
Research article

Fusing photovoltaic data for improved confidence intervals

Ansgar Steland

Institute of Statistics, RWTH Aachen University, Wüllnerstr. 3, D-52056 Aachen, Germany

Correspondence: Email: steland@stochastik.rwth-aachen.de; Tel: +49-241-8094571.

Abstract: Characterizing and testing photovoltaic modules requires carefully made measurements on important variables such as the power output under standard conditions. When additional data is available, which has been collected using a different measurement system and therefore may be of different accuracy, the question arises how one can combine the information present in both data sets. In some cases one even has prior knowledge about the ordering of the variances of the measurement errors, which is not fully taken into account by commonly known estimators. We discuss several statistical estimators to combine the sample means of independent series of measurements, both under the assumption of heterogeneous variances and ordered variances. The critical issue is then to assess the estimator's variance and to construct confidence intervals. We propose and discuss the application of a new jackknife variance estimator devised by [1] to such photovoltaic data, in order to assess the variability of common mean estimation under heterogeneous and ordered variances in a reliable and nonparametric way. When serial correlations are present, which usually affect the marginal variances, it is proposed to construct a thinned data set by downsampling the series in such a way that autocorrelations are removed or dampened. We propose a data adaptive procedure which downsamples a series at irregularly spaced time points in such a way that the autocorrelations are minimized. The procedures are illustrated by applying them to real photovoltaic power output measurements from two different sun light flashers. In addition, focusing on simulations governed by real photovoltaic data, we investigate the accuracy of the jackknife approach and compare it with other approaches. Among those is a variance estimator based on Nair's formula for Gaussian data and, as a parametric alternative, two Bayesian models. We investigate the statistical accuracy of the resulting confidence resp. credible intervals used in practice to assess the uncertainty present in the data.

Keywords: Bayes; credible interval; decorrelation; jackknife; nonparametric estimation; power output; resampling; round robin; variance estimation

1. Introduction

The characterization and testing of photovoltaic (PV) modules in terms of their electrical properties and, especially, their energy yield as measured by the power output is an important but expensive task. There are various factors which influence those electric properties such as the irradiance and its spectrum, the module temperature and, for some module technologies, to which extent the module has been recently exposed to light. For tasks such as certification and testing, power output measurements are taken in a flasher under standard conditions (STC), in order to evaluate the nominal specifications. There are a couple of situations where *two* data sets are available, such that the question arises whether one can fuse them (or the statistics calculated from them), in order to improve the estimation and uncertainty quantification of the true but unknown power output:

- Manufacturers flash each module right at the production line for the purposes of quality control and monitoring, and in order to sort the modules into power rating classes. Frequently, they hand out those data to laboratories and customers.
- When a laboratory updates or improves its testing equipment, measurements taken with the new equipment have smaller variances than observations collected under the old conditions. The distributional shape of the measurement error may change as well.
- When remeasuring a lot of modules, aging and wearing of measurement devices and modules may result in measurement series with different variances and distributional shapes.
- In round robin experiments (interlaboratory tests) the same modules are analyzed by two (or more) laboratories. The resulting data sets are expected to have the same mean but different variances. The shape of the measurement errors' distribution may also be different.

Those settings have in common that two data sets of the same physical quantity are available, such that one can assume (at least after careful calibration of the measurement systems) that they have the same mean. But, since the data are taken under different measurement systems whose accuracy differs and can not be calibrated for, their variances typically differ. In addition, the shapes of the distributions of the measurement errors may differ as well and, especially, may be non-normal. Then the question arises how one can combine the sample means calculated from the data sets, in order to obtain an estimator of the true power output, which improves upon the estimators calculated from a single data set.

Estimators designed for that purpose are called *common mean estimators* and typically make use of the sample variances to combine the sample means appropriately. Although it is known that their theoretical performance improves upon the sample mean calculated from a single data set, the question how to estimate their variance or standard error is delicate and less well studied. In general, more accurate variance estimators will lead to more accurate confidence intervals, which are highly recommended to quantify the uncertainty present in the data. Known results such as those in [2], where an exact formula for the variance and bounds for the distribution of a widely used common mean estimator are established, are restricted to the case of normally distributed measurements. In photovoltaics, however, power output observations are usually non-normal, such that those results can not be used for data analysis and interpretation. This is also confirmed by the simulations presented in this paper.

In many cases it is a priori known that one series of measurement is more accurate in the sense that its variance is less or equal to the variance of the other series or even strictly smaller. This knowledge

can be translated into the constraint of ordered variances instead of the constraint of unequal variances. Then the question arises how one can incorporate that knowledge in the statistical analysis. This issue is known in the statistical literature as common mean estimation under ordered variances. Estimation under order constraints is, in general, more involved, but for the problem studied here the resulting formulas are easy to implement in practice.

The aim of this article is threefold: First, we review and discuss the above mentioned topics and problems, i.e. i) how to combine sample means from two data sets satisfying the common mean assumption and, especially, ii) how to assess the estimator's uncertainty (i.e. its variance resp. standard error), focusing on data as arising in photovoltaics. Second, we apply a new nonparametric jackknife-type variance estimator to such data. Due to the form of the common mean estimators for ordered variances and the two-sample nature of the problem, the well known classical jackknife variance estimator can not be applied directly and its statistical consistency does not follow from the literature. Here we employ new results about common mean jackknife variance estimators as established in [1], apply it to photovoltaic data and investigate its accuracy by means of a simulation study, which is governed by real photovoltaic data. Lastly, we propose and investigate a novel data adaptive downsampling method to decorrelate a data set.

Photovoltaic data, especially power output measurements, may have quite different distributional shapes. One major reason for this phenomenon is the fact that PV modules are sorted into power rating classes. Therefore, methods which are only valid under a specific model for the distribution of measurement error are not well suited. Consequently, we propose to rely on nonparametric approaches which adapt automatically to the underlying distribution of the data without constraining the distribution to belong to a certain parametric class. For comparison, however, we also review, discuss, apply and investigate a Bayesian approach to the problem.

We present three variance estimators for common mean estimators. As a particular versatile approach we study jackknife variance estimation, which dates back, in its basic form as a method for one-sample statistics, to the work of [3]; see also [4], [1] and the references given therein. Jackknifing represent an easy-to-use and reliable statistical tool, which does neither require restrictive assumptions such as normally distributed measurement error, nor knowledge of variance formulas for the statistic of interest, as it estimates the variance of a statistic by recomputing it from certain subsamples. It is well studied for smooth one sample statistics and has been applied successfully in diverse areas of application, e.g. to analyze fish production rates, see [5]. The application to common mean estimation for photovoltaic data discussed and investigated in the present work, however, requires to jackknife a two sample statistic which is not continuous. So this paper draws on the recent results of [1], where the jackknife variance estimator is generalized to common mean estimators allowing for unequal sample sizes (a situation called unbalanced samples) and ordered variances.

When measurements are taken at a dense time grid as, for instance, under high-throughput production conditions, one may take this into account by downsampling the data appropriately. We propose two approaches which aim to eliminate or dampen autocorrelations. In addition to equidistant downsampling, we propose a data adaptive data thinning procedure, which downsamples the series at irregularly spaced time points, in such a way that autocorrelations present in the downsampled series are minimized. This procedure aims at decorrelation of the measurements based on a lower overall downsampling frequency, in order to obtain larger sample sizes. Both the data analysis and the simulations indicate that this procedure works well.

We investigate the performance of the estimators under discussion and the corresponding confidence intervals resp. credible intervals for the true power output by a simulation study. Instead of studying general but restrictive models for the measurement error, we conduct a case-study approach where the measurement errors follows empirical distributions obtained from real data sets. This, of course, limits the scope of the conclusion one may draw from the study, but it is based on the grounds of real empirical data and distributions. We study two sampling strategies in the simulations. In addition to i.i.d. sampling we consider a resampling procedure related to the block bootstrap, where the drawn subsamples preserve the dependence structure of the series of measurements. Whereas the classical bock bootstrap, however, then calculates the statistics from those blocks, we apply the proposed downsampling procedure at irregularly spaced grid points, in order to decorrelate the subsamples. The statistics are then calculated from those decorrelated data.

The organization of the paper is as follows. In Section 2, we review the common mean estimation problem for unequal as well as ordered variances and discuss the most commonly used estimators. Further, we discuss and elaborate three approaches to estimate their variability, as required for assessing the estimation uncertainty and constructing confidence intervals. The procedure based on the jackknife method draws on recent new results from [1]. Lastly, we review and discuss a Bayesian approach to the common mean problem as a parametric alternative going beyond the likelihood approach. Section 3 is devoted to the data analysis and presents results from simulations. We analyze the performance of the various estimation approaches in terms of bias, variance, root mean squared error and the coverage probability of the associated confidence intervals. In addition, we compare our nonparametric proposal with the Bayesian approach to analyze the data and assess the common mean in terms of a Bayesian credible interval. The simulations also investigate in some detail the accuracy of the proposed data adaptive thinning procedure based on downsampling. Discussion and conclusions are given in Section 4.

