluation of *non-asymptotic* confidence regions for the estimation of $\mathbf{p}^*$. Non-asymptotic indicates that the estimate is evaluated from a finite amount of measurements (unlike classical Cramer-Rao bound-like asymptotic confidence regions).

In the presence of a central processing unit the characterization of the non-asymptotic confidence regions can be accomplished using the results in [6]–[9]. Specifically, the methods in [6], [7] allow to derive a lower bound on the probability that $\mathbf{p}^*$ falls within the confidence region, whereas

the *exact* probability can be obtained using the sign perturbed sums (SPS) algorithm in [8]. In [9] efficient computation of the confidence regions is obtained using interval analysis techniques. These methods do not require precise statistical knowledge of the noise, and work under very mild assumptions on its distribution.

The aim of this work is to provide fully distributed solutions for the in-node computation of the non asymptotic confidence regions, based on the SPS algorithm [8]. The SPS method is easily amenable to a distributed form, which converges in time to the centralized algorithm. We consider three distributed approaches (data flooding and parallel in-node processing, distributed processing via average consensus, and mixed flooding+consensus) able to provide exact confidence regions at each node, even for early truncation of the communication protocol. The performance is evaluated, for a target confidence level, in terms of the size of the confidence regions expressed as a function of the amount of data to be exchanged, and hence the energetic cost. This allows to compare the energetic efficiency of the proposed solutions.

II. THE CENTRALIZED SPS ALGORITHM

Let us recall, first of all, the centralized SPS algorithm [8] for the computation of non-asymptotic confidence regions. Consider a network of N nodes taking different noisy observations $\{y_i\}$, for $i = 1, 2, \ldots, N$, of a parameter $\mathbf{p}^*$ belonging to the parameter space $\mathbb{P}$, having dimension n_p. The following local linear scalar measurement model is considered for the ith node:

$$y_i = \boldsymbol{\psi}_i^T \mathbf{p}^* + w_i \,, \tag{1}$$

where $\boldsymbol{\psi}_i$ is the regressor vector, supposed known only at the corresponding node i, and w_i is the measurement noise. The only assumption we make on the measurement noise is that it is independent from node to node and its distribution is symmetric with respect to zero, this setting a quite general domain of applications.

The SPS algorithm returns the exact confidence region around the least squares estimate $\hat{\mathbf{p}}$ of $\mathbf{p}^*$, obtained as the solution of the normal equations $\sum_{i=1}^{N} \boldsymbol{\psi}_i \left(y_i - \boldsymbol{\psi}_i^T \mathbf{p} \right) = \mathbf{0}$. Specifically, define the unperturbed sum

$$\mathbf{s}_0(\mathbf{p}) = \sum_{i=1}^{N} \boldsymbol{\psi}_i \left(y_i - \boldsymbol{\psi}_i^T \mathbf{p} \right) \tag{2}$$

and the $m-1$ sign perturbed sums

$$\mathbf{s}_j(\mathbf{p}) = \sum_{i=1}^{N} \alpha_{j,i} \boldsymbol{\psi}_i \left(y_i - \boldsymbol{\psi}_i^T \mathbf{p} \right), \qquad (3)$$

where $j = 1, \ldots, m-1$ and $\alpha_{j,i} \in \{\pm 1\}$ are independent, identically distributed and equiprobable random signs, drawn at each node i. Given the squared norm of all previous quantities

$$z_j(\mathbf{p}) = ||\mathbf{s}_j(\mathbf{p})||_2^2, \qquad j = 0, \ldots, m-1, \qquad (4)$$

the confidence region $\boldsymbol{\Sigma}_q$ for the true parameter value $\mathbf{p}^*$ is defined as

$$\boldsymbol{\Sigma}_q = \{\mathbf{p} \in \mathbb{P} | z_0(\mathbf{p}) \text{ is not among the } q \text{ largest } z_j(\mathbf{p})\}$$
$$= \left\{ \mathbf{p} \in \mathbb{P} \;\middle|\; \sum_{j=1}^{m-1} \tau_j(\mathbf{p}) \geq q \right\}, \qquad (5)$$

where $\tau_j(\mathbf{p}) = 1$ if $z_j(\mathbf{p}) - z_0(\mathbf{p}) > 0$ and 0 otherwise. The rationale is that, for the true parameter value, the unperturbed sum is closely related to the measurement noise and it is therefore small, while for parameter values that are *far* from the true one, the perturbed sums are larger than the unperturbed one. In [8] it was proven that the confidence level is

$$\text{Prob}(\mathbf{p}^* \in \boldsymbol{\Sigma}_q) = 1 - \frac{q}{m}. \qquad (6)$$

