

# Modified Maximum Entropy Method and Estimating the AIF via DCE-MRI Data Analysis

Zahra Amini Farsani <sup>1,2,\*</sup>  and Volker J. Schmid <sup>2</sup> <sup>1</sup> Statistics Department, School of Science, Lorestan University, Khorramabad 68151-44316, Iran<sup>2</sup> Bayesian Imaging and Spatial Statistics Group, Institute of Statistics, Ludwig-Maximilians-Universität München, Ludwigstraße 33, 80539 Munich, Germany; volker.schmid@lmu.de

\* Correspondence: zahra.farsani@stat.uni-muenchen.de or aminifarsani.z@lu.ac.ir; Tel.: +98-66-3312-0097

**Abstract: Background:** For the kinetic models used in contrast-based medical imaging, the assignment of the arterial input function named AIF is essential for the estimation of the physiological parameters of the tissue via solving an optimization problem. **Objective:** In the current study, we estimate the AIF relayed on the modified maximum entropy method. The effectiveness of several numerical methods to determine kinetic parameters and the AIF is evaluated—in situations where enough information about the AIF is not available. The purpose of this study is to identify an appropriate method for estimating this function. **Materials and Methods:** The modified algorithm is a mixture of the maximum entropy approach with an optimization method, named the teaching-learning method. In here, we applied this algorithm in a Bayesian framework to estimate the kinetic parameters when specifying the unique form of the AIF by the maximum entropy method. We assessed the proficiency of the proposed method for assigning the kinetic parameters in the dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI), when determining AIF with some other parameter-estimation methods and a standard fixed AIF method. A previously analyzed dataset consisting of contrast agent concentrations in tissue and plasma was used. **Results and Conclusions:** We compared the accuracy of the results for the estimated parameters obtained from the MMEM with those of the empirical method, maximum likelihood method, moment matching (“method of moments”), the least-square method, the modified maximum likelihood approach, and our previous work. Since the current algorithm does not have the problem of starting point in the parameter estimation phase, it could find the best and nearest model to the empirical model of data, and therefore, the results indicated the Weibull distribution as an appropriate and robust AIF and also illustrated the power and effectiveness of the proposed method to estimate the kinetic parameters.



**Citation:** Amini Farsani, Z.; Schmid, V.J. Modified Maximum Entropy Method and Estimating the AIF via DCE-MRI Data Analysis. *Entropy* **2022**, *24*, 155. <https://doi.org/10.3390/e24020155>

Academic Editor: Udo Von Toussaint

Received: 6 December 2021

Accepted: 17 January 2022

Published: 20 January 2022

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** kinetic model; modified maximum entropy method; arterial input function; optimization method

## 1. Introduction

Determining the probability density function of a random variable based on observations is a major and old issue in statistics. In recent years, various parametric and non-parametric methods have been introduced for the estimation of the probability density function for a random variable based on observations, but there is very limited work reported on the optimization methods. The maximum entropy method (MEM) is one of the major methods for estimating and determining the probability density with a high level of accuracy and efficiency and minimum bias. It is applied to gain the unknown density via resolution of an optimization problem. The principle of maximum entropy, as a method of statistical inference, is due to Jaynes [1]. His idea is that this principle leads to the selection of a probability density function that is consistent with our knowledge and introduces no unwarranted information. Any probability density function satisfying the constraints that has smaller entropy will contain more information (less uncertainty), and

thus says something stronger than what we are assuming [1–4]. Entropy maximization or related concepts has been frequently utilized in the past ten years to analyze large biological datasets in various fields. These fields range from determining macromolecular interactions and structures [5–20] to inferring signaling [21–25] and regulatory networks [26–28] and the coding organization in neural populations [28–43] based on DNA sequence analyzes (the detection of specific binding sites, for instance) [42–46]. MEM is a powerful vehicle to reconstruct images based on various datasets. It is also commonly used in radio astronomical interferometry, which deals routinely with images with high dynamic range and up to a million pixels [47,48].

In this paper, a concise and basic introduction to entropy maximization and its applicability for deriving models from biological datasets—especially in kinetic models and image processing via DCE-MRI—is provided. DCE-MRI is a fast and noninvasive method for quantitative perfusion analysis in soft tissue, using a contrast agent (CA). It is used extensively for the analysis of microvascular blood in a variety of clinical usages like detecting, characterizing, and therapeutic monitoring of different diseases [49–54]. In DCE-MRI, quantitative analysis is often applied on a whole tumor region of interest (ROI) [55], in which the contrast agent concentration time curve for all voxels in the tumor is used to estimate a single set of kinetic parameters (like—volume transfer constant between blood plasma and extracellular extravascular space per minute—and—rate constant between extracellular extravascular space and blood plasma per minute [56]) for each patient study.

In recent years, several approaches have been proposed to quantify the perfusion of CA into tissues and to estimate the related parameters of perfusion (indices) from concentration-time or signal-curves [57–60]. In DCE-MRI, quantification of the perfusion involves measuring the concentration of the CA in tissue over time. This time curve is then modeled using kinetic processes, where the kinetic parameters are of clinical interest [55]. The kinetic model of the tissue is explicit as an ordinary-differential equation, dissolved analytically result in a nonlinear format in the contrast agent concentration [61–69].

One fundamental necessity for the Tofts pharmacokinetic analysis is the knowledge of the arterial input function (AIF), that is the time curve of CA concentration on the left-side ventricular blood pool. Although the AIF itself is not clinically relevant, its correct determination is very important for the correct estimation of kinetic parameters [62,70]. Since the obtained rate constants are heavily dependent on the AIF [71–75], an accurate and precise measurement is necessary for their absolute and reliable quantification. Instead, a simplified method, such as a population averaged AIF can be used. However, (large) variabilities in cardiac output—between patients and within patients over time—are no longer taken into account with this method. If this variability in cardiac output can be accounted for by precise measurement of the AIF, the accuracy and repeatability of the kinetic parameters should be superior over use of a population averaged AIF. Some researchers have shown that a population averaged AIF can result in better repeatability [76,77], whereas others report the opposite [78,79]. It is possible that repeatability depends on the imaged body part and imaging sequence parameters, but also on the choice of the artery for AIF measurement.

However, in many imaging applications, e.g., for patients with breast cancer, it is not possible to directly measure the AIF from imaging, as no large vessel is in the field of view. Thus, assumed AIFs from the literature are often used, e.g., bi-exponential functions with parameters derived by [80] or [81] or a mix of the two Gaussian with an exponential [61–69,76].

We have previously developed a method for estimating the AIF and the kinetic parameters in DCE-MRI [82]. This method was developed in response to a need in the medical imaging community for the objective comparison of estimations made using different statistical methods, for example, the Bayesian method and MLE. The main problem of our previous algorithm was the dependence of Newton's method on the starting point, which was a uniform random number.

The previous algorithm was the maximum entropy method in combination with Newton's method in order to estimate the AIF using the CA time curve data in plasma,

also called “blind estimation” of the AIF and the maximum a posterior approach (MAP) to determine the kinetic parameters [82]. In this paper, we propose an improved algorithm for blind AIF estimation using a mixture of the maximum entropy method (MEM) and teaching-learning based optimization in the step of  $\lambda$ 's estimation for assessing observer performance in the classification tasks using available information. This circumvents issues with the random start points in the previous algorithm. This approach works on the effect of the influence of a teacher on students. Like other nature-inspired algorithms, TLBO is also a population-based method and uses a population of solutions to proceed to the global solution. The population is considered as a group of learners or a class of learners. The process of TLBO has two parts: the first part consists of the ‘Teacher Phase’ and the second part consists of the ‘Learner Phase’. ‘Teacher Phase’ means learning from the teacher, and ‘Learner Phase’ means learning by the interaction between learners [83,84]. The proposed algorithm is, therefore, more robust. It finally proposed the Weibull distribution in between all other selective models as a model for the AIF via MEM approach. We performed extensive studies using empirical data to better understand the performance of our method. In addition, a comparison was conducted among four other different estimation methods in DCE-MRI dataset, and the new recommended method results were compared with the previous work [55,82].

The rest of this paper is structured as follows: Section 2 describes the basic structure of the proposed method for DCE-MRI analysis: the proposed modified maximum entropy method (MMEM), TLBO algorithm, and MAP. Section 3 includes alternative approaches for estimating of the parameters. Section 4 gives an example to show the step-by-step analysis of the dataset. Section 5 contains the application of a complete DCE-MRI study using the proposed method and their evaluations. Section 6 concludes the aim of the present work.

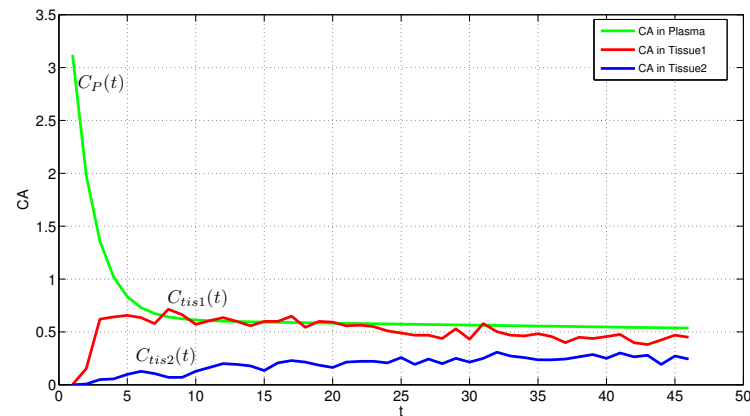
#### Data Description

As an example data set, we use a previously analyzed breast cancer data set [55]. The data were provided by the Paul Strickland Scanner Center at Mount Vernon Hospital in Northwood, UK. Pre-treatment DCE-MRI scans of twelve patients were available. In each case, 46 images were recorded every 11.9 s after administration of the contrast agent Gadolinium-DTPA. For the calculation of  $T_1$  values, we used a two-point measurement with calibration curves as described in [85,86]. In DCE-MRI,  $T_1$  is the relaxation time, also known as the spin-lattice relaxation time. In addition, it is a measure of how quickly the net magnetization vector recovers and its ground state or it is the time constant for regrowth of longitudinal magnetization. The  $T_1$  values are computed as a ratio of a  $T_1$ -weighted fast low-angle shot (FLASH) image and a proton-density-weighted FLASH image.

