# A Statistical Measure for Map Consistency in SLAM

Mladen Mazuran Gian Diego Tipaldi Luciano Spinello Wolfram Burgard Cyrill Stachniss

***Abstract*— Map consistency is an important requirement for applications in which mobile robots need to effectively perform autonomous navigation tasks. While recent SLAM techniques provide an increased robustness even in the context of bad initializations or data association outliers, the question of how to determine whether or not the resulting map is consistent is still an open problem. In this paper, we introduce a novel measure for map consistency. We compute this measure by taking into account the discrepancies in the sensor data and leverage it to address two important problems in SLAM. First, we derive a statistical test for assessing whether a map is consistent or not. Second, we employ it to automatically set the free parameter of dynamic covariance scaling, a robust SLAM back-end. We present an evaluation of our approach on over 50 maps sourced from 16 publicly available datasets and illustrate its capability for the inconsistency detection and the tuning of the parameter of the back-end.**

## I. Introduction

Building consistent maps is an essential capability for autonomous robots, as maps are useful for a large variety of applications that requires the robot to navigate. It turns out that in SLAM there actually is no consistent notion of consistency. For example, in robot mapping, one often uses the term global consistency if the map is in line with the ground truth or at least resembles the topological structure of the environment. Local consistency, in contrast, is used to describe a locally correct alignment of the scans — despite a potential misalignment at the global scale. In this paper, we consider a map to be consistent from a sensor data point of view, i.e., when no noticeable artifacts are present and the topology of the environment is correctly represented.

A large number of mapping approaches have been proposed in the past, see [3, 6] for an overview. A popular approach to SLAM comes from its graph-based formulation. Corresponding methods represent the map as a graph that consists of poses and constrains between the poses. They compute the maximum likelihood map by performing error minimization. Modern techniques apply robust kernels or related techniques during the minimization, with the aim of mitigating data association errors and computing a consistent map [1, 16, 13, 14]. Most robust kernels, such as dynamic covariance scaling (DCS) [1] and switchable constraints (SC) [16], depend on an additional parameter, which, loosely spoken, defines if a constraint is an inlier or an outlier. This parameter, typically is sensitive to the particular environment and needs to be set manually.

All authors are with the University of Freiburg, Institute of Computer Science, 79110 Freiburg, Germany. Cyrill Stachniss is also with the University of Bonn, Inst. of Geodesy and Geoinformation, 53115 Bonn, Germany. This work has partly been supported by the European Commission under FP7-600890-ROVINA, ERC-AG-PE7-267686-LIFENAV and FP7-610917-STAMINA.

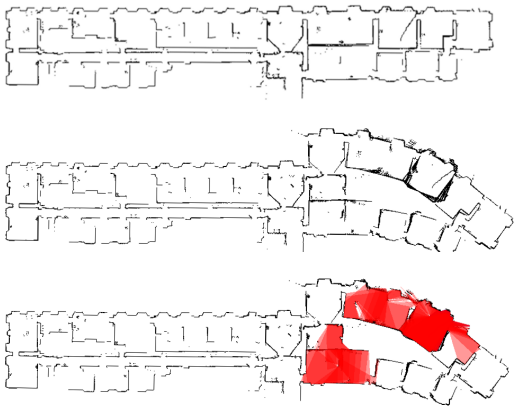


Fig. 1. Inconsistencies detected with our method in a corrupted Belgioioso castle dataset: (top) ground-truth, (middle) inconsistent alignment, (bottom) regions detected as inconsistent by our approach.

The contribution of this paper is a novel method to automatically compute the consistency of static maps by taking into account the discrepancy between sensor readings. We introduce a score function that quantifies the mismatch in the sensor data *after optimization*. It is partially inspired by the lazy data association approach of Hähnel et al. [11] in which a data association tree was computed during mapping. We combine the use of inconsistencies in observations with a cascaded hypothesis test on the entire map to test for global map consistency. Our score function also supports to compare changes in the graph configuration and to evaluate the quality of the result. This allows us to formulate an optimization problem, which sits on top of DCS to tune its parameter automatically.

We formulated both the hypothesis test and the parameter tuning, for 2D laser data. We further implemented and thoroughly evaluated our work on a large set of publicly available datasets. We show that the proposed hypothesis test allows to reliably identify inconsistent maps, without forsaking our ability to detect consistent ones.

## II. Related work

Over the last two decades, a large number of SLAM approaches have been proposed [3, 6]. A recently popular technique to estimate a consistent map is the graph-based SLAM paradigm. A variety of approaches for minimizing the error in the corresponding constraint graph have been proposed, including relaxation methods [7], stochastic gradient descent and its variants [15, 8], smoothing techniques [5, 12], and hierarchical techniques [4, 10]. A comprehensive tutorial

on graph-based SLAM has been published by Grisetti et al [9].