It is worth noting that this algorithm requires a central entity with full knowledge of regressors $\boldsymbol{\psi}_i$, measurements y_i, and coefficients $\alpha_{j,i}$. In the following section we will investigate three different solutions to overcome this limitation and let each node capable of computing (5) locally.

III. INFORMATION DIFFUSION STRATEGIES

This section presents three different solutions for the distributed evaluation of confidence regions. In all cases no central processing unit is assumed, hence confidence regions are locally evaluated at each node.

The first solution here investigated is based on the flooding of all required quantities across the entire network and the subsequent in-node computation of the SPS algorithm. This will not only ensure that the computed confidence region is the same at all nodes, but also that it is exactly the same as in the centralized version of the algorithm.

In the second solution, the information needed for the in-node computation of the SPS algorithm are, instead, exchanged according to an average consensus algorithm [10].

The third solution, whose details will be provided in the following, is a combination of flooding and consensus.

Let us observe, finally, that for energy saving reasons, truncation of the information diffusion algorithms is often desirable. The effects of truncations in the considered algorithms is also investigated in subsection III-D. This constitutes, to the author knowledge, a relevant advance with respect to the current literature.

A. Information Diffusion Through Flooding

In a flooding algorithm, the generic node i initially broadcasts its own state $\mathbf{x}_i^{(0)} = \left[\boldsymbol{\psi}_i^T, y_i \right]$ at time 0. At the successive step, it collects the states from neighbors, then forwards a new data packet containing the aggregated state $\mathbf{x}_i^{(1)} = \left[\left(\boldsymbol{\psi}_i^T, y_i \right), \left\{ \left(\boldsymbol{\psi}_j^T, y_j \right) \right\}_{j \in \mathcal{N}_i} \right]$, where $\left\{ \left(\boldsymbol{\psi}_j^T, y_j \right) \right\}_{j \in \mathcal{N}_i}$ denotes the set of regressors and measurements pairs collected from nodes j in the neighborhood $\mathcal{N}_i$ of node i. To avoid network collapse due to cycles, duplicated states are discarded. This process is repeated until each node in the network has collected the state from all nodes. Afterwards, each node is able to compute the perturbed and unperturbed sums in (2) and (3) for any $\mathbf{p}$, and hence derive the confidence region. No transmission of $\alpha_{j,i}$ is necessary, provided that all nodes agree on seeds for their random generators. Note that each node has to transmit a packet containing $D_{\text{flood}}^{(0)} = n_p + 1$ values in the first iteration and $D_{\text{flood}}^{(\text{last})} = N(n_p + 1)$ values in the last iteration.

B. Information Diffusion Through Consensus

Looking at (2) and (3) it is evident that the computation of $\mathbf{s}_0(\mathbf{p})$ and $\mathbf{s}_j(\mathbf{p})$ does not necessarily require the knowledge of each single quantity, but rather of aggregated values. This suggests the adoption of consensus strategies that are well suited for distributed computation of sums or averages.

A consensus scheme may be viewed as the following discrete time evolving system [10]–[12]

$$\mathbf{x}^{(k+1)} = \mathbf{W} \mathbf{x}^{(k)} \qquad (7)$$

where $\mathbf{x}^{(k)} = \left[\mathbf{x}_1^{(k)}, \mathbf{x}_2^{(k)}, \ldots, \mathbf{x}_N^{(k)} \right]^T$ is the global state at time k, whose ith entry is the local state at node i. $\mathbf{W}$ is the system dynamic matrix that depends on the consensus algorithm as well as on network topology. Proper structures of $\mathbf{W}$ asymptotically lead to a global state whose entries are all equal to the average of the initial quantities $\mathbf{x}^{(0)}$ [11], [12]. In particular, the convergence of the consensus algorithm is assured by the following three necessary and sufficient conditions:

$$\mathbf{1}^T \mathbf{W} = \mathbf{1}^T, \quad \mathbf{W}\mathbf{1} = \mathbf{1}, \quad \rho(\mathbf{W} - \mathbf{1}\mathbf{1}^T/N) < 1 \quad (8)$$

where ρ denotes the spectral radius (see [11]). We choose the Metropolis matrix, which allows a faster convergence than others. Each entry of the matrix is

$$w_{i,j} = \begin{cases} \frac{1}{1+\max\{d_i, d_j\}} & \text{if } (i,j) \in \mathcal{E}, \\ 1 - \sum_{(i,q) \in \mathcal{E}} w_{i,q} & \text{if } i = j, \\ 0 & \text{otherwise}, \end{cases} \qquad (9)$$

where $d_i = |\mathcal{N}_i|$ denotes the degree of node i (i.e., the number of neighbors) and $\mathcal{E}$ is the set of edges in the network topology. According to (7), each node performs its local state update as follows

$$\mathbf{x}_i^{(k+1)} = \sum_{j=1}^{N} w_{i,j} \mathbf{x}_j^{(k)}. \qquad (10)$$

To the purpose of applying the described consensus algorithm to the problem of distributed calculation of (2) and (3), define the following averages that do not depend on $\mathbf{p}$

$$\mathbf{b}_0 = \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{\psi}_i y_i \qquad \mathbf{b}_j = \frac{1}{N} \sum_{i=1}^{N} \alpha_{j,i} \boldsymbol{\psi}_i y_i \qquad (11)$$

$$\mathbf{A}_0 = \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{\psi}_i \boldsymbol{\psi}_i^T \qquad \mathbf{A}_j = \frac{1}{N} \sum_{i=1}^{N} \alpha_{j,i} \boldsymbol{\psi}_i \boldsymbol{\psi}_i^T \qquad (12)$$

for $j = 1, 2, \ldots, m - 1$.

The consensus algorithm is launched on all $\mathbf{A}_j$'s and $\mathbf{b}_j$'s (including $\mathbf{b}_0$ and $\mathbf{A}_0$). At step $k = 0$, the local state at node i is given by $\mathbf{x}_i^{(0)} = \left[(\boldsymbol{\psi}_i y_i)^T, \{\boldsymbol{\psi}_i \boldsymbol{\psi}_i^T\}, \{(\alpha_{j,i} \boldsymbol{\psi}_i y_i)^T\}_j, \{\alpha_{j,i} \boldsymbol{\psi}_i \boldsymbol{\psi}_i^T\}_j \right]$, with $j = 1, 2, \ldots, m - 1$, that is, the single addends in (11) and (12). At each successive step the ith node updates its own state according to (10). Once a consensus on $\mathbf{A}_j$ and $\mathbf{b}_j$ is reached, each node is able to locally evaluate (2) and (3) for any value $\mathbf{p}$ in the parameter search space. We want to underline that no particular value of the parameter $\mathbf{p}$ has to be transmitted and that getting averages, instead of the true sums, does not affect the SPS algorithm since the comparison of rescaled norms or norms gives the same ordering in (5). Therefore the algorithm works also without the knowledge of N. Note that the state is composed of $D_{\mathrm{cons}} = m(3n_p + n_p^2)/2$ values, where symmetry of $\boldsymbol{\psi}_i \boldsymbol{\psi}_i^T$ is exploited. The state dimension is constant during the entire running of the algorithm but is larger than that initially required by flooding.

C. Mixed Flooding+Consensus Approach

As observed, the pure average consensus algorithm requires an amount of data to be transmitted that is initially strictly larger than that required by flooding. However, in few iterations the amount of data transmitted with flooding exceeds D_{cons} due to data accumulation at nodes. This fact suggested the introduction of a third mixed strategy conceived as follows: flooding, as described in Sec. III-A, works until the amount of data flooded by nodes exceeds D_{cons}. Afterwards, nodes switch to a consensus strategy as the one described in Sec. III-B. The correct initialization for the consensus state $\mathbf{x}_i^{(0)}$, in this case, will be obtained as the average of the already received quantities during the initial flooding. The numerical results will show the benefits of this approach when the maximum amount of data exchanged is limited due to energy constraints.