2. Methodology

In this section we introduce the common mean estimation problem under heterogeneous as well as ordered variances, discuss estimators aiming at combining the sample means appropriately, and, especially, propose a nonparametric model-free approach to assess their variability. In addition, we propose a procedure to decorrelate the samples, when autocorrelations are present. Lastly, we discuss a Bayesian approach to the common mean problem as a parametric alternative.

2.1. Common mean estimation

Let us review the basic setting of common mean estimation. Suppose we are given two random samples X_1, \dots, X_n and Y_1, \dots, Y_m of sample sizes n and m . We intentionally allow for the case of unequal sample sizes. The common mean assumption asserts that the expectations μ_X and μ_Y of both data sets coincide, i.e. $\mu_X = \mu_Y$, whereas their variances σ_X^2 and σ_Y^2 may differ such that $\sigma_X^2 \neq \sigma_Y^2$. To simplify notation, we shall denote the common mean by μ . The classical common mean model assumes normally distributed measurements. In the present paper, we shall, however, avoid, by purpose, this restrictive assumption that is frequently violated in photovoltaic data analysis, and allow for more general distributional shapes. The nonparametric approach allows for quite general distributional shapes as long as higher order moments are finite, see [1] for details, and the Bayesian approach introduces

a prior distribution for the parameters which leads to more general distributional shapes.

Denote by \bar{X}_n and \bar{Y}_m the sample means and let $S_{X,n}^2$ and $S_{Y,m}^2$ be the corresponding unbiased variance estimators. The classical common mean problem under normality seeks convex combinations $\widehat{\mu} = w\bar{X}_n + (1-w)\bar{Y}_m$ of the sample means, which provide unbiased estimators for μ and minimize some optimality criterion such as mean squared error (MSE)

$$\text{MSE}(\widehat{\mu}; \mu) = E(\widehat{\mu} - \mu)^2$$

or equivalently the root mean squared error (RMSE) defined by

$$\text{RMSE}(\widehat{\mu}; \mu) = \sqrt{E(\widehat{\mu} - \mu)^2}.$$

For known variances the optimal, i.e. minimum variance unbiased estimator, can be calculated and is given by the weight $w = n\sigma_Y^2 / (n\sigma_Y^2 + m\sigma_X^2)$. When the variances are unknown and therefore need to be estimated, a natural choice is the Graybill-Deal (GD) estimator, see [6], which uses the random weight $w = nS_{Y,m}^2 / (nS_{Y,m}^2 + mS_{X,n}^2)$. It can be expressed as

$$\widehat{\mu} = \frac{nS_{Y,m}^2}{nS_{Y,m}^2 + mS_{X,n}^2} \bar{X}_n + \frac{mS_{X,n}^2}{nS_{Y,m}^2 + mS_{X,n}^2} \bar{Y}_m.$$

Observe that this estimator is obtained from the minimum variance unbiased estimator by substituting the unknown variances by their unbiased estimators.

2.2. Estimation under ordered variances

In photovoltaics one has often a priori knowledge about the precision of the measurements. For instance, when combining data sets from two laboratories one of them may systematically produce better measurements due to better equipment, longer experience or superior expertise. Then we wish to combine the sample means under the additional constraint of ordered variances $\sigma_X^2 \leq \sigma_Y^2$. In this case it is natural to consider the ordering of the variance estimators $S_{X,n}^2$ and $S_{Y,m}^2$. It has been shown that, under the assumption of normality, the so-called Nair estimator

$$\widehat{\mu}_n^{(Nair)} = \begin{cases} \widehat{\mu}, & S_{X,n}^2 \leq S_{Y,m}^2, \\ \frac{n}{n+m} \bar{X}_n + \frac{m}{n+m} \bar{Y}_m, & S_{X,n}^2 > S_{Y,m}^2, \end{cases}$$

improves upon the GD-estimator $\widehat{\mu}$ in terms of the MSE. This means, if the samples are in agreement with the order constraint, one uses the GD-estimator. Otherwise, one switches to the simpler estimator which weights the sample means by the sample sizes, which is also formally obtained when $S_{X,n}^2 = S_{Y,m}^2$, i.e. when projecting $(S_{X,n}^2, S_{Y,m}^2)$ onto the set $\{(s_1, s_2) : 0 \leq s_1 \leq s_2 < \infty\}$ of admissible values (under the order constraints) for the pair of variances. There exist more complicated estimators for ordered variances, but we confine ourselves to the above proposal, due to its simplicity for practical use.

2.3. Variance estimation

Recall that the variance of the sample mean \bar{X}_n can be easily estimated by the standard error $S_{X,n}^2/n$. For non-random weights the variance of a convex combination $w\bar{X}_n + (1-w)\bar{Y}_m$ can be estimated

by $w^2 \frac{S_{X,n}^2}{n} + (1-w)^2 \frac{S_{Y,m}^2}{m}$. For the common mean estimator $\widehat{\mu}$ it is, however, much more complicated to assess and estimate its variability, since the weights are random. Assuming normally distributed measurements, [2] obtained an exact formula for the variance in terms of an infinite series,

$$\text{Var}(\widehat{\mu}) = \frac{\sigma_x^2/n}{\beta(n/2, m/2)} \sum_{i=0}^{\infty} (i+1)(1-a)^i [\beta(n/2+i, m/2+2) + (m/n)a\beta(n/2+i+2, m/2)],$$

where $\beta(x, y)$ denotes the β -function. This formula suggests the estimator

$$\widehat{\sigma}_1^2 = \frac{S_{X,n}^2/n}{\beta(n/2, m/2)} \sum_{i=0}^M (i+1)(1-a)^i [\beta(n/2+i, m/2+2) + (m/n)(S_{X,n}^2/S_{Y,m}^2)\beta(n/2+i+2, m/2)],$$

where the infinite sum is replaced by a finite sum taking into account the terms up to the index M . The constant M can be rather small; indeed $M = 10$ suffices for practical purposes, since the series in the exact formula converges quickly. We investigate in the simulation study whether the estimator $\widehat{\sigma}_3^2$ nevertheless works for practical purposes, i.e. non-normal data sets.

The exact formula behind the estimator $\widehat{\sigma}_1^2$, however, no longer holds for non-normal data. An alternative approach makes use of the fact that the central limit theorem for $\widehat{\mu}$, which holds true under rather weak assumptions and, especially, for non-normal measurements, suggest the variance estimator

$$\widehat{\sigma}_2^2 = \frac{\left(\frac{nS_{Y,m}^2}{nS_{Y,m}^2+mS_{X,n}^2}\right)^2 \frac{n+m}{n} S_{X,n}^2 + \left(\frac{mS_{X,n}^2}{nS_{Y,m}^2+mS_{X,n}^2}\right)^2 \frac{n+m}{m} S_{Y,m}^2}{n+m},$$

which is consistent as long as the observations have a finite fourth moment.

The jackknife is a versatile and useful statistical tool for statistical problems such as bias and variance estimation. It has been studied extensively for one-sample problems and has been recently extended by [1], on which our exposition draws, to two-sample problems and the common mean problem under ordered variances as discussed here.

The basic idea is to create replicates of the statistic of interest by leaving out one of the available observation and recalculate the test statistic from that subsample. In this way, one obtains n replications of the test statistics which can be used to estimate its variability. For equal sample sizes, $n = m$, the classical jackknife approach can be used as follows: First, pair the observations, $(X_1, Y_1), \dots, (X_n, Y_n)$. Now omit the i th pair and calculate $\widehat{\mu}_{-i} = \widehat{\mu}((X_1, Y_1), \dots, (X_{i-1}, Y_{i-1}), (X_{i+1}, Y_{i+1}), \dots, (X_n, Y_n))$ for $i = 1, \dots, n$. Then the jackknife variance estimator for $\text{Var}(\widehat{\mu})$ is defined by

$$\widehat{\text{Var}}(\widehat{\mu}) = \frac{n-1}{n} \sum_{i=1}^n \left(\widehat{\mu}_{-i} - \frac{1}{n} \sum_{j=1}^n \widehat{\mu}_{-j} \right)^2.$$

There are two drawbacks of this approach: First, pairing of the observations is arbitrary and the estimate may depend on how data points have been paired. Second, the approach fails for unequal sample sizes.