To measure contrast agent concentration  $C_t(t)$ , the signal intensity is converted to  $T_1$  relaxation time values using  $T_1$ -weighted images, proton density weighted images and data from calibration phantoms with known  $T_1$  relaxation times [87]. The Gd-DTPA concentration can then be computed via  $C_t(t) = \frac{1}{r_1} \left[ \frac{1}{T_1(t)} - \frac{1}{T_{10}} \right]$ , where  $T_{10}$  is the  $T_1$  value without contrast, computed as mean value of the first four images, and  $r_1 = 4.241/\text{s}/\text{mmol}$  is the longitudinal relativity of protons in vivo due to Gd-DTPA. The imaging parameters of the  $T_1$ -weighted FLASH images were  $\text{TR} = 11$  ms,  $\text{TE} = 4.7$  ms,  $\alpha = 35$ , the parameters of the proton density-weighted image were  $\text{TR} = 350$  ms,  $\text{TE} = 4.7$  ms,  $\alpha = 6$ . Field of view was the same for all scans,  $260 \times 260 \times 8$  mm per slice, so voxel dimensions are  $1.016 \times 1.0168$  mm. A scan includes three sequential slices of  $256 \times 256$  voxels and one slice placed in the contralateral breast as control, which we do not use for our analysis. A dose of  $D = 0.1$  mmol per kg body weight of Gd-DTPA was injected after the fourth scan via a power injector with 4 mL/s with a 20 mL saline flush also at 4 mL/s.

Figure 1 shows the empirical model of data for two patients  $\vec{C}_{tis-1}$ ,  $\vec{C}_{tis-2}$  and  $\vec{C}_p$  based on the time  $t$  using Kernel distribution which actually finds an empirical density function of the sample data (See: Using “KernelDistribution” Objects and “ksdensity” in Matlab). In statistics, kernel density estimation is a non-parametric way to estimate the probability density function of a random variable and it is a smoothing function that determines the

shape of the curve used to generate the pdf, and a bandwidth value that controls the smoothness of the resulting density curve [88]. This is the primary model of data  $C_p(t)$  changing the inverse problem to forward. It is not clear the sample data belongs to which family of distributions, it is the main issue in here we try solving here by the maximum entropy method.



**Figure 1.** Empirical PDF of the contrast agent in plasma ( $C_p(t)$ ) named empirical AIF and in tissue ( $C_{tis}(t)$ ) for two patients.

## 2. Theory and Methods

In this study, we estimated different probability density functions for AIF using our proposed algorithm. The accurate estimated model of AIF is the cornerstone of the present work, because its correct determination is very important for the correct estimation of kinetic parameters. In the previous study [82], we have estimated the gamma and exponential distributions using the maximum entropy method and Newton's approach. Both models were acceptable approximated AIF in comparison to the literature models.

Here, the modified algorithm examines different moments constraints in MEM to build the best probability model fit to data. In addition, besides TLBO, the various parameter estimation methods then suggest the estimated parameters which help to find more appropriate AIF. Kinetic parameters are estimated in the next step via MAP.

### 2.1. Kinetic Model

The kinetic process in the tissue can be modeled using an ordinary-differential equation, dissolved analytically resulting in a nonlinear model for the contrast agent concentration [61–69]. In this study, we adopted the commonly used pharmacokinetic model [89] that assumes the CA resides in and exchanges between two compartments in the tissue: the vascular space and EES.

Considering the kinetic properties of the contrast agent (CA) in the tissue of interest ( $C_{tis}$ ) using DCE-MRI, we apply the differential equation system as follows:

$$\frac{dC_{tis}(t)}{dt} = K_1 C_p(t) - K_2 C_{tis}(t), \quad (1)$$

in which  $C_p(t)$  is the CA concentrations in the vascular blood pool, that is, the arterial input function AIF. Both  $K_1$  and  $K_2$  are the rate constants of the CA exchanges between extravascular-extracellular space (EES) and plasma. Subject to  $C_p(0) = 0$ , Equation (1) can be solved with the following result

$$C_{tis}(t) = K_1 \int_0^t C_p(u) e^{-K_2(t-u)} du. \quad (2)$$

This equation has been used to analyze MR data in a number of studies [81]. Murase [90] proposed a different way to solve Equation (1) using discretization:

$$C_{tis}(t) = K_1 \int_0^t C_p(u)du - K_2 \int_0^t C_{tis}(u)du, \quad (3)$$

This can be written in matrix form as follows:

$$\vec{C} = \vec{A} \times \vec{K}, \quad (4)$$

where matrix  $A_{n \times 2}$ :  $\vec{A} = \{A(1), \dots, A(n)\}'$  for each row of  $I = 1, 2, \dots, n$ :

$$A(I) = (\int_0^{t_I} C_p(u)du, - \int_0^{t_I} C_{tis}(u)du), \quad (5)$$

$$\vec{K} = \begin{pmatrix} K_1 \\ K_2 \end{pmatrix} \quad (6)$$

and

$$\vec{C} = \begin{pmatrix} C_{tis}(t_1) \\ C_{tis}(t_2) \\ \vdots \\ C_{tis}(t_n) \end{pmatrix}. \quad (7)$$

From the mathematical view, when  $C_{tis}(t_i)$  and  $C_p(t_i)$  are measured, it is possible to use conventional linear least-squares (LLSQ) method to determine  $\vec{K}$  and the trapezoidal rule for the elements of  $\vec{A}$ . Unfortunately, this method measures approximate values for the kinetic parameters. A great number of image processing problems can be presented as inverse problems. In here the linear system of equations which are obtained after the discretization of the integral equations which arises in various tomographic image restoration and reconstruction concerns are considered: Therefore, we write Equation (2) as follows:

$$y_{tis}(t_i) = A(i)\vec{K} + \varepsilon_i \quad \varepsilon_i \sim N(0, \sigma^2) \quad (8)$$

where  $y_{tis}(t_i)$  is the observed tissue concentration at time  $t_i$  and the measurement uncertainty (noise) which is assumed to be additive, centered, white, Gaussian and independent of  $K$  [91].

For that, the estimation procedure of the analysis here includes Bayesian methods, which its advantage is to overcome the integration process involved in estimating the model parameters. Therefore, Bayesian methods can provide exact estimates of the model parameters, not approximating them. In Bayesian statistics, parameters are viewed as random variables. Each parameter involved in a Bayesian model has a distribution attached to it in order to express the uncertainty about its true value. The distribution is named the prior distribution. They represent the prior knowledge about the parameter of interest, which is often obtained from historical data (data-based priors) [92,93].

## 2.2. Maximum a Posterior Approach

Image reconstruction belongs to the class of ill-posed inverse problems of mathematical physics [94]. In 1967, the physicist V. Turchin suggested using the Bayesian method of maximum a posteriori (MAP) for solving inverse ill-posed problems with stochastic data, naming this approach 'statistical regularization' [95]. Bayesian maximum a posteriori (MAP) approaches can be used to solve ill-posed problems as they arise in image reconstruction [96,96,97]. The solution of MAP obviously depends on the priori models. The main challenge of the Bayesian method is how to determine the a priori probability distribution of the studied image and specify its parameters using its data. In here, we assume a form of a priori information named entropy-based prior, which relies on the principle of entropy. Such approaches have been successfully used in the fields of plasma-tomography, X-ray, radio, and gamma-astronomy [93,96–98]. To estimate the kinetic parameters, we consider the general form of Equation (8) in the following as proposed by [92,93]. Estimating the positive-vector  $x$  (the pixel-intensities in an object) subject to a vector of measurement  $y$

(e.g., a degraded-image or the projections of an object) and a linear-transformation  $A$  which relates both vectors by

$$y = Ax + b, \tag{9}$$

where  $b$  is uncorrelated noise with normal distribution, and zero-mean. We consider only approximate information about the variance of the noise  $\Sigma^2$  and general information about the object.

To estimate the unknown vector  $\vec{x}$ , we use a Bayesian approach. Given the probability density functions (pdf)  $f(x)$  and  $f(y|x)$  and  $f(y)$  we obtain the pdf of the conditional distribution of  $x$  subject to  $y$  using the Bayesian formula [99]:

$$f(x|y) = \frac{f(y|x) \cdot f(x)}{f(y)}. \tag{10}$$

The MAP estimator  $\hat{x}$  maximizes the posterior pdf  $f(x|y)$  obtained by Bayes formula. In the Equation (10),  $f(y)$  is independent of  $x$ ,  $f(y|x)$  relates to the noise probability distribution and finally,  $f(x)$  is a prior distribution on  $x$ .