D. Effects of Information Diffusion Truncation

When energy or traffic constraints are present, as typically happens in WSNs, whatever algorithm is used it has to be truncated after a certain number of steps or when the total amount of data exchanged has reached a maximum tolerable threshold. In this case, the confidence region is derived before collecting the whole set of measurements and regressors, in the flooding approach, or before reaching the consensus on the $\mathbf{A}_j$'s and $\mathbf{b}_j$'s, in the consensus approach. Our aim is to determine the characteristics of the confidence region defined by SPS when truncation is considered.
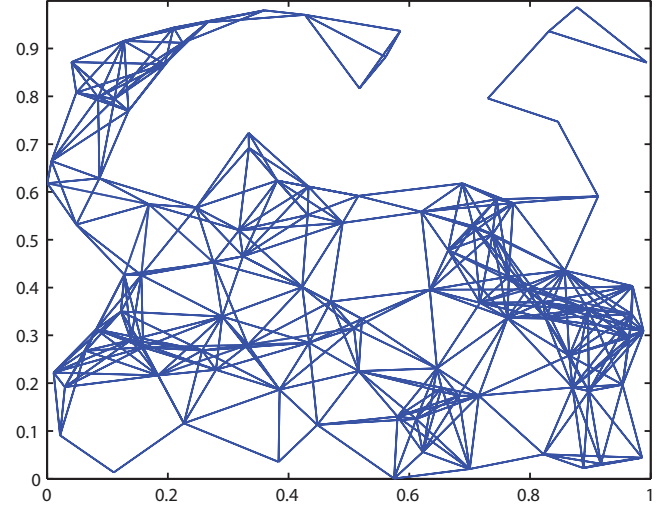

Fig. 1. Network of 100 nodes considered in the simulations.

Truncating the consensus algorithm entails that (2) and (3) are estimated taking into account only the data actually received at node k, that is:

$$\widehat{\mathbf{s}}_{k,0}(\mathbf{p}) = \sum_{i=1}^{N} c_{k,i} \boldsymbol{\psi}_i \left(y_i - \boldsymbol{\psi}_i^T \mathbf{p} \right) \qquad (13)$$

$$\widehat{\mathbf{s}}_{k,j}(\mathbf{p}) = \sum_{i=1}^{N} c_{k,i} \alpha_{j,i} \boldsymbol{\psi}_i \left(y_i - \boldsymbol{\psi}_i^T \mathbf{p} \right), \qquad (14)$$

where $j = 1, \ldots, m - 1$, and $c_{k,i} \in [0, 1]$. Coefficients $c_{k,i}$ do not depend on index j since they depend only on the effectively covered communication links, that are the same for all j's.

As for the truncated flooding algorithm, the estimations of (2) and (3) have the same expressions (13) and (14) that hold for the consensus case, but with $c_{k,i}$ being exactly 1 or 0 if data corresponding to node i have, respectively, been received or not at node k.

In any case, (13) is the normal equation that would be obtained in a centralized context, considering a weighted least-squares estimator, with a diagonal weight matrix $C_k = \mathrm{diag}\,(c_{k,1}, \ldots, c_{k,N})$. Similarly, (14) is the sign perturbed sum that would be obtained when considering weighted least-squares. The confidence region, obtained considering (13) and (14) in (5), is a non-asymptotic confidence region associated to the weighted least-squares estimate. Upon reaching completion of the flooding or convergence of the consensus algorithm, the $c_{k,i}$ are either all equal to one (flooding) or equal to $1/N$ (consensus). As a consequence, in a distributed setup, the same asymptotic result as in the centralized approach can be obtained (asymptotic equivalence). This is no more true when the truncated flooding or consensus algorithms are adopted.

Fortunately, it is possible to assert that any time-truncation of the information-diffusing algorithms still produces confidence regions with the same level of confidence of the asymptotic one. Extended details for this will appear in a
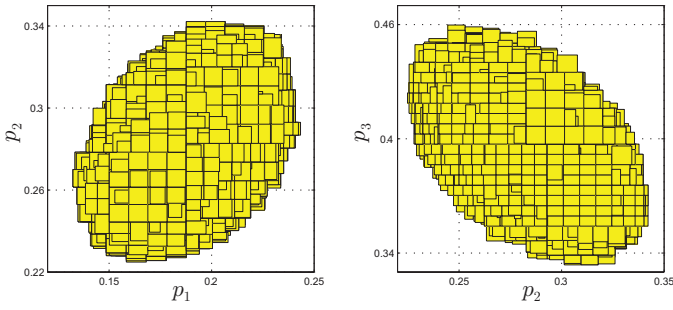
Fig. 2. Projection on the (p_1, p_2) and (p_2, p_3)-planes of the 90% confidence region at node #1 after 4 consensus iterations. Coordinates in the parameter space are denoted as p_1, p_2, p_3.