All these techniques assume Gaussian errors in the constraints of the SLAM graph but show decreased performance with an increasing number of data association outliers. Over the last years, researchers invented SLAM back-ends that can deal with a substantial number of outliers [1, 16, 13, 14]. The method of Sünderhauf and Protzel [16] is able to switch off potential outlier constraints and the control of the switching behavior is handled within the optimizer. Related to that, Agarwal et al. [1] proposed dynamic covariance scaling, which is similar to switchable constraints [16] as it also re-weights constraints, but without the need to explicitly compute switching variables. Olson and Agarwal [14] introduced a maximum operator on Gaussian mixtures in the optimization process. It allows for dealing with multi-modal constraints and rejecting outliers, while maintaining computational efficiency. Latif et al. [13] proposed an approach that handles outliers by finding the maximum set of clustered edges that are consistent with each other.

Although these methods for outlier rejection improve the robustness of SLAM back-ends, it is not yet possible to guarantee consistency of the resulting map. The ability to determine if a map is consistent has not received a lot of attention in the literature, with the notable exception of the work by Hähnel et al. [11]. Despite finding the correct data associations, Hähnel et al. determine the log-likelihood of a pairwise scan matching, given their pose estimate. The log-likelihood is computed by superimposing a scan onto a local occupancy grid map build by the other scan.

Our approach can be seen as related to the work of Hähnel et al. as we provide a measure for the likelihood of a pairwise mismatch, but we further extend it in a multiple-overlapping-scans scenario, while abandoning the need for discretization. We furthermore exploit these concepts for performing a cascaded statistical test and apply them to automatically tune the parameter of dynamic covariance scaling [1].

## III. Pairwise Inconsistency Measure

In this section, we propose a measure of compatibility between scans. Our approach makes the assumption that the environment is static, although we will illustrate experimentally that it is still robust with respect to a certain degree of dynamics. In the following, we will describe the methodology for 2D range scans.

Our goal is to provide a value describing the discrepancy between two scans. To achieve this, we determine how much two laser range scans occlude each other's free space. Let $\mathcal{S}_1$ and $\mathcal{S}_2$ be two laser scans, taken from two different poses and expressed in global coordinates. For each scan, we compute a polygon using its end points and the position of the robot. Each polygon describes the free space that is covered by the scan. The intuition is that $\mathcal{S}_1$ and $\mathcal{S}_2$ are considered to be consistent with each other if none of the end points of $\mathcal{S}_1$ lies inside the polygon of $\mathcal{S}_2$ and vice versa. We call the points of $\mathcal{S}_1$ lying inside $\mathcal{S}_2$ the *inconsistent points* of $\mathcal{S}_1$ w.r.t. $\mathcal{S}_2$.

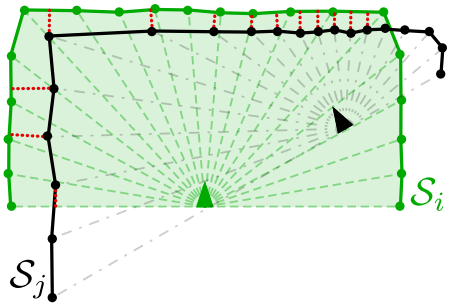



Fig. 2. Example showing inconsistency distances in $\mathcal{S}_i$ w.r.t. $\mathcal{S}_j$. The set of green and black polygonal chains identify the observable boundary of $\mathcal{S}_i$ and $\mathcal{S}_j$, while the shaded green area represents the visibility polygon of $\mathcal{S}_i$. The lengths of the dotted red lines represent the inconsistency distancies in $\mathcal{S}_i$ w.r.t. $\mathcal{S}_j$, for all inconsistent points w.r.t. $\mathcal{S}_j$ falling into the visibility polygon of $\mathcal{S}_i$.