If we are not able to directly determine  $f(x)$  and  $f(y|x)$ , we can apply the maximum entropy method. For the MEM, knowledge of some constraints on  $f(x)$  can be used. Among all probability distributions satisfying these constraints, we select the one which has maximum entropy [92,92,93], see Section 2.3. To determine  $f(y|x)$ , the noise pdf, we have

$$\begin{aligned} f(y|x) &\approx \exp[-T(x)], \\ T(x) &= [y - Ax]^t [y - Ax] / \sigma^2. \end{aligned}$$

A possible way to select a priori-distribution,  $f(x)$ , is to apply the MEM where the general model is in the form of the exponential family. The advantage of applying MEM for finding a priori is, that this method is the most objective, and maximally uncommitted [99].

### 2.3. Maximum Entropy Method

The maximum entropy principle allows one to determine the least biased probability distribution function when the information available is limited by some macroscopic constraints [100]. The MEM determines the randomness of the primary data by the concept of information entropy. It is the mathematical expectation of the uncertainty associated with an outcome in terms of its occurrence probability. It is suggested that the most likely probability distribution should be the one that maximizes the information entropy. Maximizing Shannon’s entropy is the basic idea of MEM [4]:

$$h(X) = - \int f(x) \log f(x) dx, \tag{11}$$

subject to known constraint, typically moment constraints

$$E(\phi_k(x)) = \int \phi_k(x) f(x) dx = \mu_k. \tag{12}$$

Here  $\phi_0(x) = 1$ , and  $\phi_k(x)$ ,  $k = 0, \dots, N$  are  $N + 1$  known-functions. These could be, for example,  $x^n, \log(x)$ ,  $x \log(x)$  or trigonometric or geometric functions. The idea is to assign the appropriate known-function relating the ME distribution to the exponential family via the mentioned constraints [2,3,101]. Using the method of Lagrange multipliers method, where the objective function is Shannon’s entropy Equation (11),  $J(f)$  is as follows:

$$J(f) = - \int f(x) \log f(x) dx + \lambda_0 \int f(x) dx + \sum_{k=1}^N \lambda_k \int f(x) \phi_k(x) dx. \tag{13}$$

For obtaining  $f(x)$ , we differentiate  $J$  subject to  $f(x)$ :

$$\frac{\Delta J(f)}{\Delta f(x)} = -\log f(x) - 1 + \lambda_0 + \sum_{k=1}^N \lambda_k \phi_k(x) = 0. \tag{14}$$

Adopting Taylor’s theorem (based on using a Taylor series approximation), the required expected values  $\mu_1, \dots, \mu_m$  can be obtained numerically from the data set [101]. Applying an optimization method with Shannon’s entropy as the objective function, and setting Equation (14) equal to zero, the general form of ME distribution will be [4]:

$$f(x) = e^{-\sum_{k=0}^N \lambda_k \phi_k(x)}, \quad x \in S, \tag{15}$$

where  $\lambda_k$  can be selected such that,  $f(x)$  in Equation (15) satisfies the known constraints in Equation (12). The parameters  $\lambda = [\lambda_0, \dots, \lambda_N]$  are determined to fit an appropriate class of the ME distributions. To determine the  $N + 1$  unknown-parameters, the set of  $N + 1$  nonlinear equations are solved as follows ( $1 \leq k \leq m$ ):

$$G_k(\lambda) = \int \phi_k(x) e^{-\sum_k \lambda_k \phi_k(x)} dx = \mu_k. \tag{16}$$

#### 2.4. Teaching-Learning Based Optimization

To solve Equation (13) we suppose a population base algorithm named teaching-learning based optimization (TLBO), which resolves the problem of random starting points. Instead, it can measure the mean of all possible parameter estimations in order to fit a better model to data. The idea of this algorithm is to assume the relationships between a teacher and all learners in a classroom [83]. It considers the same probability for learners to get information from others. TLBO has two steps to find the best solutions, which named teaching and, learning. The most important features of this algorithm is the easy implementation, and rapid convergence [102].

It considered the population as a group of students and their related subjects are the design variables of the problem. In TLBO, the different presented courses for students are supposed to be different variables and their scores are like the objective function. It is very important for teachers to share their knowledge between students to improve the class’s level of knowledge, then it can cause achieving acceptable scores by students based on their talents. The teacher is assumed to be the most informed person in the classroom, which distributes her/his knowledge to the students. For that, the teacher will be the best solution (the best person in the population) among all. There is a fact that the student’s level of knowledge highly depends on the teacher’s quality of the teaching, and the quality of the others in the class. Therefore, the two main steps of learning for the students are implemented After the creation of the initial population and calculation of the objective value for each individual as follows:

##### 2.4.1. Teacher Phase

In this step, the teacher tries to improve the mean scores of the students subject to her/his situation. The random procedure here is to produce a new solution instead of the old one:

$$X_{new,D} = X_{old,D} + r(X_{teacher,D} - T_f M_D) \tag{17}$$

where,  $D$  shows the number of courses,  $X_{old,D}$  (a vector  $1 \times D$ ) is the old solution, when there is no contribution between the students to improve their knowledge, and it includes the results of each specific course, a random number  $r$  is in  $[0, 1]$ ,  $X_{teacher,D}$  is the best solution of the whole population,  $T_f$  is a teaching factor changes from 1 to 2 randomly with the same probability, and  $M_D$  is a vector ( $1 \times D$ ) involving the mean values of the classroom results for each specific course. The new solution  $X_{new,D}$  is accepted if it is better than the old one [102].

### 2.4.2. Learner Phase

To improve the knowledge of each student when randomly cooperating with other students, Equation (18) is implemented to all of them, so that, one can achieve the new information in the situation when the other learner has more knowledge than her/him.

$$X_{new,i} = X_{old,i} + r_i(X_j - X_k) \quad (18)$$

where  $i = 1, 2$  is the total number of solutions,  $X_{old,i}$  when there is no cooperation with other students,  $r_i$  is a random number in the range  $[0, 1]$ , and  $X_j$  and  $X_k$  are two learners randomly chosen with  $j \neq k$  and in which  $X_j$  presents a better objective value than  $X_k$ . The solution  $X_{new,i}$  is accepted if it is better than the old solution  $X_{old,i}$ .

### 2.5. Implementation

We have implemented the code for the proposed method in MATLAB. These are the main steps of the algorithm:

- (1) Determining  $\phi_k(x)$  and their numerical expectations using dataset via Taylor's theorem [4],
- (2) Using TLBO (or an alternative optimization method, see below) to determine the unknown function with the Shannon's entropy as target function. The general form is given in Equation (15), ( $f_{C_{tis}}(t)$  and  $f_{C_p}(t)$ ),
- (3) Applying the proposed method to find  $\lambda_k$  in which  $f(x)$  (Equation (15)) matches the constraints in (Equation (12)), ( $\hat{f}_{C_{tis}}(t)$  and  $\hat{f}_{C_p}(t)$ ),
- (4) Estimating the kinetic parameters  $\vec{K}$ , we replace  $\hat{f}_{C_{tis}}(t)$  and  $\hat{f}_{C_p}(t)$  in Equation (8) and resolve them via MAP,
- (5) Using the Kullback–Leibler divergence  $D_{K-L}(f||g)$  to check the accuracy of the estimated AIF,  $\hat{f}_{C_p}(t)$  in comparison with the empirical distribution of dataset  $g(C_p)$ ,

$$D_{K-L}(\hat{f}||g) = \int_s \hat{f}_{C_p}(t) \log \frac{\hat{f}_{C_p}(t)}{g(C_p)} dt. \quad (19)$$

- (6) With the predicted values  $\hat{x}_1, \dots, \hat{x}_m$  and the observed values  $x_1, \dots, x_m$ :

$$RMSE = \left[ \frac{1}{m} \sum_{i=1}^m (x_i - \hat{x}_i)^2 \right]^{1/2}, \quad (20)$$

$$\chi^2 = \frac{\sum_{i=1}^m (x_i - \hat{x}_i)^2}{m - n}, \quad (21)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (x_i - \hat{x}_i)^2}{\sum_{i=1}^N (x_i - \bar{x})^2}, \quad (22)$$

## 3. Alternative Parameter Estimation

### 3.1. Weibull Distribution

The Weibull distribution is widely used in reliability and life data analysis due to its versatility. It is also established a close approximation to the probability laws of various natural-phenomena. The pdf of the Weibull distribution has two parameters:

$$f(x) = \frac{k}{c^k} x^{k-1} e^{-\left(\frac{x}{c}\right)^k}, \quad (23)$$

where  $k$  is the shape and  $c$  is the scale parameter [103,104].

Using the Weibull model, we can also use different approaches to determine the parameters  $k$  and  $c$ :



### 3.2. Methods of Moments

The method of moments (MM) calculates the first and second moments to estimate shape and scale parameters. The algorithm relies on the mean, variance of the gamma-function of  $(1 + 1/k)$  [105]. The sample mean and standard error are

$$\bar{x} = c\Gamma(1 + 1/k), \sigma = c(\Gamma(1 + 1/k) - \Gamma^2(1 + 1/k))^{1/2}, \tag{24}$$

where  $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}dt$  is the gamma function.

### 3.3. Empirical Measurement Method

The empirical measurement method is a special case of MM [105–107]:

$$k = \left(\frac{\sigma}{\bar{x}}\right)^{-1.086}, c = \frac{\bar{x}}{\Gamma(1 + \frac{1}{k})}, \tag{25}$$

where  $\sigma$  is the sample standard deviation.