Fig. 3. Projection on the (p_1, p_2) and (p_2, p_3)-planes of the 90% confidence region at node #1 after 30 consensus iterations.

companion paper; for now it suffices to say that following a similar approach as in [8], the evaluation of (13) and (14) for $\mathbf{p}^*$ gives

$$\hat{\mathbf{s}}_{k,0}(\mathbf{p}^*) = \sum_{i=1}^{N} c_{k,i} \boldsymbol{\psi}_i w_i \qquad (15)$$

and

$$\hat{\mathbf{s}}_{k,j}(\mathbf{p}^*) = \sum_{i=1}^{N} c_{k,i} \alpha_{j,i} \boldsymbol{\psi}_i w_i, \qquad (16)$$

with $j = 1, \ldots, m-1$. The truncation effect is therefore a deterministic rescaling of measurement noises w_i, since it only depends on the number of performed iterations. This rescaling preserves independence as well as symmetry of noise distribution which are the only hypothesis necessary to make (5) and (6) valid. The conclusion is that the confidence region obtained in the presence of truncation is a region that contains the true parameter value with the same exact level of confidence as in the not truncated case, the only difference being the region shape. This will be further investigated in the numerical results and is a consequence of the fact that the asymptotic and the truncated amounts of collected information are not the same.

IV. NUMERICAL RESULTS

In order to show the effects of truncation on the computation of confidence regions and compare the proposed distributed schemes, simulations were performed using Matlab along with the Intlab package [13].

We considered a network of $N = 100$ nodes randomly deployed with uniform distribution in a square area of unit width, as shown in Fig. 1. Each node takes a single scalar measurement according to (1) with the true parameter $\mathbf{p}^* = [p_1^*, p_2^*, p_3^*]^T = [0.2, 0.3, 0.4]^T$. White Gaussian measurement noise was considered, with variance $\sigma^2 = 115$. The regressors are composed of random equiprobable and independent elements with values in $\{-1, 1\}$. No unit of measurement is specified because we do not want to restrict $\mathbf{p}^*$ to any specific domain. The communication link distance is fixed to $d = 0.18$ to ensure network connectivity. Any two nodes under this
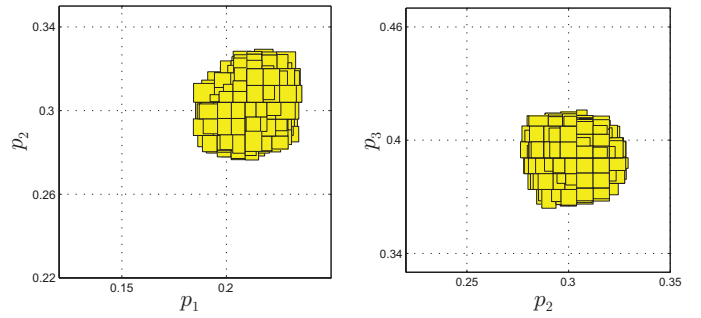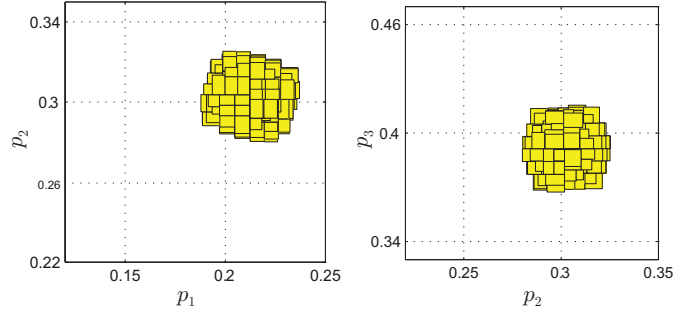

Fig. 4. Projection on the (p_1, p_2) and (p_2, p_3)-planes of the 90% confidence region at node #1 after a complete flooding run.

distance are considered to be communicating in a error-free way.