In [1] jackknife variance estimation for two sample statistics and especially common mean estimators has been studied from a theoretical viewpoint and extended to unequal sample sizes and the case of ordered variances. That new jackknife variance estimator is calculated as follows: First, fuse the data sets and write

$$(Z_1, \dots, Z_{n+m}) = (X_1, \dots, X_n, Y_1, \dots, Y_m),$$

i.e. Z_1, \dots, Z_n are the measurements of the X -sample and Z_{n+1}, \dots, Z_{n+m} are the measurements of the Y -sample. Let $\widehat{\mu}_{-i}$ denote the common mean estimator obtained when leaving out the i th measurement. From those leave-one-out estimators we calculate the pseudo values

$$\xi_i = n\widehat{\mu} - (n-1)\widehat{\mu}_{-i}, \quad i = 1, \dots, n+m.$$

The jackknife variance estimator is then given by

$$\widehat{\sigma}_3^2 = \frac{n}{n+m}\widehat{\tau}_1^2 + \frac{m}{n+m}\widehat{\tau}_2^2,$$

with

$$\widehat{\tau}_1^2 = \frac{1}{n-1} \sum_{i=1}^n (\xi_i - \bar{\xi})^2$$

and

$$\widehat{\tau}_2^2 = \frac{1}{m-1} \sum_{i=n+1}^{n+m} (\xi_i - \bar{\xi})^2.$$

Observe that the recalculation of the estimator also covers the weights used to combine the sample means. In this way the jackknife automatically takes into account the variability of the common mean estimator which is due to random weights.

Let us briefly discuss consistency issues. Recall that a statistical estimator $\widehat{\theta}_n$ for a parameter (or characteristic) $\theta = \theta(F)$ depending on the underlying distribution F of the data is called consistent, if $\widehat{\theta}_n$ converges in probability to θ , as the sample size n tends to ∞ . Consistency is an indispensable statistical property. It allows, amongst other properties, to replace the unknown asymptotic variance, $\sigma^2(T_n)$, of a statistic T_n satisfying a central limit theorem such that $T_n \sim_{approx} N(\mu_n, \sigma_n^2(T)/n)$ for large n , by such a consistent estimator $\widehat{\sigma}^2(T_n)$. This allows to employ the normal approximation to construct procedures for statistical inference. [4] provide general theoretical results for one-sample statistics which are asymptotically equivalent to a linear statistic in the sense that the remainder term is of negligible order in terms of second moments. Indeed, for the one sample problem, consistency of the jackknife has been established for general classes of statistics such as M estimators, rank statistics and continuously Gâteaux differentiable statistical functionals, see e.g. [7] and [8]. In [1] it is shown that the new two-sample jackknife variance estimator $\widehat{\sigma}_3^2$ is consistent for the variance of a large class of common mean estimators including estimations such as the Nair estimator, which is based on a discontinuous function of the unbiased variance estimators, under weak assumptions.

Although the jackknife represents a quite general statistical tool, its validity in the sense of consistency as a statistical estimator relies on the assumption of independent and identically distributed random samples. This restriction is, of course, not limited to the jackknife discussed above, but also applies to other statistical approaches designed for i.i.d. random samples such as the Bayesian approach discussed below. In order to handle correlated series of measurements, we propose below a downsampling procedure to decorrelate the data.

2.4. Confidence interval

As shown in [1], the common mean estimators discussed above are asymptotically normal, which allows us to construct confidence intervals for the true unknown mean (power output) given a valid

variance estimator. Given a variance estimator $\widehat{\sigma}^2$ for $\text{Var}(\widehat{\mu})$, one calculates the interval

$$CI = [\widehat{\mu} - q_{1-\alpha/2}\widehat{\sigma}, \widehat{\mu} + q_{1-\alpha/2}\widehat{\sigma}],$$

where $q_{1-\alpha/2}$ denotes the $(1-\alpha/2)$ -quantile of the standard normal distribution, i.e. $\Phi(q_{1-\alpha/2}) = 1-\alpha/2$. Then for large enough sample sizes it holds

$$P(\mu \in CI) \approx 1 - \alpha,$$

such that CI represents an approximate confidence interval with nominal coverage probability $1 - \alpha$. This follows from the validity of the central limit theorem for common mean estimators with random and possibly non-smooth weights (under ordered variances), see [1].

2.5. Data thinning for decorrelation and massive data sets

Depending on how the PV modules are characterized, there may be serial correlation present in the data, i.e. we are given time series and not i.i.d. samples. Such correlations can result, e.g., from the high-frequency data acquisition required for high-throughput production processes. By appropriately downsampling the data one can obtain *thinned data sets* which aim at eliminating or at least substantially damping autocorrelations without changing the marginal distribution of the observations. Going beyond a simple downsampling approach, one may downsample the series at an irregularly spaced grid, in order to obtain a downsampled subseries which minimizes autocorrelations. To the best of the author's knowledge, this approach is novel and has not yet been studied in the literature. In the simulation study we investigate to which extent this data-adaptive approach leads to accurate results.

Downsampling, sometimes also called subsampling (such that it can be confused with the subsampling method related to the bootstrap for time series, see e.g. [9]), means that one re-samples a series of observations at a lower sampling frequency by taking only each ℓ th observation, where $\ell > 1$ is a natural number. Often, one assumes that the time series has been obtained by sampling from a continuous time signal $\mathcal{X}(t)$ at equidistant time points $t_i = i\Delta$, such that

$$X_i = \mathcal{X}(t_i) = \mathcal{X}(i\Delta), \quad i = 1, \dots, n,$$

for some sampling period Δ corresponding to a sampling frequency $f = 1/\Delta$; In this context downsampling aims at sampling that signal at a lower sampling frequency $f' = 1/\Delta' < f$.

If the data is ℓ -dependent, i.e. two observations are independent whenever their time lag is larger than ℓ , then the resulting downsampled data set consists of independent and identically distributed data. An important special case of a ℓ -dependent series is a moving average process of order $q = \ell - 1$ with i.i.d. innovations. For other forms of dependence, serial correlations could still be present in the downsampled series, but usually the autocorrelations will be substantially dampened, depending on how quickly the autocorrelation functions decays. Downsampling can also be an effective tool when dealing with massive production data, i.e. too large data sets, whose analysis would require too much computing resources.

Let us denote the original and possibly serially correlated data set by X_1, \dots, X_n , assuming the observations are ordered in time. Downsampling outputs the series

$$X_{thin,i} = X_{\ell i}, \quad i = 1, \dots, m,$$

where m is the largest integer less or equal to n/ℓ . n/m is called the *downsampling rate*.

The thinned data set $X_{thin,1}, \dots, X_{thin,m}$ obtained by this downsampling approach now represents a subsample of size m from the original data set with sample mean and variance close to the sample mean and variance of the original data sets, such that a common mean estimator calculated from the thinned data sets will be close to the estimate calculated from the original series as well. But since the downsampled series can be assumed to be (approximately) i.i.d, the assessment of the estimator's variability can be based on simpler formulas which assume i.i.d. data.

The question arises whether one can draw a larger subsample with the above properties by downsampling at a lower downsampling rate, in such a way that the autocorrelations are minimized. This can be achieved as follows by downsampling at an irregular grid: Fix $m < n$ and let $1 \leq i_1 < \dots < i_m \leq n$ be m integers. Then the associated thinned (downsampled) series is $X_{i_1}, X_{i_2}, \dots, X_{i_m}$. Consider the optimization problem to select the grid $i_1 < \dots < i_m$ of size m in such a way that the first, say, r autocorrelations of the resulting subseries are minimized. This means, $1 \leq i_1^* < \dots < i_m^* \leq n$ are selected such that

$$M_{X,m}(i_1^*, \dots, i_m^*) \leq M_{X,m}(i_1, \dots, i_m)$$

for all $1 \leq i_1 < \dots < i_m \leq n$, where $M_{X,m}(i_1, \dots, i_m)$ is a measure which penalizes autocorrelations, e.g.

$$M_{X,m}(i_1, \dots, i_m) = \max_{h=1, \dots, r} |\widehat{\rho}_{X,m}(h; i_1, \dots, i_m)| \quad (1)$$

where

$$\widehat{\rho}_{X,m}(h; i_1, \dots, i_m) = \widehat{\gamma}_{X,m}(h; i_1, \dots, i_m) / \widehat{\gamma}_{X,m}(0; i_1, \dots, i_m)$$

with

$$\widehat{\gamma}_{X,m}(h; i_1, \dots, i_m) = \frac{1}{m} \sum_{j=1}^{m-h} (X_{i_j} - \bar{X}_m)(X_{i_{j+h}} - \bar{X}_m)$$

and $\bar{X}_m = \frac{1}{m} \sum_{j=1}^m X_{i_j}$ is the lag- h autocorrelation coefficient of the series sampled at the grid $i_1 < \dots < i_m$. Instead of the maximal autocorrelation one could also use a sum of squares criterion in (1). The above combinatorial optimization problem over the set of ordered sequences of m integers from the set $\{1, \dots, n\}$ is characterized by its highly nonlinear objective function $M_{X,m}(i_1, \dots, i_m)$. An exhaustive search over the whole solution space $\mathcal{D} = \{(i_1, \dots, i_m)' \in \mathbb{N}^m : 1 \leq i_1 < \dots < i_m \leq n\}$ is, in practice, infeasible. Therefore, we propose to conduct a random search algorithm where, say L , randomly drawn sets of ordered indices are drawn. Then one uses that set which leads to the optimal value of the objective function.