### 3.4. Maximum Likelihood Method

For the maximum likelihood estimator (MLE), an iterative algorithm can be used. With  $n$  the number of non-zero data points, the shape and scale parameters  $(k, c)$  are by iteratively solving

$$k = \left(\frac{\sum_{i=1}^n x_i^k \ln(x_i)}{\sum_{i=1}^n x_i^k} - \frac{\sum_{i=1}^n \ln(x_i)}{n}\right)^{-1},$$

and

$$c = \left(\frac{1}{n} \sum_{i=1}^n x_i^k\right)^{1/k},$$

### 3.5. Modified Maximum Likelihood Method

When the data is available in the form of the frequency distribution, we can apply the modified maximum likelihood method (MMLE).

$$k = \left(\frac{\sum_{i=1}^n x_i^k \ln(x_i)P(x_i)}{\sum_{i=1}^n x_i^k P(x_i)} - \frac{\sum_{i=1}^n \ln(x_i)P(x_i)}{P(x \geq 0)}\right)^{-1}, \quad c = \left(\frac{1}{P(x \geq 0)} \sum_{i=1}^n x_i^k P(x_i)\right)^{1/k}, \tag{26}$$

where  $P(x_i)$  represents the frequency of data  $x_i$ ,  $n$  the number non-zero data points, and  $P(x \geq 0)$  the probability of the random variable equal or exceeding zero. In Equation (26),  $k$  can be resolved iteratively, then  $c$  can be solved explicitly [108,109].

### 3.6. Non-Linear Least Squares Method

For the non-linear least squares method (NLSM) the observations are ordered in an ascendant form, and coupled to the failure probabilities, gained by the estimators. The Gauss–Newton’s algorithm is used to gain the best-fitted curve of a Weibull model [110].

Based on the individual results by the method of maximum likelihood, modified maximum likelihood, and least-squares regression, Seguro and Lambert [111] concluded that maximum likelihood or modified maximum likelihood proposed here provided more reasonable and accurate values for parameter estimation of Weibull distribution than least-squares regression. Later it is proven once again by Cook [112].

## 4. Example of Application

In this section, we show the application of the modified maximum entropy method (MMEM). Due to the limitation of length, this paper only provides a brief description of formulas and fitting curves.

An additional challenge is the fact that usually there are an infinite number of statistical models that are consistent with a given set of global properties measured from data.

Therefore, one needs an additional criterion to decide which model to use. The benefit of using the maximum entropy method is to find the simplest model with the lowest bias, which maximizes the entropy. For the same dataset, there could be many complicated models to describe data, but finding the best fitted model with a few numbers of constraints, a few steps of computations, and finally, with the known family of distribution, is the advantage of choosing this method. For that, the maximum entropy method, searches, examines and, applies different moment constraints (see Section 2.3) [3], and adopts the minimum number of them to form an appropriate probability density model for the sample data. There could be a large difference between the maximum entropy estimated model with two constraints and three constraints, but for the estimated ME model with four or five cases, it would not be a big difference in the estimated models. In other words, there is no guarantee to estimate the better maximum entropy model when adding more constraints.

After examining several known functions [3] and their exceptions (constraints) which are computed numerically based on Taylor’s theorem from the sample data, the estimated probability density function of the data fits well to the Weibull distribution (Figure 2). For data  $\vec{C}_p(t)$

$$\begin{aligned} \int_t f_{C_p}(t) dt &= 1, \\ \int_t \log(C_p) f_{C_p}(t) dt &= -0.4465, \\ \int_t C_p^3 f_{C_p}(t) dt &= 1.0930 \end{aligned}$$

using the general form of the maximum entropy distribution Equation (15),  $\hat{f}_{C_p}(t)$  can be as follows:

$$f_{C_p}(t) = e^{-\lambda_0 - \lambda_1 \log(C_p(t)) - \lambda_2 C_p^3(t)},$$

where the final ME multipliers  $\lambda$ 's and the Weibull parameters are estimated as follows

$$\hat{f}_{C_p}(t) = \exp(-0.7466 - 1.4944 \log(C_p(t)) - 0.1128 C_p^3(t)) + 0.5. \tag{27}$$

and based on the ME form of Equation (23)

$$f(x) = e^{\log(\frac{k}{c^k}) + (k-1) \log(x) - (\frac{x}{c})^k}.$$

in which

$$\lambda_0 = -\log(\frac{k}{c^k}), \tag{28}$$

$$\lambda_1 = -(k - 1), \tag{29}$$

$$\lambda_2 = c^{-k}. \tag{30}$$

Then, according to Equations (27) and (28), the Weibull parameters will be  $c = 1.8498$ ,  $k = 3$  where the mean of absolute error,  $D_{K-L}$  divergence and entropy are 0.0470, 0.0438, and 0.2026, respectively, see Figure 2.

Table 1 lists the mean absolute error (MAE), Kullback–Leibler distance  $D_{K-L}$  and the entropy of different AIF models via MMEM described in Section 2.3 and the empirical one. In each case, we have applied evaluation methods to check the validity of the estimated model. The high measurement of entropy shows the superiority of the Weibull probability density function to fit the data, too. The MMEM can not optimize the different values of  $k$  and  $c$  by itself (based on the uniqueness of the maximum entropy distribution [4]), but it can be applied on a grid on  $k$ . However, using MAE,  $D_{K-L}$  and entropy, we achieve different optimal models with the optimal values of their parameters. Additionally, Figure 3 pictures the fit for the different CDFs.

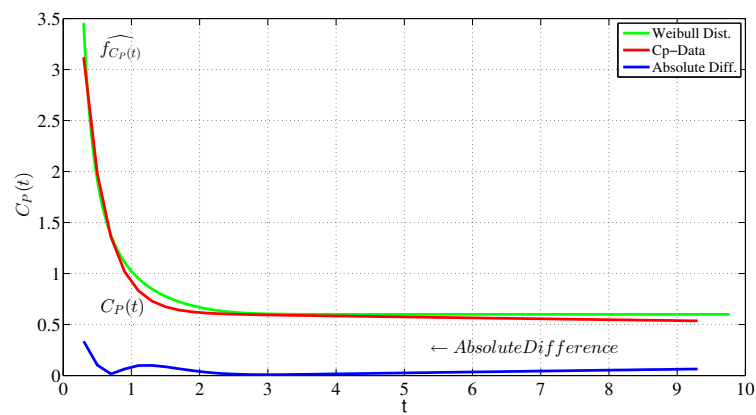


Figure 2. Maximum entropy probability density function of AIF ( $\hat{f}_{C_p}(t)$ ) and empirical AIF.

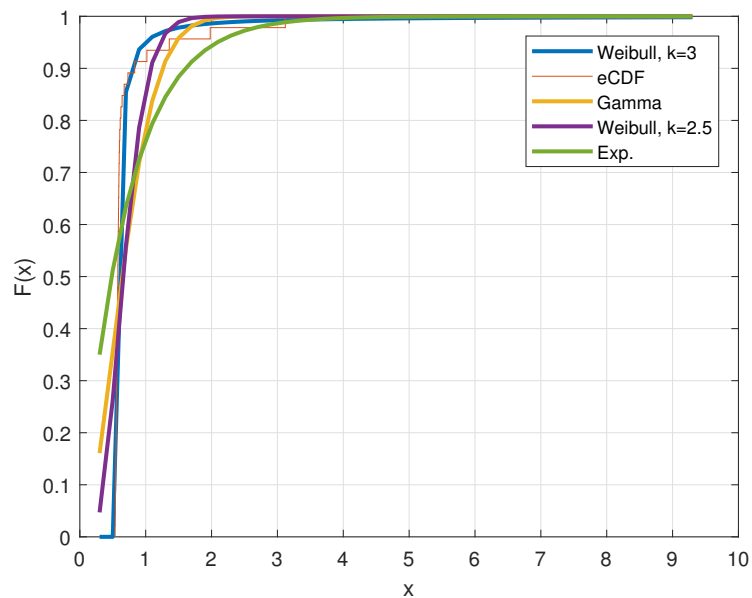


Figure 3. Maximum entropy distribution of AIF, empirical CDF of data and gamma and exponential CDFs.

Table 1. Comparison of the estimated maximum entropy AIFs, and the empirical AIF.

Estimated Distribution	MAE	$D_{K-L}$	Entropy
Gamma	0.0775	0.0285	0.0303
Exponential	0.0375	0.0363	0.0872
Weibull ( $k = 3$ )	0.0470	0.0438	0.2026
Weibull ( $k = 2.6$ )	0.0403	0.0389	0.1755
Weibull ( $k = 2$ )	0.0471	0.0342	0.1471

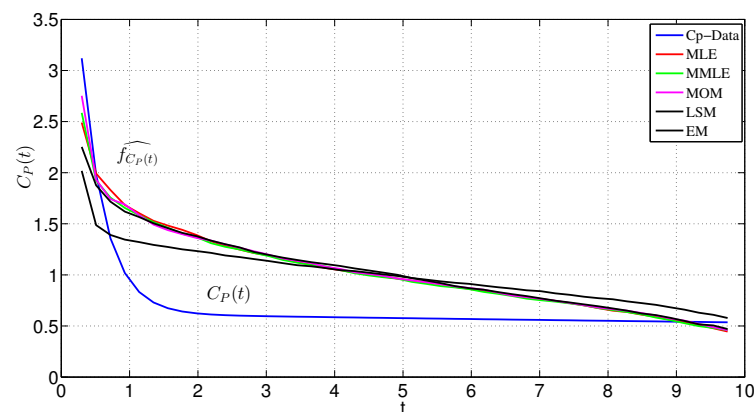
Table 2 lists the estimated Weibull parameters via different estimation methods as mentioned in Section 3. All the estimated models are presented in Figure 4. Table 3 indicates the evaluation measurements for all the mentioned method in Section 3 to investigate how the proposed method works. Among all these models, the MMEM has the best fit to the data.