Figs. 2 and 3 show the 90% confidence regions ($q = 1$, $m = 10$), computed at node #1, after 4 and 30 consensus iterations, respectively. For efficient computation of the regions, the method of contractors described in [9] was employed. The volume reduction, when a higher number of iterations is allowed, is quite evident. One can note that the true parameter is contained in the regions. The amount of data values that node #1 receives before being able to compute the confidence regions is, respectively, 3960 and 29700. The flooding algorithm is investigated in Fig. 4 where the 90% confidence region at node #1 is shown. In this case no truncation was applied and the required amount of data was 49540.

The effect of information diffusion truncation, when a constraint on the maximum amount of received data at each node is present, is investigated in the following results. As indicator of the resulting confidence region extension, we considered the diameter of the single box outer approximation of the confidence region, defined as the longest side of the outer parallelepiped that approximates the confidence region. In Fig. 5 consensus and flooding algorithms are compared, again at node #1. As can be observed, the achieved confidence region with flooding is smaller than that obtainable with consensus when the same amount of received data is considered.

As far as the mixed flooding and consensus approach is concerned, Figs. 6 and 7 show the diameter of the single box outer approximation as a function of the received number
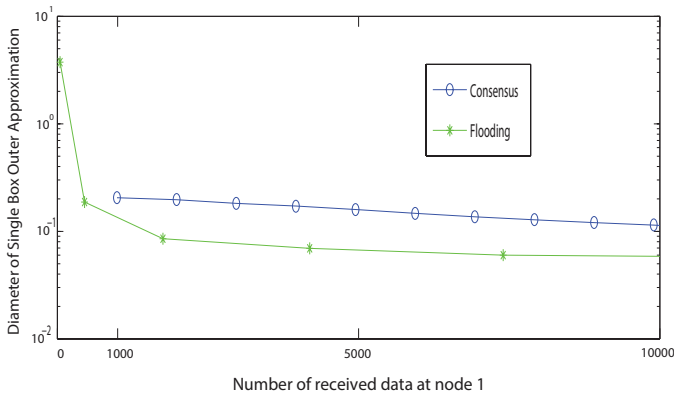
Fig. 5. Diameter of the single box outer approximation as a function of the amount of received data at node #1. Consensus curve is in blue.

Fig. 6. Diameter of the single box outer approximation as a function of the amount of received data at node #1. The mixed approach curve is in blue.

of data at node #1. Fig. 6 was obtained for the topology shown in Fig. 1, while Fig. 7 was obtained with a smaller network of 50 nodes with $d = 0.34$, not shown here due to lack of space. As can be observed, in both cases there is a range of amount of received data wherein the mixed approach slightly outperforms flooding, which is, however, the best solution when no limitation on data exchanges is present (see the asymptotic behavior in Fig. 6 and 7).

V. CONCLUSIONS

In this paper, the problem of efficient distributed evaluation of confidence regions for parameter estimations has been considered. Three different information diffusion strategies, namely flooding, average consensus and a mixed approach, have been compared in terms of confidence region outer approximation shape as a function of the amount of data required for the computation. Effects of truncation for the presented algorithms have been also addressed. Numerical results have been provided both to evaluate the truncation effect and to measure energetic efficiency of the presented algorithms. The main conclusions are that truncation does affect the region shape, but not the confidence level, and that, when the maximum amount of data exchanges is limited, an approach mixing flooding and average consensus is well suited.

Fig. 7. Diameter of the single box outer approximation as a function of the amount of received data at node #1 with $N = 50$ nodes. The mixed approach curve is in blue.

REFERENCES

[1] R. Verdone, D. Dardari, G. Mazzini, and A. Conti, in *Wireless Sensor and Actuator Networks: technologies, analysis and design*. Elsevier Ltd, London, 2008.

[2] J. Matamoros, F. Fabbri, C. Anton-Haro, and D. Dardari, "On the estimation of randomly sampled 2D spatial fields under bandwidth constraints," *IEEE Trans. Wireless Commun.*, vol. 10, no. 12, pp. 4184–4192, Dec. 2011.

[3] I. D. Schizas, G. Mateos, and G. B. Giannakis, "Distributed LMS for consensus-based in-network adaptive processing," *IEEE Trans. on Signal Processing*, vol. 57, no. 6, pp. 2365–2382, June 2009.