The measure we derive from this intuition computes the sum of the distances of each inconsistent point of the first scan to the boundary of the polygon surrounding the second scan and vice versa. Let $\mathcal{V}_i$ be the set of points defining the visibility polygon of $\mathcal{S}_i$ and $p_i^k$ its $k$-th end point. We define the *inconsistency distance* $d_i(p)$ for a point $p$ as:

$$d_i(p) = \begin{cases} \text{dist}(p, \mathcal{V}_i) & \text{if } p \text{ inside } \mathcal{V}_i \\ 0 & \text{otherwise} \end{cases}, \tag{1}$$

where $\text{dist}(p, \mathcal{V})$ is the Euclidian distance of a point $p$ to the closest point on the polygon boundary $\mathcal{V}$. By summing over all the end points of the scans, we obtain the *pairwise inconsistency measure*

$$M_{ij} = \sum_k d_i(p_j^k) + \sum_k d_j(p_i^k). \tag{2}$$

See Fig. 2 for an example. In the remainder of this paper, we refer to the total number of inconsistent points in $\mathcal{S}_i$ w.r.t. $\mathcal{S}_j$ and in $\mathcal{S}_j$ w.r.t. $\mathcal{S}_i$ as $n_{ij}$. A naïve implementation computes $M_{ij}$ in $\mathcal{O}(K^2)$, where $K$ is the number of end points.

## IV. Statistical Test for Map Consistency

In this section, we formulate a hypothesis test for the global consistency of a map, based on the pairwise measure of compatibility.

### A. Hypothesis Test for Pairs of Scans

Under the assumption of a correct alignment of two scans $\mathcal{S}_i$ and $\mathcal{S}_j$, we expect that on average 50% of the end points in $\mathcal{S}_i$ are inconsistent points of $\mathcal{S}_j$ and vice versa. This is due to the inherent sensor noise in the laser measurements.

If $\mathcal{S}_i$ and $\mathcal{S}_j$ are obtained from the exact same pose and the laser range values are normally distributed, the inconsistency distances are half-normally distributed, with the scale parameter $s^2$ given by twice the range variance of the laser. The half-normal distribution is the distribution of $Y = |X|$, with $X \sim \mathcal{N}(0, s^2)$.

We further assume that the inconsistencies are i.i.d. random variables with finite mean $\mu$ and variance $\sigma^2$. In the case of the half-normal distribution, we have

$$\mu = s\sqrt{\frac{2}{\pi}} \qquad \sigma^2 = s^2\left(1 - \frac{2}{\pi}\right). \tag{3}$$

According to the central limit theorem, $M_{ij}$ follows

$$\frac{M_{ij} - \mu\, n_{ij}}{\sigma\sqrt{n_{ij}}} \underset{n_{ij}\to\infty}{\sim} \mathcal{N}(0,1). \tag{4}$$

The former can be generalized to situations in which the random variables follow different distributions, as long as they are independent and they satisfy Lindeberg's condition [2]. In this case, the limit of the mean is still normally distributed. We found that the approximation is very accurate even for as little as 20 half-normal samples.

As a result of that, we can conduct a hypothesis test if $n_{ij}$ is sufficiently large and $\mu$ and $\sigma^2$ are known. The hypothesis test allows us to determine whether $M_{ij} \sim \mathcal{N}(\mu\, n_{ij},\, \sigma^2 n_{ij})$. It requires a one-sided test, as only large $M_{ij}$ values are relevant; a low $M_{ij}$ value only implies that larger deviations from the perfect map assumption are accepted.

Based on $N$ laser scans, we can build an $N \times N$ symmetric matrix $\mathbf{\Psi}$ containing standardized $M_{ij}$ values as follows:

$$\mathbf{\Psi} = \left[\frac{M_{ij} - \mu\, n_{ij}}{\sigma\sqrt{n_{ij}}}\right]_{\substack{1\le i\le N\\ 1\le j\le N}} \tag{5}$$

Here, the sparsity of this matix depends on the amount of overlap between laser scans.

Given $\mathbf{\Psi}$, we can infer whether two scans $\mathcal{S}_i$ and $\mathcal{S}_j$ are consistent with confidence $1-\alpha$ by verifying the inequality

$$\Psi_{ij} \le F^{-1}(1-\alpha), \tag{6}$$

where $F^{-1}(p)$ is the inverse CDF of the normal distribution.

By adopting a bounding box test and assuming that the pairwise measure of compatibility is determined in $\mathcal{O}(K^2)$ complexity, $\mathbf{\Psi}$ can be computed in $\mathcal{O}(N^2 + NRK^2)$, where $R$ denotes the average number of laser scans a scan overlaps with respect to the bounding box and $K$ is the number of laser beams per scan.

### B. One-vs-All Consistency Test for Scans

To asses global consistency, one can conduct the above described test for all pairs of scans and consider a map to be consistent if all tests are successful. The problem, however, is that a single statistical test will produce the wrong result with probability $\alpha$. Thus, testing a single scan that overlaps with $r$ other scans yields a type I error probability of $1-(1-\alpha)^r$. This renders the direct application of the pairwise approach unsuitable.