**Table 2.** Weibull parameters via different parameter estimation methods.

Methods	K	C
EM	1.6469	0.7787
MOM	1.9125	0.7850
MLE	1.8005	0.7890
MMLE	2.0201	0.7758
NLSM	2.7767	0.7518
MMEM	2.6	1.7380

**Table 3.** Evaluating methods to compare the empirical AIF and the maximum entropy PDFs of AIF.

Methods	RMSE	Chi-Square	R <sup>2</sup>	Adjust R <sup>2</sup>
EM	0.286	0.0755	0.631	0.622
MOM	0.255	0.0691	0.670	0.663
MLE	0.278	0.1191	0.570	0.580
MMLE	0.274	0.0771	0.636	0.628
NLSM	0.194	0.2854	0.535	0.525
MMEM	0.0320	7.5687 × 10 <sup>-4</sup>	0.995	0.995



**Figure 4.** Modified maximum entropy PDF of AIF, empirical AIF, maximum entropy PDFs combined with parameter estimation methods.

Based on the results of Tables 1–3 MMEM has achieved a much better fit model to the data. Actually, Table 2 shows the parameter estimations of the Weibull distribution via different methods in comparison to those via MMEM to see in which case the estimated model fits well to the data. In (Table 3), we examined the results via the root mean square error (RMSE), the goodness of fit ( $\chi^2$ ), determination coefficient ( $R^2$ ) and the adjust determination coefficient ( $R^2$ ) which highlights the MMEM. The proposed MMEM gives the estimation with the lowest absolute error and  $D_{K-L}$  divergence with the highest entropy.

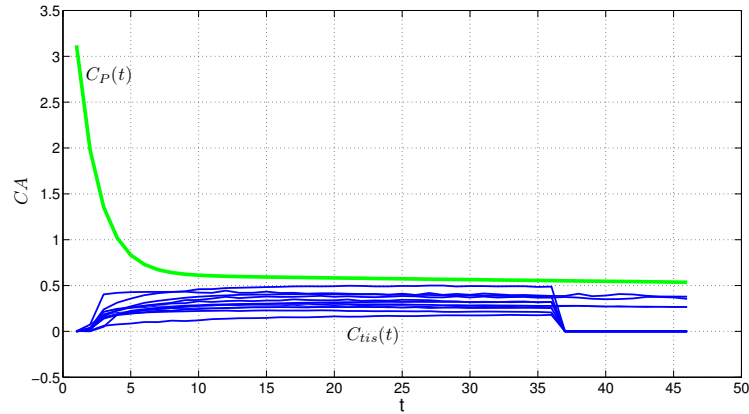
### 5. Evaluation

To better evaluate the desired method (MMEM), we considered the data of 12 more patients in total. MMEM and MAP were utilized to estimate the AIF and the kinetic parameters, and the results were compared accordingly, (Figures 5 and 6). The difference within empirical AIF and estimated AIF via the MMEM is clear.

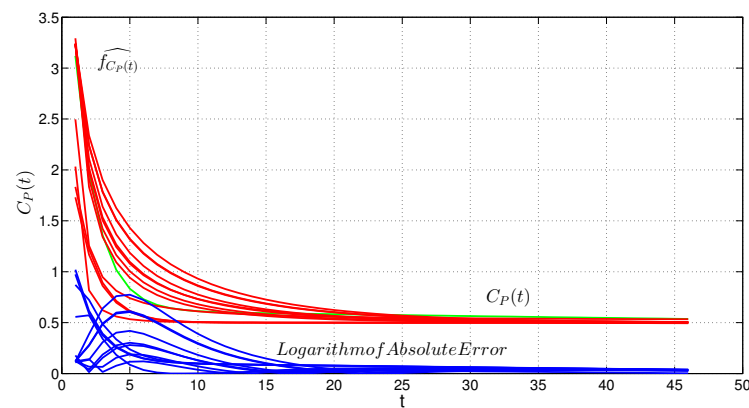
Actually, the AIF in the first two minutes is typically estimated higher than the assumed AIF, whereas there would be negligible difference after about two and a half

minutes. However, the correct estimation of the AIF at the onset is the most important for the correct estimation of the kinetic parameters. The K-L divergence measurements range from 0.001 to 0.0637 for all patients.

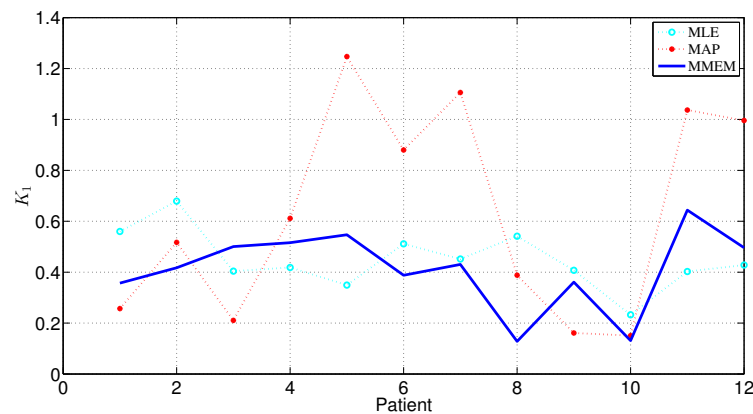
Figure 7 depicts the estimated  $k_1$  values using MMEM/MAP and assumed AIF/ML & MEM/MAP for all 12 patients. For the MMEM, the  $k$  values (Table 4) are more on the same level between patients—compared to the MEM/MAP—which biologically makes sense. Still, the AIF is estimated from the data, which makes the estimation of  $k$  more realistic than the estimation using an assumed AIF.



**Figure 5.** Empirical density model of the contrast agent in plasma ( $C_p(t)$ ) as AIF and tissue ( $C_{tis}(t)$ ) for 12 patients.



**Figure 6.** Estimating AIF via MMEM for 12 patients and empirical AIF.



**Figure 7.**  $k_1$  estimated using MMEM/MAP and assumed AIF/ML & MEM/MAP for all 12 patients.

**Table 4.** Kinetic parameters estimation via MMEM/MAP for 12 patients.

Patient	1	2	3	4	5	6
$k_1$	0.1637	0.1016	0.7175	0.1650	0.5959	1.0477
$k_2$	0.0210	0.3688	0.1073	0.2079	0.1233	0.0072
Patient	7	8	9	10	11	12
$k_1$	0.6309	0.7980	0.1085	0.4327	0.544	1.0225
$k_2$	0.0701	0.3861	0.2377	0.0839	0.235	0.0271

## 6. Discussion and Conclusions

The main purpose of this study is to connect an important problem in statistics which is to determine the probability density function of a random variable based on observations to an important problem in image processing which is to determine the AIF in situations where not enough information about the AIF is available and then accurate estimation of the kinetic parameters. In recent years, various parametric and non-parametric methods have been introduced for estimation of the probability density function for a random variable based on observations, but there is very limited work reported on the optimization methods.

Therefore, we have introduced a new algorithm which is the combination of MEM/TLBO named MMEM. The maximum entropy method (MEM) is one of the major and strong methods for estimating and determining the probability density with a high level of accuracy and efficiency and minimum bias. The core idea of this approach is to determine the statistical models agree with data. In other words, the MEM provides a method to find the least biased model that is consistent with the data, i.e., the maximally noncommittal with regard to missing information [1–4].

A number of calculations and comparisons have been conducted for the estimation of the AIF and the kinetic parameters, respectively, (Figures 5–7 and Tables 1–4). The results have revealed the characteristics of the empirical PDF of AIF as well as the fact that the modified maximum entropy approach performs adequately in fitting an appropriate model by comparing with the Weibull distributions to the AIF with consideration of its accuracy and applicability.

The aim of this work is not exactly to make decisions on the AIF or  $C_p$ , but it is very important how exactly the accurate kinetic parameters are determined. These parameters are so significant and are of clinical relevance, for that, clinicians can easily interpret them. Using the MMEM/MAP guarantees to minimize bias in the estimation of the AIF and the kinetic parameters

Since the AIF plays an important role in the analysis of DCE-MRI, in cases where the AIF could not be determined in the image, the literature AIF is a standard technique. The proposed method in this study gives an alternative way to assess the input function from the existing data. We have shown that the proposed method allows a good fit of the data and a good estimation of the kinetic parameters.

To further evaluate this method, we propose to apply it on terms of energy efficiency and system complexity.

**Author Contributions:** Supervision, V.J.S.; Writing—review & editing, Z.A.F. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Data from this study were acquired in accordance with the recommendation given by Leach et al., 2005, Br. J. Cancer, 2005, vol. 92, pp. 1599–1610. Informed consent was obtained from all patients.