[4] G. Mateos, I. D. Schizas, and G. B. Giannakis, "Distributed recursive least-squares for consensus-based in-network adaptive estimation," *IEEE Trans. on Signal Processing*, vol. 57, no. 11, pp. 4583–4588, November 2009.

[5] R. Olfati-Saber, "Kalman-consensus filter: Optimality, stability, and performance," in *IEEE Conf. on Decision and Control Proc.*, Shanghai, China, December 2009, pp. 7036–7042.
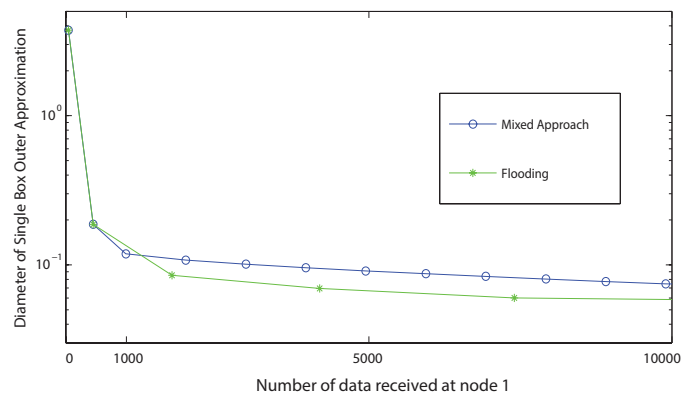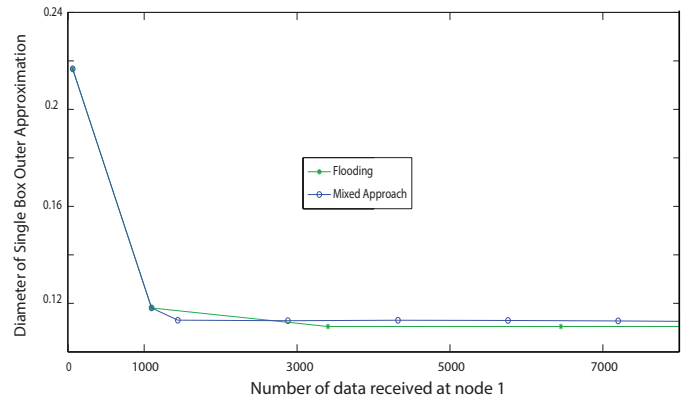
[6] M. C. Campi and E. Weyer, "Guaranteed non-asymptotic confidence regions in system identification," *Automatica*, vol. 41, no. 10, pp. 1751–1764, October 2005.

[7] M. C. Campi, S. Ko, and E. Weyer, "Non-asymptotic confidence regions for model parameters in the presence of unmodelled dynamics," *Automatica*, vol. 45, no. 10, pp. 2175–2186, October 2009.

[8] B. C. Csàji, M. C. Campi, and E. Weyer, "Non-asymptotic confidence regions for the least-squares estimate," in *Proc. IFAC SYSID*, Brussels, Belgium, 2012, pp. 227–232.

[9] M. Kieffer and E. Walter, "Guaranteed characterization of exact non-asymptotic confidence regions as defined by {LSCR} and {SPS}," *Automatica*, vol. 50, February 2014.

[10] L. Xiao, S. Boyd, and S. Lall, "A scheme for robust distributed sensor fusion based on average consensus," in *Proc. of Information Processing in Sensor Networks, IPSN 2005. Fourth International Symposium on*, April 2005, pp. 63–70.

[11] L. Xiao and S. Boyd, "Fast linear iterations for distributed averaging," *Syst. Control Lett.*, vol. 53, no. 1, pp. 65–78, September 2004.

[12] R. Olfati-Saber, J. Fax, and R. Murray, "Consensus and cooperation in networked multi-agent systems," *Proc. IEEE*, vol. 95, no. 1, pp. 215–233, January 2007.

[13] S. Rump, "INTLAB - INTerval LABoratory," in *Developments in Reliable Computing*, T. Csendes, Ed. Dordrecht: Kluwer Academic Publishers, 1999, pp. 77–104. [Online]. Available: http://www.ti3.tu-harburg.de/rump/