2.6. Bayesian approach

Let us contrast the above nonparametric approach with a Bayesian common mean model as studied in [10]. Let us briefly review the basics; for a comprehensive exposition we refer to [11]. The Bayesian approach assumes that the parameter θ of a stochastic model for the observed data is random and distributed according to a prior distribution $\pi(\theta)$ over the domain Θ . The data X is assumed to follow a known parametric model given the parameter, i.e.

$$X | \theta \sim f(x|\theta).$$

This means, the conditional distribution is modeled by a parametric family $f(\cdot|\theta)$, $\theta \in \Theta$, of distributions, typically given by densities, given the parameter θ . The joint distribution is then given by $(X, \theta) \sim f(x|\theta)\pi(\theta)$ and the posterior distribution of the parameter θ given (i.e. having observed the data) x is

$$\theta|x \sim f(\theta|x) = \frac{\int f(x|\theta)\pi(\theta)}{\int f(x|\theta')\pi(\theta')d\theta'}.$$

Usually, the prior belongs to some class of distributions, such that $\pi(\theta) = \pi(\theta; \eta)$ for some hyperparameter η , typically allowing for explicit formulas. In some case, e.g. for the important class of exponential families, a conjugate prior exists, such that the posterior distribution is of the prior form, i.e. $\pi(\theta; \eta_x)$ for some value η_x depending on the observed data. In this case, the calculation of the posterior, i.e. processing the observed data, is simply given by a transformation of the parameter η .

The common mean model as studied in [10] assumes Gaussian models, $N(\mu, \sigma_i^2)$, $i = 1, 2$, given the parameter $\theta = (\mu, \sigma_1, \sigma_2)$ or, alternatively, $(\mu, \sigma_1^2, \sigma_2^2)$. The joint density of the samples $x = (x_{11}, \dots, x_{2n_2})$ given the parameter is

$$p(x|\theta, \sigma_1, \sigma_2) \propto \prod_{i=1}^2 \sigma_i^{-n_i} \exp\left(-\frac{1}{2\sigma_i^2}[(n_i - 1)s_i^2 + n_i(\mu - \bar{x}_i)^2]\right)$$

where $s_i^2 = \frac{1}{n_i-1} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$ and $\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij}$. Let us discuss the choice of the priors. For laboratory data as studied in this article, where the precision could vary to some extent but is controlled for, such that large values are highly unlikely, the conjugate prior for the precisions is not adequate, since it has a long right tail. Instead, vague priors such as $p(\tau_i) \propto 1/\tau_i$, $i = 1, 2$, are more appropriate and a constant (improper) prior for the common mean. The prior $\pi(\theta, \sigma_1, \sigma_2) = 1/(\sigma_1\sigma_2)$, corresponding to independence of the variances, is also called righ-Haar prior, see e.g. [12]. It arises as the limit of the conjugate prior for $\alpha, \beta \rightarrow 0$. Alternatively, one could use uniform distributions, implementing the principle of indifference over some finite interval or use improper priors, which are, however, not implemented in standard statistical software and thus have to be approximated by proper flat priors. Jeffreys priors are given by $I(\theta)^{1/2}$ where $I(\theta)$ is the information matrix. For a Gaussian model this leads to $\pi(\mu, \sigma^2) = 1/\sigma^2$, a prior which is often not recommended, whereas the single-parameter Jeffrey priors are $\pi(\mu) = 1$ and $\pi(\sigma) = 1/\sigma$. The choice of the prior is a great concern in a Bayesian analysis and may have a noticeable effect on the analysis. See [12] for a recent thorough discussion and references to the literature.

Bayesian inference relies on the posterior distributions of the parameter θ given the observed data, $f(\theta|x)$, which is used for prediction, point estimation and the construction of credibility intervals. The latter is usually constructed by central quantiles of the posterior distribution, e.g. $CI = [q(\theta|x)_{\alpha/2}, q(\theta|x)_{1-\alpha/2}]$, where $q(\theta|x)_p$ denotes the p -quantile of the posterior distribution $f(\theta|x)$. As explicit formulas are only available in special cases, one relies on simulation methods, especially Monte Carlo Markov Chain (MCMC), in order to sample from the posterior distributions. Those samples are then used to make inference.

The question arises how the posterior distribution behaves when the sample size increases. A posterior, $\pi(\theta|x_n)$, is called consistent, if $\pi(\theta|x_n)(U)$ converges to 1, as $n \rightarrow \infty$, for any neighborhood of the true parameter value θ_0 governing the data (i.e. $X \sim f(x|\theta_0)$). For smooth priors, Schwartz's theorem ensures consistency, provided for every neighborhood U one may find a statistical test for $H_0 : \theta = \theta_0$

against $H_1 : \theta \notin U$ with power strictly greater than size and if the prior assigns positive probability to the set of parameter values for which the Kullback-Leibler distance between $f(\cdot|\theta)$ and $f(\cdot|\theta_0)$ is small. But this result does not apply to uniform priors or improper priors. Here results of Ibragimov and Has'minskii ensure consistency under general conditions, see [13], on which our discussion draws, for further discussion and references.

Consistency is an important aspect and a Bayesian model for which consistency does not hold should certainly not be used, since then the procedure does not learn the correct model even for infinite sample size and yields a biased guess of the distribution of the data. However, in the presence of consistency, for large sample sizes the dependence of the posterior distribution on the prior vanishes and the posterior distribution is close to the true but unknown model $f(x|\theta_0)$. If the model is, however, misspecified, i.e. if the true distribution $f_0(\cdot)$ of the data differs from $f(x|\theta')$ for all $\theta' \in \Theta$, then the Bayesian analyst works with a misspecified model for the data.

The Bayesian notion corresponding to the frequentist's notion of a confidence interval is the concept of a credible interval (or set): A set $C_B = C_{B,x}$, depending on the observed data x is called α -credible set, for a given prior π , if $P(\theta \in C_{B|x}) \geq 1 - \alpha$. Notice that here the posterior distribution of the parameter given the observed data x is used to set up the coverage probability, whereas the frequentist's approach is to calculate that probability with respect to the distribution of the data and requiring that it is larger or equal to the nominal confidence level, for each fixed value of the non-random but unknown parameter.

3. Data Analysis and Simulations

3.1. Data analysis

The data used in this paper for simulation and data analysis consists of power output measurements under STC of 385 modules obtained from two sun flashers. One of those series are taken at TÜV Rheinland Energy GmbH, the other series was made by the manufacturer using another flasher device. Analyzing the data shows that $\bar{x} = 178.6614$ and $s_x^2 = 3.543361$, whereas $\bar{y} = 178.2499$ and $s_y^2 = 4.917217$. Although the sample means are quite close, the measurement uncertainty differs substantially. Both data sets are non-normal, as statistically confirmed by the Shapiro-Wilk test for normality.

Figure 1 depicts a quantile-quantile plot of both series (bold line), i.e. the sample quantile functions are plotted against each other. It can be seen that the shapes of distributions differ substantially. This data set is an example where the flash data from the manufacturer can not be easily used to infer the distributional shape of measurements taken in the laboratory and would require a deeper investigation going beyond the scope of the present article.

In order to illustrate the data thinning procedure discussed in Section 2.5, both data sets were down-sampled by taking every 6th data point, i.e. $\ell = 6$. Figure 1 compares the quantile-quantile plot of the original series and the corresponding curve of the equidistantly downsampled series (dashed line).

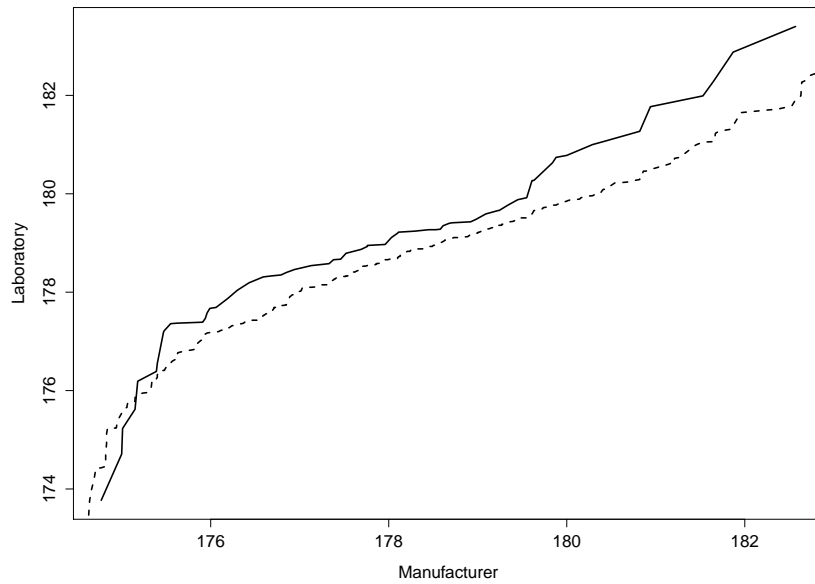


Figure 1. Quantile-quantile plots of the original data sets and the equidistantly thinned series (dashed line).