**Data Availability Statement:** Patient data are not made available.

**Acknowledgments:** The authors thank the institute of Statistics, Ludwig-Maximilian university (LMU) of Munich, for the financial support of this publication.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Jaynes, E.T. Information Theory and Statistical Mechanics. *Phys. Rev.* **1957**, *106*, 620–630. [[CrossRef](#)]
2. Pougaza, D.B.; Djafari, A.M. Maximum Entropy Copulas. *AIP Conf. Proc.* **2011**, *1305*, 2069–2072.
3. Ebrahimi, N.; Soofi, E.S.; Soyer, R. Multivariate maximum entropy identification, transformation, and dependence. *J. Multivar. Anal.* **2008**, *99*, 1217–1231. [[CrossRef](#)]
4. Thomas, A.; Cover, T.M. *Elements of Information Theory*; John Wiley: Hoboken, NJ, USA, 2006.
5. Cofré, R.; Herzog, R.; Corcoran, D.; Rosas, F.E. A comparison of the maximum entropy principle across biological spatial scales. *Entropy* **2019**, *21*, 1009. [[CrossRef](#)]
6. Jaynes, E.T. *Probability Theory: The Logic of Science*; Cambridge University Press: Cambridge, UK, 2003.
7. Ozer, H.G. Residue Associations in Protein Family Alignments. Ph.D. Thesis, The Ohio State University, Columbus, OH, USA, 2008.
8. Seno, F.; Trovato, A.; Banavar, J.R.; Maritan, A. Maximum entropy approach for deducing amino acid interactions in proteins. *Phys. Rev. Lett.* **2008**, *100*, 078102. [[CrossRef](#)]
9. Weigt, M.; White, R.A.; Szurmant, H.; Hoch, J.A.; Hwa, T. Identification of direct residue contacts in protein–protein interaction by message passing. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 67–72. [[CrossRef](#)]
10. Pitera, J.W.; Chodera, J.D. On the use of experimental observations to bias simulated ensembles. *J. Chem. Theory Comput.* **2012**, *8*, 3445–3451. [[CrossRef](#)] [[PubMed](#)]
11. Hopf, T.A.; Colwell, L.J.; Sheridan, R.; Rost, B.; Sander, C.; Marks, D.S. Three-dimensional structures of membrane proteins from genomic sequencing. *Cell* **2012**, *149*, 1607–1621. [[CrossRef](#)]
12. Roux, B.; Weare, J. On the statistical equivalence of restrained-ensemble simulations with the maximum entropy method. *J. Chem. Phys.* **2013**, *138*, 02B616. [[CrossRef](#)] [[PubMed](#)]
13. Cavalli, A.; Camilloni, C.; Vendruscolo, M. Molecular dynamics simulations with replica-averaged structural restraints generate structural ensembles according to the maximum entropy principle. *J. Chem. Phys.* **2013**, *138*, 03B603. [[CrossRef](#)]
14. Jennings, R.C.; Belgio, E.; Zucchelli, G. Does maximal entropy production play a role in the evolution of biological complexity? A biological point of view. *Rendiconti Lincei Scienze Fisiche e Naturali* **2020**, *31*, 259–268. [[CrossRef](#)]
15. Ekeberg, M.; Lövkvist, C.; Lan, Y.; Weigt, M.; Aurell, E. Improved contact prediction in proteins: Using pseudolikelihoods to infer Potts models. *Phys. Rev. E* **2013**, *87*, 012707. [[CrossRef](#)]
16. Boomsma, W.; Ferkinghoff-Borg, J.; Lindorff-Larsen, K. Combining experiments and simulations using the maximum entropy principle. *PLoS Comput. Biol.* **2014**, *10*, e1003406. [[CrossRef](#)] [[PubMed](#)]
17. Zhang, B.; Wolynes, P.G. Topology, structures, and energy landscapes of human chromosomes. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 6062–6067. [[CrossRef](#)] [[PubMed](#)]
18. Cesari, A.; Reißer, S.; Bussi, G. Using the maximum entropy principle to combine simulations and solution experiments. *Computation* **2018**, *6*, 15. [[CrossRef](#)]
19. Farré, P.; Emberly, E. A maximum-entropy model for predicting chromatin contacts. *PLoS Comput. Biol.* **2018**, *14*, e1005956. [[CrossRef](#)]
20. D'haeseleer, P.; Liang, S.; Somogyi, R. Genetic network inference: From co-expression clustering to reverse engineering. *Bioinformatics* **2000**, *16*, 707–726. [[CrossRef](#)] [[PubMed](#)]
21. Lezon, T.R.; Banavar, J.R.; Cieplak, M.; Maritan, A.; Fedoroff, N.V. Using the principle of entropy maximization to infer genetic interaction networks from gene expression patterns. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 19033–19038. [[CrossRef](#)]
22. Dhadialla, P.S.; Ohiorhenuan, I.E.; Cohen, A.; Strickland, S. Maximum-entropy network analysis reveals a role for tumor necrosis factor in peripheral nerve development and function. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 12494–12499. [[CrossRef](#)] [[PubMed](#)]
23. Remacle, F.; Kravchenko-Balasha, N.; Levitzki, A.; Levine, R.D. Information-theoretic analysis of phenotype changes in early stages of carcinogenesis. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 10324–10329. [[CrossRef](#)]
24. Sanguinetti, G.; Huynh-Thu, V.A. Gene regulatory network inference: An introductory survey. In *Gene Regulatory Networks*; Springer: New York, NY, USA, 2019; pp. 1–23.
25. Locasale, J.W.; Wolf-Yadlin, A. Maximum entropy reconstructions of dynamic signaling networks from quantitative proteomics data. *PLoS ONE* **2009**, *4*, e6522. [[CrossRef](#)]
26. Graeber, T.; Heath, J.; Skaggs, B.; Phelps, M.; Remacle, F.; Levine, R.D. Maximal entropy inference of oncogenicity from phosphorylation signaling. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 6112–6117. [[CrossRef](#)] [[PubMed](#)]
27. Sharan, R.; Karp, R.M. Reconstructing Boolean models of signaling. *J. Comput. Biol.* **2013**, *20*, 249–257. [[CrossRef](#)]
28. Schneidman, E.; Berry, M.J.; Segev, R.; Bialek, W. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature* **2006**, *440*, 1007–1012. [[CrossRef](#)]
29. Shlens, J.; Field, G.D.; Gauthier, J.L.; Grivich, M.I.; Petrusca, D.; Sher, A.; Litke, A.M.; Chichilnisky, E. The structure of multi-neuron firing patterns in primate retina. *J. Neurosci.* **2006**, *26*, 8254–8266. [[CrossRef](#)]

30. Quadeer, A.A.; McKay, M.R.; Barton, J.P.; Louie, R.H. MPF–BML: A standalone GUI-based package for maximum entropy model inference. *Bioinformatics* **2020**, *36*, 2278–2279. [[CrossRef](#)]
31. Tang, A.; Jackson, D.; Hobbs, J.; Chen, W.; Smith, J.L.; Patel, H.; Prieto, A.; Petrusca, D.; Grivich, M.I.; Sher, A.; et al. A maximum entropy model applied to spatial and temporal correlations from cortical networks in vitro. *J. Neurosci.* **2008**, *28*, 505–518. [[CrossRef](#)] [[PubMed](#)]
32. Cocco, S.; Leibler, S.; Monasson, R. Neuronal couplings between retinal ganglion cells inferred by efficient inverse statistical physics methods. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 14058–14062. [[CrossRef](#)] [[PubMed](#)]
33. Roudi, Y.; Nirenberg, S.; Latham, P.E. Pairwise maximum entropy models for studying large biological systems: When they can work and when they can not. *PLoS Comput. Biol.* **2009**, *5*, e1000380. [[CrossRef](#)]
34. Tkačik, G.; Prentice, J.S.; Balasubramanian, V.; Schneidman, E. Optimal population coding by noisy spiking neurons. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 14419–14424. [[CrossRef](#)]
35. Ohiorhenuan, I.E.; Mechler, F.; Purpura, K.P.; Schmid, A.M.; Hu, Q.; Victor, J.D. Sparse coding and high-order correlations in fine-scale cortical networks. *Nature* **2010**, *466*, 617–621. [[CrossRef](#)]
36. Yeh, F.C.; Tang, A.; Hobbs, J.P.; Hottowy, P.; Dabrowski, W.; Sher, A.; Litke, A.; Beggs, J.M. Maximum entropy approaches to living neural networks. *Entropy* **2010**, *12*, 89–106. [[CrossRef](#)]
37. Granot-Atedgi, E.; Tkačik, G.; Segev, R.; Schneidman, E. Stimulus-dependent maximum entropy models of neural population codes. *PLoS Comput. Biol.* **2013**, *9*, e1002922. [[CrossRef](#)]
38. Tkačik, G.; Marre, O.; Mora, T.; Amodei, D.; Berry II, M.J.; Bialek, W. The simplest maximum entropy model for collective behavior in a neural network. *J. Stat. Mech. Theory Exp.* **2013**, *2013*, P03011. [[CrossRef](#)]
39. Ferrari, U.; Obuchi, T.; Mora, T. Random versus maximum entropy models of neural population activity. *Phys. Rev. E* **2017**, *95*, 042321. [[CrossRef](#)]
40. Rostami, V.; Mana, P.P.; Grün, S.; Helias, M. Bistability, non-ergodicity, and inhibition in pairwise maximum-entropy models. *PLoS Comput. Biol.* **2017**, *13*, e1005762. [[CrossRef](#)]
41. Nghiem, T.A.; Teleńczuk, B.; Marre, O.; Destexhe, A.; Ferrari, U. Maximum entropy models reveal the correlation structure in cortical neural activity during wakefulness and sleep. *bioRxiv* **2018**, 243857. [[CrossRef](#)]
42. Yeo, G.; Burge, C.B. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J. Comput. Biol.* **2004**, *11*, 377–394. [[CrossRef](#)]
43. Mora, T.; Walczak, A.M.; Bialek, W.; Callan, C.G. Maximum entropy models for antibody diversity. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 5405–5410. [[CrossRef](#)]
44. Santolini, M.; Mora, T.; Hakim, V. A general pairwise interaction model provides an accurate description of in vivo transcription factor binding sites. *PLoS ONE* **2014**, *9*, e99015. [[CrossRef](#)]
45. Fariselli, P.; Taccioli, C.; Pagani, L.; Maritan, A. DNA sequence symmetries from randomness: The origin of the Chargaff’s second parity rule. *Brief. Bioinform.* **2020**, *22*, 2172–2181. [[CrossRef](#)]
46. Fernandez-de Cossio-Diaz, J.; Mulet, R. Maximum entropy and population heterogeneity in continuous cell cultures. *PLoS Comput. Biol.* **2019**, *15*, e1006823. [[CrossRef](#)]
47. Jackson, A.; Constable, C.; Gillet, N. Maximum entropy regularization of the geomagnetic core field inverse problem. *Geophys. J. Int.* **2007**, *171*, 995–1004. [[CrossRef](#)]
48. De Martino, A.; De Martino, D. An introduction to the maximum entropy approach and its application to inference problems in biology. *Heliyon* **2018**, *4*, e00596. [[CrossRef](#)]
49. Khalifa, F.; Soliman, A.; El-Baz, A.; Abou El-Ghar, M.; El-Diasty, T.; Gimel’farb, G.; Ouseph, R.; Dwyer, A.C. Models and methods for analyzing DCE-MRI: A review. *Med. Phys.* **2014**, *41*, 124301. [[CrossRef](#)]
50. Fennessy, F.M.; McKay, R.R.; Beard, C.J.; Taplin, M.E.; Tempny, C.M. Dynamic contrast-enhanced magnetic resonance imaging in prostate cancer clinical trials: Potential roles and possible pitfalls. *Transl. Oncol.* **2014**, *7*, 120–129. [[CrossRef](#)]
51. Huang, W.; Li, X.; Chen, Y.; Li, X.; Chang, M.C.; Oborski, M.J.; Malyarenko, D.I.; Muzi, M.; Jajamovich, G.H.; Fedorov, A.; et al. Variations of dynamic contrast-enhanced magnetic resonance imaging in evaluation of breast cancer therapy response: A multicenter data analysis challenge. *Transl. Oncol.* **2014**, *7*, 153. [[CrossRef](#)]
52. Sobhani, F.; Xu, C.; Murano, E.; Pan, L.; Rastegar, N.; Kamel, I.R. Hypo-vascular liver metastases treated with transarterial chemoembolization: Assessment of early response by volumetric contrast-enhanced and diffusion-weighted magnetic resonance imaging. *Transl. Oncol.* **2016**, *9*, 287–294. [[CrossRef](#)]
53. Usuda, K.; Iwai, S.; Funasaki, A.; Sekimura, A.; Motono, N.; Matoba, M.; Doai, M.; Yamada, S.; Ueda, Y.; Uramoto, H. Diffusion-weighted magnetic resonance imaging is useful for the response evaluation of chemotherapy and/or radiotherapy to recurrent lesions of lung cancer. *Transl. Oncol.* **2019**, *12*, 699–704. [[CrossRef](#)]
54. Stoyanova, R.; Huang, K.; Sandler, K.; Cho, H.; Carlin, S.; Zanzonico, P.B.; Koutcher, J.A.; Ackerstaff, E. Mapping tumor hypoxia in vivo using pattern recognition of dynamic contrast-enhanced MRI data. *Transl. Oncol.* **2012**, *5*, 437. [[CrossRef](#)]
55. Schmid, V.J.; Whitcher, B.; Padhani, A.R.; Taylor, N.J.; Yang, G.Z. Bayesian methods for pharmacokinetic models in dynamic contrast-enhanced magnetic resonance imaging. *IEEE Trans. Med. Imaging* **2006**, *25*, 1627–1636. [[CrossRef](#)]
56. Tofts, P.S.; Brix, G.; Buckley, D.L.; Evelhoch, J.L.; Henderson, E.; Knopp, M.V.; Larsson, H.B.; Lee, T.Y.; Mayr, N.A.; Parker, G.J.; et al. Estimating kinetic parameters from dynamic contrast-enhanced T1-weighted MRI of a diffusible tracer: Standardized quantities and symbols. *J. Magn. Reson. Imaging Off. J. Int. Soc. Magn. Reson. Med.* **1999**, *10*, 223–232. [[CrossRef](#)]