To overcome this problem, we model the outcome of a pairwise hypothesis test as a Bernoulli-distributed random variable with parameter $\alpha$. Thus, the number of failed tests follows a binomial distribution with parameters $\alpha$ and $r$. Given that, we can compute the maximum number $\hat{\xi}$ of tests that can fail for a confidence level $1-\alpha'$ as

$$\hat{\xi} = \min_{0\le\xi\le r}\left\{\xi \;\middle|\; \sum_{i=\xi+1}^{r}\binom{r}{i}\alpha^i(1-\alpha)^{r-i} \le \alpha'\right\}. \tag{7}$$

Computing $\hat{\xi}$ according to (7) is numerically unstable. Thus, we exploit

$$\begin{aligned}\binom{r}{i}\alpha^r(1-\alpha)^{r-i} = \\ \exp\big(i\log\alpha + (r-i)\log(1-\alpha) + \log\Gamma(r+1) \\ - \log\Gamma(i+1) - \log\Gamma(r-i+1)\big),\end{aligned} \tag{8}$$

where $\Gamma(\cdot)$ is the gamma function.

This allows for computing a cascaded hypothesis test for the consistency of a scan with respect to all scans it overlaps with. We first perform all pairwise hypothesis tests. Then, if the number of failed tests is smaller than $\hat{\xi}$, the overall consistency test is positive.

### C. Estimating Distribution Parameters from Data

In theory, one could compute the mean $\mu$ and variance $\sigma^2$ of the half-normal distribution directly from the variance of the laser range finder. In practice, however, this approach does not take into account the bias of the scan matcher nor artifacts due to the incidence angle of the laser beams. To improve robustness, we estimate them directly from data, assuming that the incremental solution, i.e., without loop closures, is consistent. Moreover, to account for the heteroscedasticity of the test variables $M_{ij}$ due to the scan matcher bias, we adaptively estimate a mean $\mu_i$ and a variance $\sigma_i^2$ for each scan.

We assume here that the incremental scan matcher produces correctly aligned scans locally. For the $i$-th scan, we define $L_i$ to be the set of scan indices in its local neighborhood. In our current implementation, this contains up to 20 scans.

Formally, in order for all $M_{ij}$ to pass the hypothesis test with confidence $1-\alpha$, the following inequalities need to be satisfied:

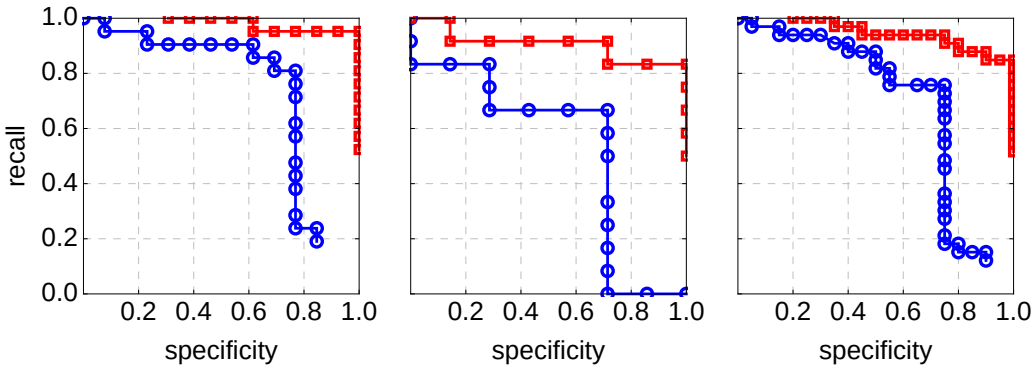
$$\frac{M_{ij} - \mu_i\, n_{ij}}{\sigma_i\sqrt{n_{ij}}} \le F^{-1}(1-\alpha) \quad \forall j \in L_i \quad \forall i \tag{9}$$

Recall that the inconsistency distances are half-normally distributed with mean and variance given by the scale parameter $s_i^2$ according to (3). By substituting (3) into (9) we obtain

$$s_i \ge \max_{j\in L_i} \frac{M_{ij}}{n_{ij}\sqrt{\frac{2}{\pi}} + F^{-1}(1-\alpha)\sqrt{n_{ij}\left(1-\frac{2}{\pi}\right)}}. \tag{10}$$

There is one degree of freedom to choosing $s_i^2$, we thus consider the limiting behavior and replace the inequality with an equality.

During the compatibility test of two scans on the overall map, we choose the largest scale parameter between $s_i^2$ and $s_j^2$ for the pairwise $M_{ij}$ test. This formulation, to some extent, allows to deal with dynamic environments, although the larger the amount of dynamics the worse the ability to detect inconsistencies.