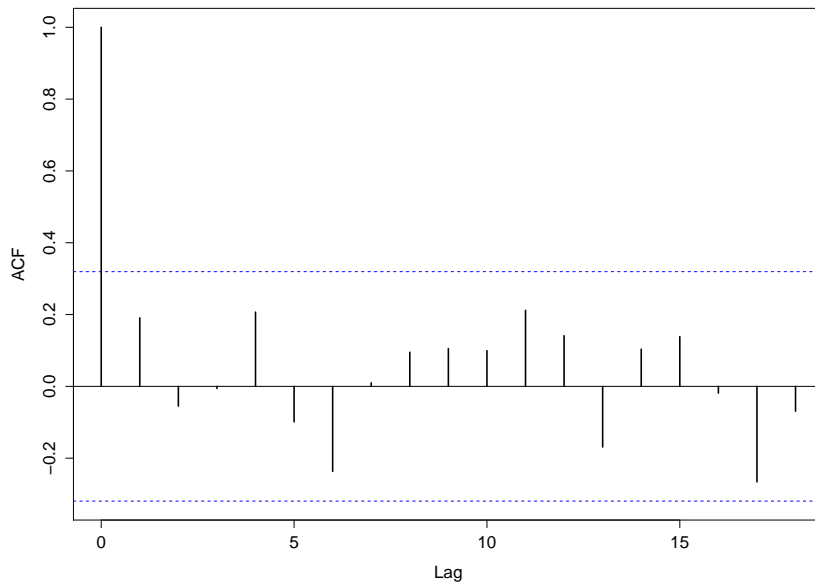


Figure 2. Autocorrelation function of the equidistantly thinned manufacturer data.

The autocorrelation functions of the downsampled manufacturer and laboratory data are shown in Figure 2 and 3. Those plots also show confidence intervals for a nominal confidence level of 99%, in order to allow for multiple testing of the $\ell - 1 = 5$ hypothesis tests about the lag h autocorrelations for

$h = 1, \dots, 5$, against the strong white noise null hypothesis. Note that the global significance level is then given by $\alpha = 5\%$. One can notice that both the manufacturer data and the laboratory data show no significant serial correlations.

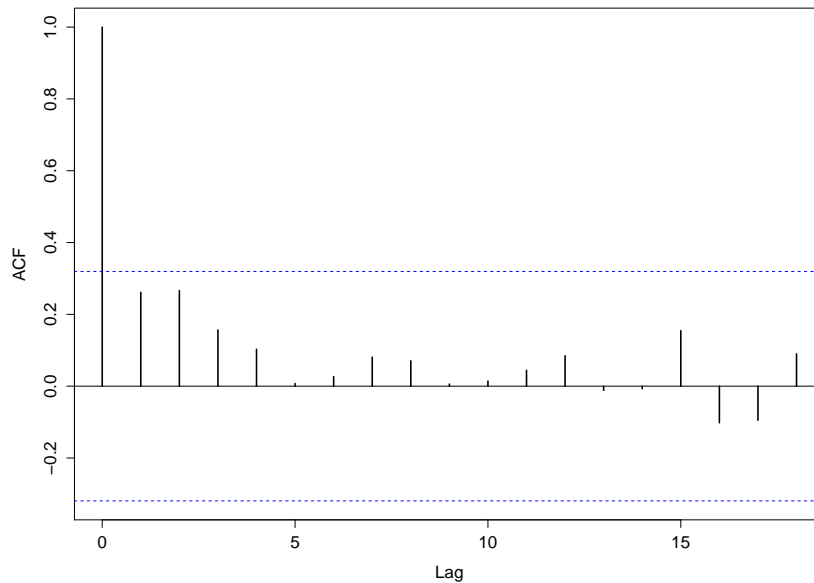


Figure 3. Autocorrelation function of the equidistantly thinned laboratory data.

The above preprocessing results in data sets of size 65. In order to obtain larger thinned data sets, we downsampled the series at an irregular grid of size $m = n/3 = 128$. As described in the previous section, the grid was obtained by a random search. Figure 4 shows the trajectory of the objective function for 1,000 iterations. This number of iterations was also used in the simulations.

The associated quantile-quantile plot is shown in Figure 5 and the autocorrelation functions are depicted in Figure 6 and 7. It can be seen that even for a smaller downsampling rate of $n/m = 3$ instead of 6 the proposed data-adaptive approach minimizing autocorrelations leads to samples whose quantile-quantile plot is quite similar to the quantile-quantile plot of the original data. Simultaneously, the irregularly space downsampled series show no significant autocorrelations, such that an i.i.d. assumption is justifiable from a statistical point of view.

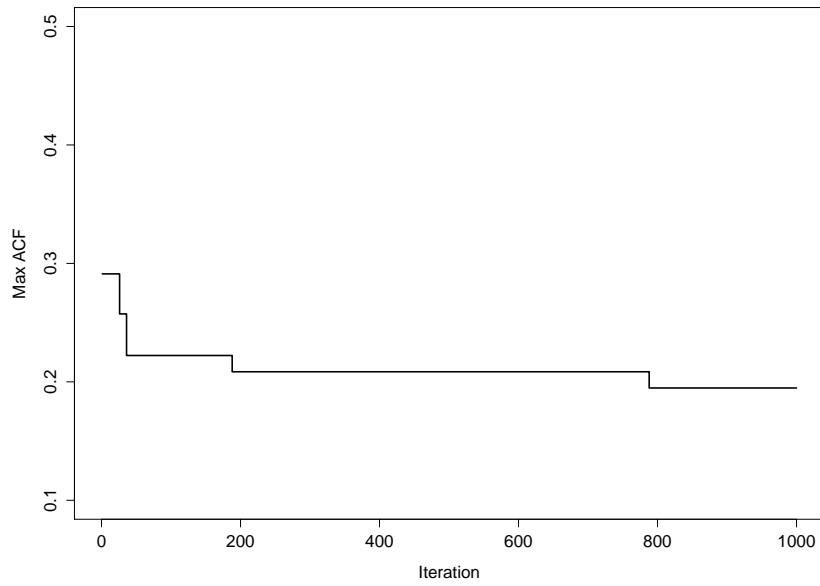


Figure 4. Illustration of the random search algorithm to find an optimal irregularly spaced grid.

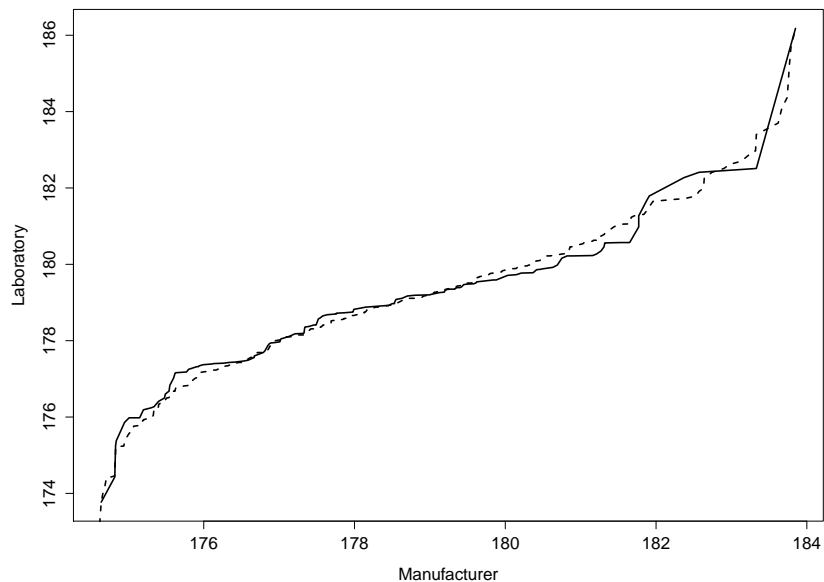


Figure 5. Quantile-quantile plot of the data sets thinned by downsampling at an irregularly spaced grid of size 128.