57. Shao, J.; Zhang, Z.; Liu, H.; Song, Y.; Yan, Z.; Wang, X.; Hou, Z. DCE-MRI pharmacokinetic parameter maps for cervical carcinoma prediction. *Comput. Biol. Med.* **2020**, *118*, 103634. [[CrossRef](#)]
58. Lingala, S.G.; Guo, Y.; Bliesener, Y.; Zhu, Y.; Lebel, R.M.; Law, M.; Nayak, K.S. Tracer kinetic models as temporal constraints during brain tumor DCE-MRI reconstruction. *Med. Phys.* **2020**, *47*, 37–51. [[CrossRef](#)]
59. Zou, J.; Balter, J.M.; Cao, Y. Estimation of pharmacokinetic parameters from DCE-MRI by extracting long and short time-dependent features using an LSTM network. *Med. Phys.* **2020**, *47*, 3447–3457. [[CrossRef](#)]
60. Dikaos, N. Stochastic Gradient Langevin dynamics for joint parameterization of tracer kinetic models, input functions, and T1 relaxation-times from undersampled k-space DCE-MRI. *Med. Image Anal.* **2020**, *62*, 101690. [[CrossRef](#)]
61. Tofts, P.S.; Kermode, A.G. Measurement of the blood-brain barrier permeability and leakage space using dynamic MR imaging. 1. Fundamental concepts. *Magn. Reson. Med.* **1991**, *17*, 357–367. [[CrossRef](#)]
62. Larsson, H.B.W.; Tofts, P.S. Measurement of blood-brain barrier permeability using dynamic Gd-DTPA scanning—A comparison of methods. *Magn. Reson. Med.* **1992**, *24*, 174–176. [[CrossRef](#)]
63. Brix, G.; Kiessling, F.; Lucht, R.; Darai, S.; Wasser, K.; Delorme, S.; Griebel, J. Microcirculation and microvasculature in breast tumors: Pharmacokinetic analysis of dynamic MR image series. *Magn. Reson. Med.* **2004**, *52*, 420–429. [[CrossRef](#)]
64. Berg, B.; Stucht, D.; Janiga, G.; Beuing, O.; Speck, O.; Thovenin, D. Cerebral Blood Flow in a Healthy Circle of Willis and Two Intracranial Aneurysms: Computational Fluid Dynamics Versus Four-Dimensional Phase-Contrast Magnetic Resonance Imaging. *ASME J. Biomech. Eng.* **2014**, *15*, 041003. [[CrossRef](#)]
65. Orton, M.R.; Collins, D.J.; Walker-Samuel, S.; d’Arcy, J.A.; Hawkes, D.J.; Atkinson, D.; Leach, M.O. Bayesian estimation of pharmacokinetic parameters for DCE-MRI with a robust treatment of enhancement onset time. *Phys. Med. Biol.* **2007**, *52*, 2393–2408. [[CrossRef](#)]
66. Dikaos, N.; Arridge, S.; Hamy, V.; Punwani, S.; Atkinson, D. Direct parametric reconstruction from undersampled (k, t)-space data in dynamic contrast enhanced MRI. *Med. Image Anal.* **2014**, *18*, 989–1001. [[CrossRef](#)]
67. Bender, R.; Heinemann, L. Fitting nonlinear regression models with correlated errors to individual pharmacodynamic data using SAS software. *J. Pharmacokinet. Biopharm.* **1995**, *23*, 87–100. [[CrossRef](#)]
68. Cheng, H.L.M. T1 measurement of flowing blood and arterial input function determination for quantitative 3D T1-weighted DCE-MRI. *J. Magn. Reson. Imaging JMRI* **2007**, *25*, 1073–1078. [[CrossRef](#)]
69. Gauthier, M. Impact of the arterial input function on microvascularization parameter measurements using dynamic contrast-enhanced ultrasonography. *World J. Radiol.* **2012**, *4*, 291. [[CrossRef](#)]
70. Cheng, H.L.M. Investigation and optimization of parameter accuracy in dynamic contrast-enhanced MRI. *J. Magn. Reson. Imaging* **2008**, *28*, 736–743. [[CrossRef](#)]
71. Lavini, C. Simulating the effect of input errors on the accuracy of Tofts’ pharmacokinetic model parameters. *Magn. Reson. Imaging* **2015**, *33*, 222–235. [[CrossRef](#)]
72. Peled, S.; Vangel, M.; Kikinis, R.; Tempny, C.M.; Fennessy, F.M.; Fedorov, A. Selection of fitting model and arterial input function for repeatability in dynamic contrast-enhanced prostate MRI. *Acad. Radiol.* **2019**, *26*, e241–e251. [[CrossRef](#)]
73. Huang, W.; Chen, Y.; Fedorov, A.; Li, X.; Jajamovich, G.H.; Malyarenko, D.I.; Aryal, M.P.; LaViolette, P.S.; Oborski, M.J.; O’Sullivan, F.; et al. The impact of arterial input function determination variations on prostate dynamic contrast-enhanced magnetic resonance imaging pharmacokinetic modeling: A multicenter data analysis challenge. *Tomography* **2016**, *2*, 56–66. [[CrossRef](#)]
74. Huang, W.; Chen, Y.; Fedorov, A.; Li, X.; Jajamovich, G.H.; Malyarenko, D.I.; Aryal, M.P.; LaViolette, P.S.; Oborski, M.J.; O’Sullivan, F.; et al. The impact of arterial input function determination variations on prostate dynamic contrast-enhanced magnetic resonance imaging pharmacokinetic modeling: A multicenter data analysis challenge, part II. *Tomography* **2019**, *5*, 99–109. [[CrossRef](#)]
75. Keil, V.C.; Mädler, B.; Gieseke, J.; Fimmers, R.; Hattingen, E.; Schild, H.H.; Hadizadeh, D.R. Effects of arterial input function selection on kinetic parameters in brain dynamic contrast-enhanced MRI. *Magn. Reson. Imaging* **2017**, *40*, 83–90. [[CrossRef](#)]
76. Parker, G.J.; Roberts, C.; Macdonald, A.; Buonaccorsi, G.A.; Cheung, S.; Buckley, D.L.; Jackson, A.; Watson, Y.; Davies, K.; Jayson, G.C. Experimentally-derived functional form for a population-averaged high-temporal-resolution arterial input function for dynamic contrast-enhanced MRI. *Magn. Reson. Med. Off. J. Int. Soc. Magn. Reson. Med.* **2006**, *56*, 993–1000. [[CrossRef](#)]
77. Rata, M.; Collins, D.J.; Darcy, J.; Messiou, C.; Tunariu, N.; Desouza, N.; Young, H.; Leach, M.O.; Orton, M.R. Assessment of repeatability and treatment response in early phase clinical trials using DCE-MRI: Comparison of parametric analysis using MR-and CT-derived arterial input functions. *Eur. Radiol.* **2016**, *26*, 1991–1998. [[CrossRef](#)]
78. Rijpkema, M.; Kaanders, J.H.; Joosten, F.B.; van der Kogel, A.J.; Heerschap, A. Method for quantitative mapping of dynamic MRI contrast agent uptake in human tumors. *J. Magn. Reson. Imaging Off. J. Int. Soc. Magn. Reson. Med.* **2001**, *14*, 457–463. [[CrossRef](#)]
79. Ashton, E.; Raunig, D.; Ng, C.; Kelcz, F.; McShane, T.; Evelhoch, J. Scan-rescan variability in perfusion assessment of tumors in MRI using both model and data-derived arterial input functions. *J. Magn. Reson. Imaging Off. J. Int. Soc. Magn. Reson. Med.* **2008**, *28*, 791–796. [[CrossRef](#)]
80. Weinmann, H.J.; Laniado, M.; Mützel, W. Pharmacokinetics of Gd-DTPA/Dimeglumine after intravenous injection into healthy volunteers. *Physiol. Chem. Phys. Med. NMR* **1984**, *16*, 167–172.
81. Fritz-Hansen, T.; Rostrup, E.; Larsson, H.B.W.; Søndergaard, L.; Ring, P.; Henriksen, O. Measurement of the Arterial Concentration of Gd-DTPA Using MRI: A step toward Quantitative Perfusion Imaging. *Magn. Reson. Med.* **1996**, *36*, 225–231. [[CrossRef](#)]
82. Farsani, Z.A.; Schmid, V.J. Maximum Entropy Approach in Dynamic Contrast-Enhanced Magnetic Resonance Imaging. *Methods Inf. Med.* **2017**, *56*, 461–468. [[CrossRef](#)]