(a) Static-world datasets (b) Datasets with dynamics (c) All 53 datasets

Fig. 3. Pareto frontier for specificity and recall on the evaluated datasets. The method of Hähnel et al. is shown in blue while ours is in red.

## V. Automatic Parameter Tuning for DCS

One current limitation of DCS [1] is the effect of the parameter $\Phi$ on the solution. In this section, we show how to use the proposed inconsistency measure in order to find the $\Phi$ that produces the most consistent map.

The parameter $\Phi$ influences the map optimization process, thus it implicitly affects the inconsistency measure $M_{ij}$. Let $\Xi(\Phi)$ denote the number of pairs of overlapping scans, i.e., $\Xi(\Phi) = |\{(i,j) \mid \mathcal{V}_i \cap \mathcal{V}_j \neq \emptyset\}|$. We define the average inconsistency measure of a DCS solution as

$$M(\Phi) = \frac{1}{\Xi(\Phi)} \sum_{i=1}^{N} \sum_{j=1}^{N} M_{ij}(\Phi), \tag{11}$$

where $M_{ij}(\Phi)$ is the pairwise inconsistency measure computed on the DCS solution with parameter $\Phi$.

The function $M(\Phi)$ allow us to compare the outcome of the map optimization for different values of $\Phi$. Our aim is to find the $\Phi^*$ that minimizes (11). In the ideal, noise-free case $M(\Phi^*) = 0$ and by definition $M(\Phi^*) \geq 0$. Thus, we aim to minimize $M(\Phi)$:

$$\Phi^* = \underset{\Phi \in [0, +\infty)}{\arg\min} \; M(\Phi) \tag{12}$$

Unfortunately, no closed form for $M(\Phi)$ is available and it is only pointwise evaluable. Therefore, we address (12) as a search problem over the $\sqrt{\Phi}$ space ($\Phi$ is an intrinsically quadratic value), by iteratively computing the DCS solution and its $M(\Phi)$ value. We empirically found that a comparably simple search strategy is sufficient. Rather than adopting procedures such as simulated annealing, a standard grid search approach followed by a bisection method for fine tuning is successful to approach $\Phi^*$.

If a $\Phi$ such that the resulting map is consistent does not exist, $M(\Phi)$ may favor a few large $M_{ij}(\Phi)$ values over many small $M_{ij}(\Phi)$ values. Intuitively, it may favor few strong inconsistencies, e.g., intersecting corridors, over many weak ones, e.g., misaligned corridors. While this effect is to some extent balanced by the number of overlapping scans $\Xi(\Phi)$, we do not exclude the possibility of such an instance arising. In those cases, we recommend to perform the statistical hypothesis test on the resulting map, to assess its consistency.

## VI. Experimental Evaluation

This evaluation is designed to illustrate (a) the effectiveness of our statistical consistency test, and (b) the ability to automatically find the $\Phi$ parameter in DCS for computing consistent maps.

### A. Map Consistency

To evaluate the performance of the proposed hypothesis test, we executed it on 16 publicly available, 2D laser datasets[1] including ACES3, Belgioioso castle, Bicocca 25b, Bicocca 26a, MIT CSAIL, Edmonton, FHW, Freiburg 079, Freiburg 101, Intel lab, MIT Killian, Mexico Acapulco, Orebro, Intel Oregon, UW Seattle, Stanford gates, and University of Bremen. We partitioned the 16 datasets into two groups: static and dynamic. For each group, we have generated consistent and inconsistent maps, resulting in 53 configurations.

We compared our method with the one by Hähnel et al. [11] on each configuration. For our method, we consider a map to be inconsistent if at least one cascaded hypothesis test failed. For the method of Hähnel et al. a map is inconsistent if the log-likelihood is below a set threshold. We run the experiments for multiple values of our confidence parameters and the threshold of the method by Hähnel et al. Each parameter configuration results in a point in the specificity-recall space. We compute the optimal Pareto frontiers of the two methods and report them in Fig. 3.

In static environments, the plots show that our method improves both, recall and specificity, compared to the one of Hähnel et al, see Fig. 3(a). Our approach is close to the optimal result point $(1, 1)$ in the specificity-recall space. The results show the advantage of using a cascaded hypothesis test rather than relying on a greedy parameter choice. Note that, the method proposed by Hähnel et al. only obtains high specificity scores at low recall values. Fig. 1 shows, as an example, the inconsistent regions detected by our method on a corrupted version of the Belgioioso castle.