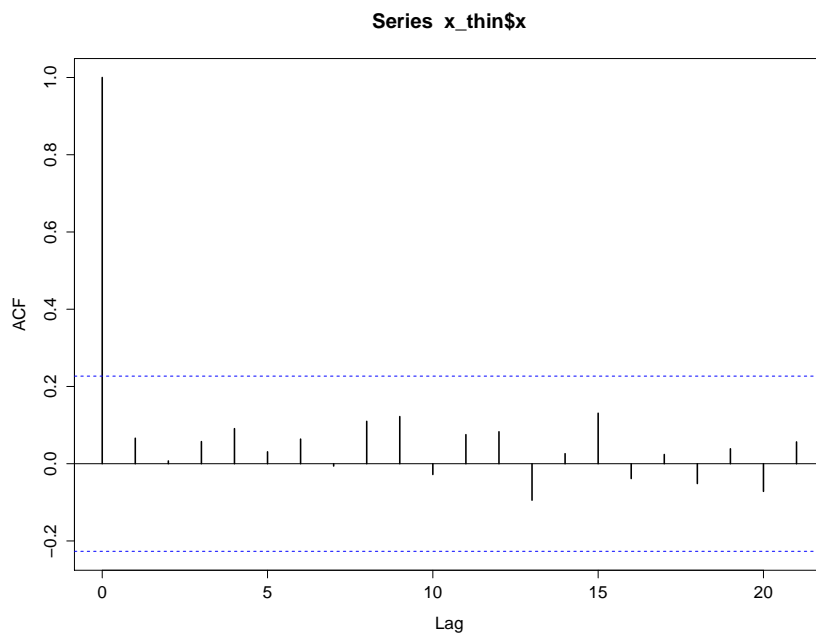


Figure 6. Autocorrelation function of the manufacturer data downsampled at an irregular grid of size 128.

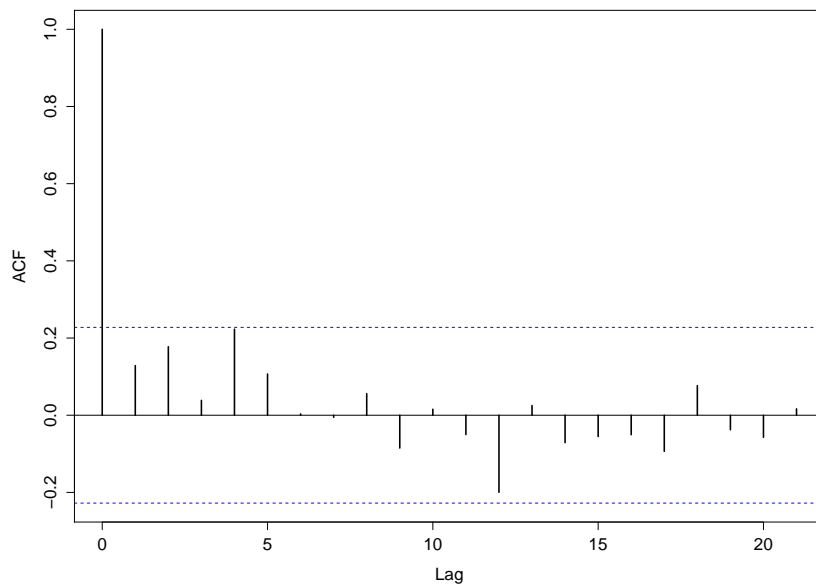


Figure 7. Autocorrelation function of the laboratory data downsampled at an irregular grid of size 128.

The common mean is estimated by 178.40 with jackknife standard error 0.194 when using the Graybill-Deal estimator and by 178.36 with a jackknife standard error 0.171 when relying on the Nair

estimator for ordered variances. Here, for illustration, we applied Nair's estimator with the second sample having the smaller variance estimate, since otherwise both estimators coincide.

Fixing the nominal confidence level at 95%, let us illustrate the confidence intervals discussed above. The asymptotic confidence interval for the common mean using the Nair estimator is given by

$$CI_{as} = [177.86, 178.72]$$

whereas the jackknife variance estimator leads to the *tighter* confidence interval

$$CI_{jack} = [178.03, 178.78].$$

The simulation study below shows that, overall, the jackknife confidence intervals are more accurate from a statistical point of view and should therefore be preferred in data analyses.

As suggested by a referee, we also investigated the Bayesian approach. We studied two Bayesian models with non-conjugate priors. For model B1 the priors were selected as

$$\theta \sim N(180, 20^2), \quad \pi(\sigma_1, \sigma_2) = 1/(\sigma_1\sigma_2).$$

The flat normal prior for θ is used as a proxy for an improper constant prior. The specification for the second Bayesian model (B2) assumes that the uncertainty about the standard deviations can be modeled by uniform distributions over the interval [1.5; 3], i.e.

$$\theta \sim N(180, 20^2), \quad \pi(\sigma_j) \sim U(1.5; 3), \quad j = 1, 2.$$

Computations were done using R and JAGS. Markov Chain Monte Carlo (MCMC) simulations to sample from the posterior distributions were conducted with a burn-in-period of 1,000 runs. Then a sample of size 5,000 was simulated from the posterior distribution of the common mean parameter given the data, in order to determine the credible interval. Figure 8 shows the Gelman-Rubin diagnostics plots for model B1, which are passed, when running four MCMC chains. For model B2 the Gelman-Rubin plot leads to a similar result indicating convergence of the MCMC chain and is therefore omitted.

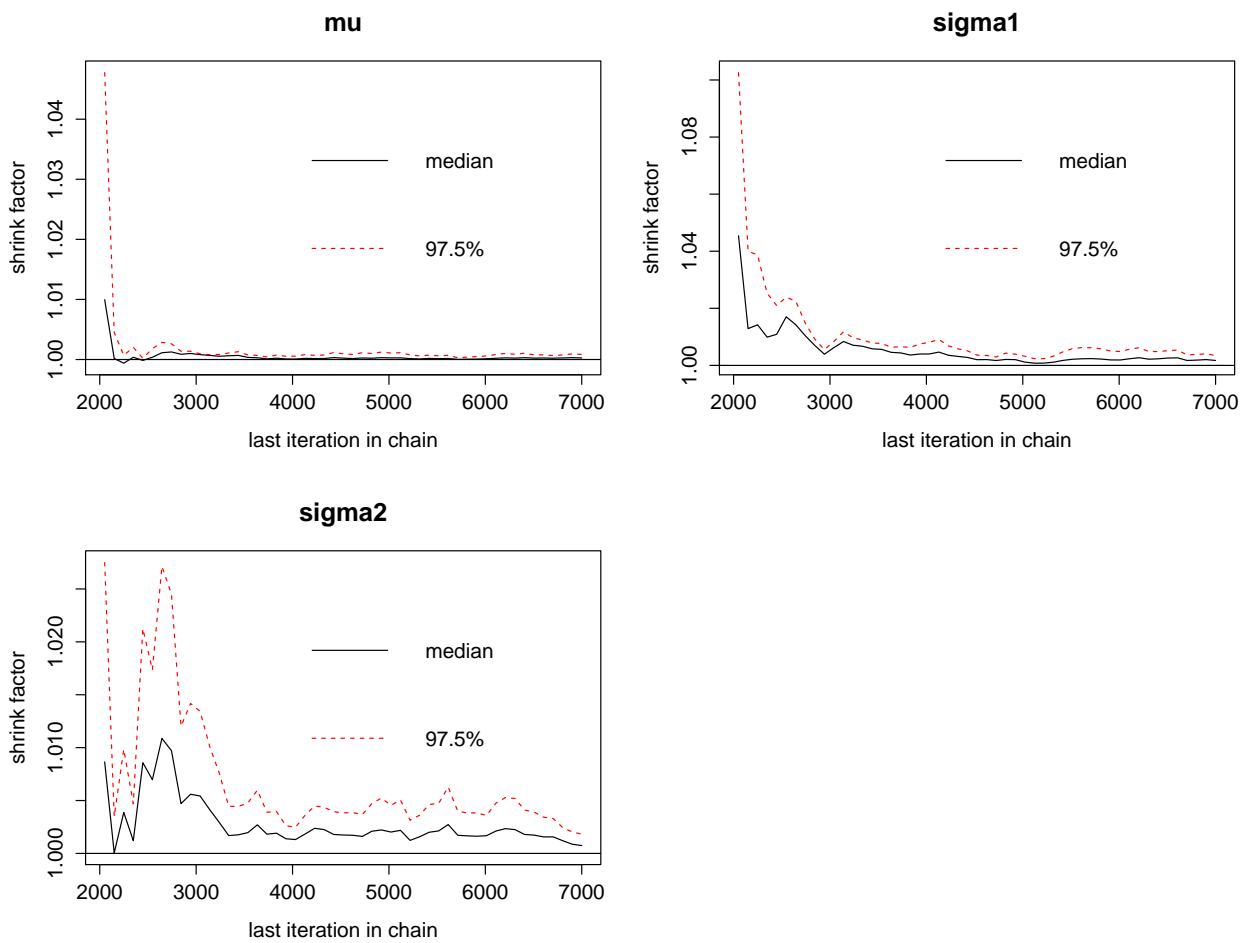


Figure 8. Gelman-Rubin diagnostics plot for model B1.