83. Rao, R.V.; Savsani, V.J.; Vakharia, D. Teaching–learning-based optimization: A novel method for constrained mechanical design optimization problems. *Comput.-Aided Des.* **2011**, *43*, 303–315. [[CrossRef](#)]
84. Zou, F.; Chen, D.; Xu, Q. A survey of teaching–learning-based optimization. *Neurocomputing* **2019**, *335*, 366–383. [[CrossRef](#)]
85. Parker, G.J.; Suckling, J.; Tanner, S.F.; Padhani, A.R.; Revell, P.B.; Husband, J.E.; Leach, M.O. Probing tumor microvasculature by measurement, analysis and display of contrast agent uptake kinetics. *J. Magn. Reson. Imaging* **1997**, *7*, 564–574. [[CrossRef](#)]
86. d’Arcy, J.A.; Collins, D.J.; Padhani, A.R.; Walker-Samuel, S.; Suckling, J.; Leach, M.O. Magnetic resonance imaging workbench: Analysis and visualization of dynamic contrast-enhanced MR imaging data. *Radiographics* **2006**, *26*, 621–632. [[CrossRef](#)]
87. Buckley, D.; Parker, G. Measuring Contrast Agent Concentration in  $T_1$ -Weighted Dynamic Contrast-Enhanced MRI. In *Dynamic Contrast-Enhanced Magnetic Resonance Imaging in Oncology*; Jackson, A., Parker, G.J.M., Buckley, D.L., Eds.; Springer: Berlin/Heidelberg, Germany; New York, NY, USA, 2005; Chapter 5, pp. 69–80.
88. Parzen, E. On estimation of a probability density function and mode. *Ann. Math. Stat.* **1962**, *33*, 1065–1076. [[CrossRef](#)]
89. Choyke, P.; Dwyer, A.; Knopp, M. Functional tumor imaging with dynamic contrast-enhanced magnetic resonance imaging. *Magn. Reson. Med.* **2003**, *17*, 509–520.
90. Murase, K. Efficient method for calculating kinetic parameters using  $T_1$ -weighted dynamic contrast-enhanced magnetic resonance imaging. *Magn. Reson. Med.* **2004**, *51*, 858–862. [[CrossRef](#)] [[PubMed](#)]
91. Mohammad-Djafari, A. Bayesian Image Processing. In Proceedings of the Fifth International Conference on Modelling, Computation and Optimization in Information Systems and Management Sciences (MCO 2004), Metz, France, 1–3 July 2004.
92. Mohammad-Djafari, A.; Demoment, G. Estimating priors in maximum entropy image processing. In Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, Albuquerque, NM, USA, 3–6 April 1990; pp. 2069–2072.
93. Mohammad-Djafari, A. A full Bayesian approach for inverse problems. In *Maximum Entropy and Bayesian Methods*; Springer: Berlin/Heidelberg, Germany, 1996; pp. 135–144.
94. Hadamard, J. *Le Probleme de Cauchy et les Équations aux Dérivées Partielles Linéaires Hyperboliques*; Paris Russian translation: 1932; Volume 220.
95. Turchin, V.F. Solution of the Fredholm equation of the first kind in a statistical ensemble of smooth functions. *USSR Comput. Math. Math. Phys.* **1967**, *7*, 79–96. [[CrossRef](#)]
96. Denisova, N. Bayesian maximum-a posteriori approach with global and local regularization to image reconstruction problem in medical emission tomography. *Entropy* **2019**, *21*, 1108. [[CrossRef](#)]
97. Sparavigna, A.C. Entropy in image analysis. *Entropy* **2019**, *21*, 502. [[CrossRef](#)] [[PubMed](#)]
98. Skilling, J. The axioms of maximum entropy. In *Maximum-Entropy and Bayesian Methods in Science and Engineering*; Springer: Berlin/Heidelberg, Germany, 1988; pp. 173–187.
99. Elfving, T. An Algorithm for Maximum Entropy Image Reconstruction from Noisy Data. *Mathl. Comput. Model.* **1989**, *12*, 729–745. [[CrossRef](#)]
100. Akpınar, S.; Akpınar, E.K. Wind energy analysis based on maximum entropy principle (MEP)-type distribution function. *Energy Convers. Manag.* **2007**, *48*, 1140–1149. [[CrossRef](#)]
101. Casella, G.; Berger, R. *Statistical Inference 2*; Duxbury: Belmont, CA, USA, 2002.
102. García, J.A.M.; Mena, A.J.G. Optimal distributed generation location and size using a modified teaching-learning based optimization algorithm. *Int. J. Electr. Power Energy Syst.* **2013**, *50*, 65–75. [[CrossRef](#)]
103. Bain, L.J.; Antle, C.E. Estimation of parameters in the weibull distribution. *Technometrics* **1967**, *9*, 621–627. [[CrossRef](#)]
104. Stevens, M.; Smulders, P. The estimation of the parameters of the Weibull wind speed distribution for wind energy utilization purposes. *Wind Eng.* **1979**, *3*, 132–145.
105. Justus, C.; Hargraves, W.; Mikhail, A.; Graber, D. Methods for estimating wind speed frequency distributions. *J. Appl. Meteorol.* **1978**, *17*, 350–353. [[CrossRef](#)]
106. Morgan, E.C.; Lackner, M.; Vogel, R.M.; Baise, L.G. Probability distributions for offshore wind speeds. *Energy Convers. Manag.* **2011**, *52*, 15–26. [[CrossRef](#)]
107. Akdağ, S.A.; Dinler, A. A new method to estimate Weibull parameters for wind energy applications. *Energy Convers. Manag.* **2009**, *50*, 1761–1766. [[CrossRef](#)]
108. Werapun, W.; Tirawanichakul, Y.; Waewsak, J. Comparative study of five methods to estimate Weibull parameters for wind speed on Phangan Island, Thailand. *Energy Procedia* **2015**, *79*, 976–981. [[CrossRef](#)]
109. Zhang, H.; Yu, Y.J.; Liu, Z.Y. Study on the Maximum Entropy Principle applied to the annual wind speed probability distribution: A case study for observations of intertidal zone anemometer towers of Rudong in East China Sea. *Appl. Energy* **2014**, *114*, 931–938. [[CrossRef](#)]
110. Li, T.; Griffiths, W.; Chen, J. Weibull modulus estimated by the non-linear least squares method: A solution to deviation occurring in traditional Weibull estimation. *Metall. Mater. Trans. A* **2017**, *48*, 5516–5528. [[CrossRef](#)]
111. Seguro, J.; Lambert, T. Modern estimation of the parameters of the Weibull wind speed distribution for wind energy analysis. *J. Wind. Eng. Ind. Aerodyn.* **2000**, *85*, 75–84. [[CrossRef](#)]
112. Cook, N.J. “Discussion on modern estimation of the parameters of the Weibull wind speed distribution for wind speed energy analysis” by J.V. Seguro, T.W. Lambert. *J. Wind. Eng. Ind. Aerodyn.* **2001**, *89*, 867–869. [[CrossRef](#)]