[1] http://www.informatik.uni-freiburg.de/~mazuran/files/tagged-datasets-icra14.tar.gz

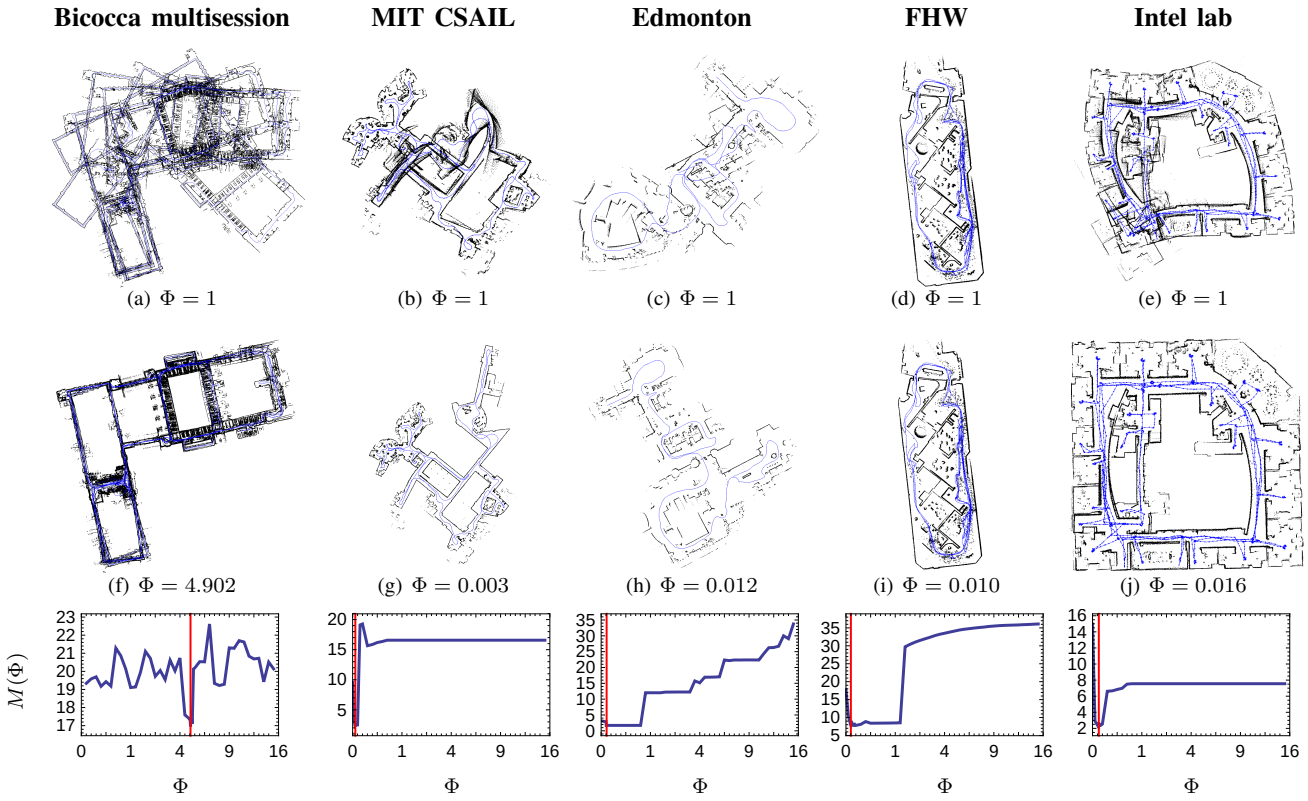



Fig. 4. Φ sensitivity for different datasets: the top row shows the maps for the $\Phi = 1$ baseline, while the middle row shows the maps for the Φ values computed by our algorithm. The last row displays the value of $M(\Phi)$ on the bounded interval $[0, 16]$; the red line denotes the Φ coordinate of the minimum of $M(\Phi)$.

In case of dynamic environments, the difference between the two approaches becomes even more evident (see Fig. 3(b)). The MIT-Killian dataset, for instance, contains a substantial amount of fast dynamics cased by walking people. Our approach is able to correctly identify those scans as being consistent, apart from two instances, which are depicted in Fig. 5. For these two scans, our algorithm detects two significant instances of quasi-static elements, in the form of two objects that were absent when the robot first visited the area, but were present in subsequent visits. By inspecting other dynamic datasets, we found that our method handles fast dynamics robustly, as long as their impact is moderate, i.e., does not cover large portions of the scans.