The mean of the (simulated) posterior distribution is 178.32 and for model B1 the credible interval is simulated as [178.15, 178.47]. For model B2 we obtain the interval [178.19, 178.46]. Both intervals are tighter than the nonparametric intervals, but the simulations below show that the nominal coverage probability of 95% is not reached.

3.2. Simulations

The aim of the simulation study was to investigate the statistical quality of the discussed variance estimators for common mean estimation and, especially, the accuracy of the associated confidence resp. credible intervals for the common mean in terms of their *coverage probabilities*. Further, the aim was to analyze how the proposed data adaptive procedure to decorrelate time series measurements works for the problem at hand.

3.2.1. I.I.D. Sampling

The simulations address all three variance estimators for the Graybill-Deal common mean estimator. The jackknife variance estimator was also applied to the Nair estimator for ordered variance, but we

do not report the results, since they were quite similar to the results for the Graybill-Deal estimator. In particular, we are interested in investigating the accuracy of the jackknife approach. We focus on the case that the measurements follow the empirical distribution of the given industrial PV data set of power output measurements. This means, we benchmark the methods with real PV distributions instead of using simulation models for the measurement errors. The data sets were calibrated in such a way that both of them have the same mean by transforming the Y -data using the calibration function $k(y) = y - 0.4415$ calculated from the full original data. Then, for various sample sizes random samples were drawn from both data sets to mimic future testing of modules.

Since usually large flasher lists are available from the manufacturer, we were interested in highly unbalanced samples where the sample size of the Y -sample (from the manufacturer or previous experiments) is much larger than the sample size of the X -sample taken in the laboratory. Here the question is to which extent measurements taken at the production line can substitute lab measurements of higher quality. A similar situation occurs when a data set of a lot is available and additional test measurements are made, in order to check the current power output. If there is no evidence for a loss of power, one can combine the data to estimate the true power output using both data sets. The sample size required to decide, on statistical grounds, whether or not the lot is in agreement with the specs, can be determined using the PV-specific acceptance sampling methods provided by [14], [15], [16] and [17].

We evaluated the variance estimators in terms of their bias, variance and RMSE. Those quantities reflect systematic deviations (bias), estimation error (SE) and the overall statistical quality (RMSE). The main purpose, however, of estimating the variance of the common mean estimator is to calculate a confidence interval, in order to assess the precision resp. uncertainty of the point estimate for the mean power output given the available data. Since for a confidence interval the real coverage probability that the calculated confidence interval indeed contains the unknown true power output is of main interest. Therefore, we determined, by simulation, the true coverage probabilities of the confidence intervals, for each method of variance estimation.

The sample size n of the X -sample representing high-quality laboratory measurements was chosen between 10 and 30. The sample size m of the second data set representing measurements from the manufacturer (or previous data of lower accuracy) was set equal to the sample size from the lab and then further increased up to 200. For each combination obtained in this way, the statistical measures discussed above were determined for the variance estimator $\hat{\sigma}_1^2$ based on Nair's formula, the estimator $\hat{\sigma}_2^2$ arising from the central limit theorem and the jackknife variance estimator $\hat{\sigma}_3^2$. Each quantity was estimated based on 10,000 simulation runs. The nominal confidence level for our study was chosen as 95%.

The simulation study also covers the two Bayesian models discussed above. For each simulation run samples of size 5,000 from the posterior distribution were simulated by the MCMC approach with a burn-in period of 1,000, in order to determine the credible intervals for the common mean.

The results are shown in Table 1. Concerning the interpretation of the numbers given there, it is important to note that the columns SE and RMSE provide measures for the accuracy of the variance estimator and not of the common mean estimator. Especially, SE is the square root of the variance of the variance estimator of the common mean estimator and not its standard error. The coverage probabilities of the confidence resp. credible intervals for the common mean are denoted by p_{Nair} for Nair's method, p_{as} when using the variance estimator based on the asymptotic normality, p_{jack} for the confidence interval based on the jackknife variance estimator of the GD common mean estimator, and

the coverage probabilities for the two Bayesian credible intervals are denoted by p_{CI-B1} for model B1 and p_{CI-B2} for model B2.

It can be seen that the variance estimator based on Nair's formula does not work properly. The asymptotic variance estimator provides very good results in terms of the RMSE and is better in this respect than the jackknife variance estimator. But the accuracy of asymptotic confidence intervals, p_{as} is poor compared to the highly accurate confidence intervals based on the jackknife, p_{jack} .

The Bayesian approach is disappointing. The coverage probabilities p_{CI-B1} for model B1 are noticeably lower than the nominal coverage probability 95% even for large sample sizes. This indicates that the Bayesian intervals are too optimistic in terms of the uncertainty. This effect is even more pronounced for the Bayesian model B2.

In all cases we studied here the confidence intervals constructed from the jackknife variance estimator are highly accurate and definitely outperform the other approaches. Even when it comes to highly unbalanced samples the jackknife's assessment of the estimator's variability and the associated confidence interval are highly reliable from a statistical point of view.

Table 1. Accuracy of Variance Estimators and Confidence Intervals

n	m	Nair			Asymptotics			Jackknife			Bayes				
		Bias	SE	RMSE	p_{as}	Bias	SE	RMSE	CI	Bias	SE	RMSE	p_{jack}	p_{CI-B1}	p_{CI-B2}
10	10	-0.006	0.018	0.019	0.917	-0.010	0.016	0.019	0.904	0.008	0.033	0.034	0.940	0.908	0.875
10	25	0.041	0.019	0.045	0.992	-0.004	0.007	0.008	0.922	0.003	0.014	0.014	0.946	0.912	0.872
10	50	0.059	0.015	0.061	1.000	-0.002	0.003	0.004	0.930	0.001	0.008	0.008	0.946	0.920	0.882
10	100	0.048	0.008	0.048	1.000	-0.001	0.001	0.001	0.941	0.001	0.003	0.003	0.952	0.925	0.884
10	200	0.024	0.003	0.024	1.000	-0.000	0.000	0.001	0.940	0.000	0.001	0.001	0.949	0.933	0.882
20	20	-0.002	0.006	0.006	0.931	-0.003	0.006	0.006	0.925	0.001	0.008	0.008	0.944	0.910	0.870
20	50	0.021	0.006	0.022	0.997	-0.001	0.002	0.003	0.938	0.001	0.003	0.003	0.948	0.925	0.882
20	100	0.030	0.005	0.031	1.000	-0.000	0.001	0.001	0.944	0.000	0.001	0.001	0.951	0.929	0.883
20	200	0.025	0.003	0.025	1.000	-0.000	0.000	0.000	0.944	0.000	0.001	0.001	0.949	0.932	0.886
30	30	-0.001	0.003	0.003	0.939	-0.001	0.003	0.003	0.935	0.000	0.004	0.004	0.947	0.922	0.880
30	50	0.008	0.003	0.009	0.985	-0.001	0.002	0.002	0.939	0.000	0.002	0.002	0.949	0.925	0.880
30	100	0.018	0.003	0.018	0.999	-0.000	0.001	0.001	0.941	0.000	0.001	0.001	0.947	0.930	0.880
30	200	0.020	0.002	0.020	1.000	-0.000	0.000	0.000	0.947	0.000	0.000	0.000	0.950	0.936	0.887

Table 2. Accuracy for Downsampled Data

n	m	Nair			Asymptotics			Jackknife			Bayes				
		Bias	SE	RMSE	p_{as}	Bias	SE	RMSE	CI	Bias	SE	RMSE	p_{jack}	p_{CI-B1}	p_{CI-B2}
90	90	-0.012	0.001	0.012	0.720	-0.001	0.003	0.003	0.934	0.001	0.004	0.004	0.949	0.921	0.879
180	90	-0.008	0.001	0.008	0.723	-0.001	0.002	0.002	0.938	0.000	0.002	0.002	0.945	0.925	0.876
300	150	-0.004	0.000	0.004	0.736	-0.000	0.001	0.001	0.947	0.000	0.001	0.001	0.952	0.933	0.883
300	300	-0.003	0.000	0.003	0.734	-0.000	0.001	0.001	0.945	-0.000	0.001	0.001	0.949	0.934	0.885

Table 3 provides, for the same settings as above, the expectations of the variance estimators based on the asymptotic formula and the jackknife. In addition, their ratio is given. On average, the asymptotic formula provides smaller variance estimates than the jackknife, which, however, give too tight confidence intervals with lower than nominal coverage probabilities, see above. When the sample sizes increase, the ratio gets closer to 1, the limiting value due to the consistency of both estimators.

Table 3. Accuracy of Variance Estimators and Confidence Intervals

n	m	Asymptotics	Jackknife	Ratio
10	10	0.045	0.062	0.725
10	25	0.027	0.034	0.797
10	50	0.016	0.019	0.839
10	100	0.009	0.010	0.885
10	200	0.005	0.005	0.915
20	20	0.024	0.028	0.853
20	50	0.014	0.015	0.904
20	100	0.008	0.009	0.937
20	200	0.005	0.005	0.960
30	30	0.016	0.018	0.902
30	50	0.012	0.013	0.925
30	100	0.008	0.008	0.949
30	200	0.004	0.004	0.969

3.2.2. Time series

The following second set of simulations assumes that the measurements form weakly dependent time series and investigates the accuracy of the proposed downsampling approach with a random search to minimize the serial correlations. In order to get insights into the statistical accuracy, a subsampling simulation approach was used where the simulated data sets consist of real blocks of real calibrated measurements to preserve autocorrelations present in the real data, cf. [9] and the references given therein.

For each given sample sizes n and m subsamples of consecutive measurements from the original data sets were drawn, starting at a randomly chosen index. In other words, block bootstrap resampling was applied, a procedure which yields subsamples which contain the serial correlations present in a real time series of length equal to the block size. Those subsamples drawn according to the block bootstrap scheme were, however, not directly used to calculate the statistics. Instead, we applied the proposed data thinning procedure, in order to decorrelate them. Each subsample was downsampled to sample sizes $n' = n/3$ and $m' = m/3$, respectively, at irregularly spaced grid points using the proposed random search optimization algorithm, in order to minimize the autocorrelations up to the lag $r = 8$. The random search algorithm was applied with 1,000 randomly selected sets of strictly ordered grid points. The grid points leading to the minimal value of the objective function (1) among those 1,000 candidate sets was then used as the final simulated thinned data set. Then the variance estimators and confidence intervals were calculated from those thinned series of length n' and m' . Table 2 provides

the results for a couple of sample sizes leading to sample sizes of the downsampled series between 30 and 100. Each table entry is based on 5,000 simulation runs.

The simulation results demonstrate the high accuracy of the jackknife confidence intervals, which perform, for the cases studied here, better than the asymptotic confidence intervals, especially for smaller sample sizes. The Bayesian credible intervals again lead to substantially lower than nominal coverage probabilities indicating that they produce too optimistic, i.e. too narrow intervals.

4. Discussion and Conclusions

Common mean estimators allow to combine data sets collected using different measurement systems such as a flasher at the production line of a manufacturer and a flasher of a testing laboratory, e.g. to determine the power output of PV modules. Similarly, two data sets are available, when data is collected at two time points when two laboratories characterize the same modules. As both the measurement uncertainty in terms of the standard deviation of measurement errors and the distributional shape of the measurement error may differ substantially, the question arises how one can assess the standard error of such a common mean estimator. This is of substantial practical importance and is also needed to set up improved confidence intervals for the true but unknown power output of the PV modules under investigation, which combine the information contained in both series of measurements.

We discuss three statistical approaches to estimate the variance of common mean estimator and construct confidence intervals and two Bayesian models leading to Bayesian credible intervals. Those approaches were evaluated by simulating their performance, focusing on the coverage probability of the confidence resp. credible intervals as the most important statistical property. In the simulations the measurement errors follow real world distributions. Here two simulation approaches were studied, namely i.i.d. sampling assuming the original data sets satisfy the i.i.d. assumption and a nonparametric subsampling procedure with data adaptive data thinning to decorrelate the subsamples. The latter approach assumes that the measurement series are weakly dependent time series and downsamples the series at irregularly spaced time points in such a way that the autocorrelations present in the downsampled series are minimized.

The results show that, for all settings studied here, the jackknife variance estimator is preferable. It leads to accurate estimates of the estimator's standard error and, most importantly, the jackknife outperforms the other approaches in terms of the accuracy of the coverage probability of the corresponding confidence intervals, on which our study focuses. This even holds true for highly unbalanced samples where the sample sizes differ substantially. As a result, our results show that one can improve the efficiency significantly by combining a series of test measurements with an additional data set of observations with larger measurement errors. However, we do not recommend to use the extreme cases investigated in the simulations (such as $n = 20$ and $m = 200$), as any decision including estimates of the true power output should be based on a sufficiently large sample of high-quality test measurements. In addition, we compare our nonparametric proposal with a Bayesian approach where the common mean and the variances are assumed to follow prior distributions. It turns out that the Bayesian credible intervals are noticeably less accurate than the confidence intervals based on the jackknife variance estimator.

Those conclusions apply to both sampling approaches studied in this paper and, especially, demonstrate that the proposed data adaptive downsampling procedure to decorrelate the measurement series

works well and leads to accurate nonparametric confidence intervals for the photovoltaic data studied here. Over all, the jackknife confidence intervals based on the jackknife variance estimator leads to highly accurate confidence intervals and improves upon the asymptotic confidence intervals based on the central limit theorem, especially for smaller sample sizes. The Bayesian models lead to credible intervals, which are too narrow in view of estimation uncertainty inherent to the true underlying distributions, and therefore result in substantially lower than nominal coverage probabilities as evident from the Monte Carlo simulations.

We may conclude that the real-data driven simulations illustrate that the jackknife variance estimation provides an easy-to-use, reliable and powerful approach to determine the precision of the common mean estimator and calculate confidence intervals. The resulting confidence intervals are highly accurate and avoid model errors as they do not impose restrictive assumptions on the distribution of the data.

Acknowledgments

The author would like to thank Dr. Werner Herrmann, TÜV Rheinland Energy GmbH, Cologne, for providing the power output data sets used in this paper. The comments of an anonymous referee are acknowledged.

Conflict of Interest

The author declares no conflict of interest in this paper.

References

1. Chang YT, Steland A (2016) Jackknife variance estimation for common mean estimators under ordered variances and general two-sample statistics. *[preprint]*
2. Nair K (1980) Variance and distribution of the Graybill-Deal estimator of the common mean of two normal populations. *Ann Stat* 8: 212-216.
3. Tukey JW (1958) Bias and confidence in quite large samples. *Ann Math Stat* 29: 614.
4. Shao J, Wu CF (1998) A general theory for jackknife variance estimation. *Ann Stat* 17: 1176-1197.
5. Newman RM, Martin FB (1983) Estimation of fish production rates and associated variances. *Can J Fish Aquat Sci* 40: 1729-1736.
6. Graybill FA, Deal RB (1959) Combining unbiased estimators. *Biometrics* 15: 543-550.
7. Parr WC (1985) Jackknifing differentiable statistical functionals, *J Roy Stat Soc B* 47: 56-66.
8. Shao J, Tu D (1995) *The Jackknife and Bootstrap*. Springer-Verlag, New York.
9. Steland A, Rafajłowicz E (2014). Decoupling change-point detection based on characteristic functions: Methodology, asymptotics, subsampling and application, *J Stat Plan Infer* 145: 49-73.
10. Gregurich MA, Broemeling LD (1989) A Bayesian analysis for estimating the common mean of independent normal populations. *Commun Stat- Simul C*

11. Roberts CP (2007) *The Bayesian Choice: A Decision-Theoretic Motivation*. Springer Texts in Statistics, Springer, New York.
12. Berger JO, Bernardo JM, Sun D (2015) Overall objective priors. *Bayesian Anal* 10: 189-221.
13. Ghosal Y (1996) A review of consistency and convergence rates of posterior distribution. *Proceedings of Varanashi Symposium in Bayesian Inference*, Banaras Hindu University, 26: 35-51.
14. Herrmann W, Althaus J, Steland A, et al. (2006) Statistical and experimental methods for assessing the power output specification of PV modules. *Proceedings of the 21st European Photovoltaic Solar Energy Conference* 2416-2420.
15. Herrmann W, Steland A, Herff W (2010) Sampling Procedures for the Validation of PV Module Output Specification. *Proceedings of the 24th European Photovoltaic Solar Energy Conference, Hamburg* 3540-3547.
16. Herrmann W, Steland A (2010) Evaluation of Photovoltaic Modules Based on Sampling Inspection Using Smoothed Empirical Quantiles. *Prog Photovoltaics* 18: 1-9.
17. Pepelyshev A, Steland A, Avellan-Hampe A (2014) Acceptance sampling plans for photovoltaic modules with two-sided specification limits. *Prog Photovoltaics* 22: 603-611.



©2017, Ansgar Steland, licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)