The robustness w.r.t. dynamics comes from the adaptive estimation of the inconsistency distance distribution parameters presented in Section IV-C. Mean and variance are increased when temporally local scans do not match well. This also allows to mitigate the effect of incorrect laser readings, due to bumps or slight slopes of the terrain, which occur in datasets such as Bicocca 25b.

Fig. 3(c) shows the performance on all the datasets, including static and dynamic environments. The performance trend of both approaches is confirmed. Our approach outperforms the one of Hähnel et al. in all the specificity-recall space.

### B. Parameter Optimization for SLAM

This experiment is designed to show the impact of parameter search for map optimization with DCS. We compare

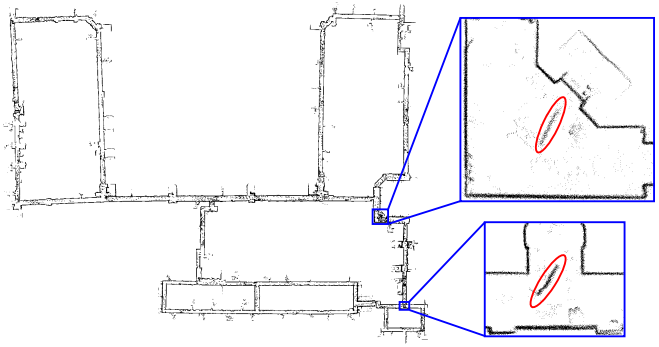


Fig. 5. Quasi-static objects in the MIT-Killian dataset marked in red.

DCS with our parameter search method vs. DCS with the suggested default parameter value of $\Phi = 1$ on five publicly available datasets[2]: Bicocca multisession, MIT CSAIL, Edmonton, FHW, and Intel lab.

The Bicocca multisession dataset is the same one considered by Latif et al. [13], including the initial guess. For all other datasets, we corrupted the initial guess to challenge our parameter tuning approach. In addition to that, we manually added data association outliers to the graphs of MIT CSAIL and Edmonton. For the FHW and Intel lab datasets, we generated false data association edges by running a SLAM

[2] http://www.informatik.uni-freiburg.de/~mazuran/files/dcs-datasets-icra14.tar.gz

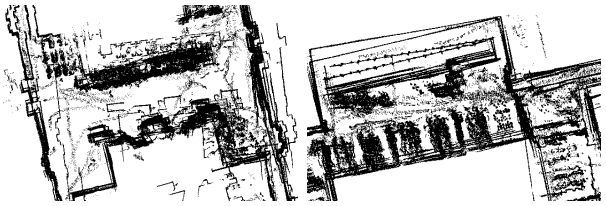


Fig. 6. Example inconsistencies in the optimized Bicocca multisession dataset. The trajectory of the robot is not shown for clarity.

TABLE I
AVG. EXECUTION TIME OF THE STATISTICAL HYPOTHESIS TEST

| **Dataset** | $N$ | $R$ | $K$ | **Runtime [s]** |
|---|---|---|---|---|
| ACES3 | 706 | 32.7 | 180 | 3.6 |
| Belgioso castle | 349 | 29.7 | 361 | 5.3 |
| Bicocca 25b | 1268 | 26.7 | 181 | 4.5 |
| Bicocca 26a | 1475 | 23.7 | 181 | 5.4 |
| MIT CSAIL | 686 | 50.7 | 361 | 13.1 |
| MIT Killian | 5489 | 56.0 | 180 | 42.0 |
| Edmonton | 630 | 35.4 | 180 | 1.3 |
| FHW | 1941 | 211.7 | 180 | 27.3 |
| Freiburg 079 | 833 | 118.4 | 360 | 29.5 |
| Freiburg 101 | 504 | 73.6 | 360 | 6.8 |
| Intel lab | 1311 | 133.7 | 180 | 13.4 |
| Intel Oregon | 485 | 62.4 | 181 | 2.9 |
| Mexico Acapulco | 1547 | 72.7 | 181 | 3.4 |
| Orebro | 237 | 19.5 | 181 | 0.7 |
| UW Seattle | 233 | 22.8 | 361 | 3.5 |
| Stanford gates | 1127 | 63.9 | 181 | 6.5 |
| U. Bremen | 119 | 21.1 | 181 | 0.4 |

front-end [17] with suboptimal parameters.

For such challenging datasets, DCS with $\Phi = 1$ provides a consistent solution only for the FHW dataset. Fig. 4 shows the resulting maps of the datasets considered during the experiment. The top row shows the output of DCS when $\Phi = 1$, while the second row the output of DCS using the automatic parameter tuning procedure. The last row shows the profile of $M(\Phi)$ for different values of $\Phi$. Our approach is able to find appropriate values for $\Phi$ and is producing consistent maps.

Note that the best result for the Bicocca multisession dataset is obtained with a value of $\Phi \approx 5$, which is in line with the work of Agarwal et al. [1]. The resulting map, however, does not pass the consistency test and it is barely usable for autonomous navigation. Fig. 6 illustrates this fact through a magnified view of two portions of the map.

### C. Computation Time

Conducting the statistical test increases the computational resources required to build a map. Table I summarizes the execution time on an Intel i7-3770K processor using 8 parallel threads for different datasets. Our hypothesis test along with the parameter tuning is highly parallelizable, hence can be speeded-up if implemented in CUDA or OpenCL.

## VII. CONCLUSION

In this paper, we addressed the problem of measuring map consistency. We introduced a pairwise measure of inconsistency between scans that calculates the amount of mismatches of globally aligned laser measurements. We furthermore described how the proposed measure enables us to derive a statistical hypothesis test to assess the consistency of a map and to provide a means for automatically tuning parameters of a SLAM back-end.

We implemented and thoroughly tested our approach. We evaluated the proposed hypothesis test on publicly available datasets and showed the effectiveness of our approach in both tasks, namely detecting the consistency of a map and automatically tuning the free parameter in the dynamic covariance scaling approach. The latter allows us to retrieve the best map achievable by DCS without the need for user intervention or manually specifying the free parameter. We believe that the combination of an automatic parameter tuning method with a consistency hypothesis test is an important tool for verifying the success of a mapping process. The source code of the approach is freely available at `www.informatik.uni-freiburg.de/~mazuran/consistency`

## REFERENCES


[1] P. Agarwal, G. Tipaldi, L. Spinello, C. Stachniss, and W. Burgard. Robust map optimization using dynamic covariance scaling. In *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, 2013.

[2] R. Ash and C. Doléans-Dade. *Probability and Measure Theory*. Harcourt/Academic Press, 2000.

[3] T. Bailey and H. Durrant-Whyte. Simultaneous localization and mapping (SLAM): Part II state of the art. *Journ. of Rob. & Aut. Systems*, 2006.

[4] M. Bosse, P. M. Newman, J. J. Leonard, and S. Teller. An ATLAS framework for scalable mapping. In *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, 2003.

[5] F. Dellaert. Factor graphs and GTSAM: A hands-on introduction. Technical report, Georgia Tech, 2012. GT-RIM-CP & R-2012-002.

[6] H. Durrant-Whyte and T. Bailey. Simultaneous localization and mapping (SLAM): Part I the essential algorithms. *Journ. of Rob. & Aut. Systems*, 2006.

[7] U. Frese, P. Larsson, and T. Duckett. A multilevel relaxation algorithm for simultaneous localisation and mapping. *IEEE Transactions on Robotics*, 21(2), 2005.

[8] G. Grisetti, C. Stachniss, and W. Burgard. Non-linear constraint network optimization for efficient map learning. *IEEE Transactions on Intelligent Transportation Systems*, 2009.

[9] G. Grisetti, R. Kümmerle, C. Stachniss, and W. Burgard. A tutorial on graph-based SLAM. *IEEE Transactions on Intelligent Transportation Systems Magazine*, 2:31–43, 2010.

[10] G. Grisetti, R. Kümmerle, C. Stachniss, U. Frese, and C. Hertzberg. Hierarchical optimization on manifolds for online 2D and 3D mapping. In *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, 2010.

[11] D. Hähnel, W. Burgard, B. Wegbreit, and S. Thrun. Towards lazy data association in SLAM. In *Proc. of the Int. Symposium of Robotics Research (ISRR)*, 2003.

[12] M. Kaess, A. Ranganathan, and F. Dellaert. iSAM: Fast incremental smoothing and mapping with efficient data association. In *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, 2007.

[13] Y. Latif, C. Cadena, and J. Neira. Robust loop closing over time. *Proc. of Robotics: Science and Systems (RSS)*, 2012.

[14] E. Olson and P. Agarwal. Inference on networks of mixtures for robust robot mapping. In *Proc. of Robotics: Science and Systems (RSS)*, 2012.

[15] E. Olson, J. Leonard, and S. Teller. Fast iterative optimization of pose graphs with poor initial estimates. In *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, 2006.

[16] N. Sünderhauf and P. Protzel. Switchable constraints for robust pose graph SLAM. In *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2012.

[17] G. D. Tipaldi, L. Spinello, and W. Burgard. Geometrical FLIRT phrases for large scale place recognition in 2D range data. In *Proc. of the IEEE Int. Conf. on Robotics & Automation (ICRA)*, 2013.