

Article

Exploratory Matching Model Search Algorithm (EMMSA) for Causal Analysis: Application to the Cardboard Industry

Richard Aviles-Lopez ^{1,*} , Juan de Dios Luna del Castillo ²  and Miguel Ángel Montero-Alonso ^{2,*} ¹ Department of Computer Science and A.I., University of Granada, 18071 Granada, Spain² Department of Statistics and Operational Research, University of Granada, 18071 Granada, Spain; jdluna@ugr.es

* Correspondence: richardaviles@correo.ugr.es (R.A.-L.); mmontero@ugr.es (M.Á.M.-A.)

Abstract: This paper aims to present a methodology for the application of matching methods in industry to measure causal effect size. Matching methods allow us to obtain treatment and control samples with their covariates as similar as possible. The matching techniques used are nearest, optimal, full, coarsened exact matching (CEM), and genetic. These methods have been widely used in medical, psychological, and economic sciences. The proposed methodology provides two algorithms to execute these methods and to conduct an exhaustive search for the best models. It uses three conditions to ensure, as far as possible, the balance of all covariates, the maximum number of units in the treatment and control groups, and the most significant causal effect sizes. These techniques are applied in the carton board industry, where the causal variable is downtime, and the outcome variable is waste generated. A dataset from the carton board industry is used, and the results are contrasted with an expert in this process. Meta-analysis techniques are used to integrate the results of different comparative studies, which could help to determine and prioritize where to reduce waste. Two machines were found to generate more waste in terms of standardized measures whose values are 0.52 and 0.53, representing 48.60 and 36.79 linear meters (LM) on average for each production order with a total downtime of more than 3000 s. In general, for all machines, the maximum average wastage for each production order is 24.98 LM and its confidence interval is [13.40;36.23] LM. The main contribution of this work is the use of causal methodology to estimate the effect of downtime on waste in an industry. Particularly relevant is the contribution of an algorithm that aims to obtain the best matching model for this application. Its advantages and disadvantages are evaluated, and future areas of research are outlined. We believe that this methodology can be applied to other industries and fields of knowledge.

Keywords: matching; exploratory matching algorithm; homologous model search algorithm; manufacturing; cardboard industry; genetic; exploratory matching model search algorithm; EMMSA

MSC: 62D20

Citation: Aviles-Lopez, R.; Luna del Castillo, J.d.D.; Montero-Alonso, M.Á. Exploratory Matching Model Search Algorithm (EMMSA) for Causal Analysis: Application to the Cardboard Industry. *Mathematics* **2023**, *11*, 4506. <https://doi.org/10.3390/math11214506>

Academic Editor: Bumba Mukherjee

Received: 23 September 2023

Revised: 16 October 2023

Accepted: 28 October 2023

Published: 31 October 2023



Copyright: © 2023 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/>).

1. Introduction

The cardboard industry has been growing steadily, attributable to increased corrugated box packaging needs for export and e-commerce from its versatility to protect, store, and transport [1]. The corrugated cardboard is made up of two sheets of paper (internal and external) and, in the middle, the corrugated paper that provides the energy-absorbing capacity [2]. The production volume is expected to reach 700 million metric tons by 2040, with a market size bordering USD 220 billion by 2028 [3]. Consequently, more waste will be generated which must be controlled because of its environmental, energy, and economic burdens [4].

One of the sources of waste in this industry is the printing subsystems, composed of several printing machines forming production lines (printing lines), where each includes

several stages or modules [1]. Each machine may have different characteristics depending on the number of colors to be printed, the volume to be produced, and the post-printing stages it includes.

The production process of an order starts with a setup stoppage, and multiple stoppages may occur due to non-programmed or programmed operations. These stoppages are registered in real time by alarm monitoring software. The waste can arise from intrinsic reasons of the product design that can be calculated for each sheet, but we are interested in the waste sheets due to stoppages in the printing line. Thus, some of these stoppages can generate more waste due to effects on unscheduled maintenance, staffing, supply of materials, resource planning, subsystem failures, or fault propagations, among others [5].

Thus, determining the waste volume produced by each stoppage is impractical, complex, and difficult to quantify. The strategy followed has been to address all sources of waste using, for example, total productive maintenance [6], but which machine generates the most waste due to downtime? In which time interval does the most waste occur? Where to start? This generates more uncertainty because it is very complex to determine which stoppages are more important than others in almost half a million stoppage records.

However, if the waste is significant, the machine will likely stop, so somehow there should be a relationship between downtime and waste. For this reason, from a causal perspective, a study is required to determine the magnitude of the effect produced by the downtime of all these causes on waste.

Causal research in the manufacturing industry has been concentrated on the alarm root cause analysis, presenting multiple data-driven methods, such as those oriented to time-series data (Granger causality or entropy transfer), as well as methods oriented towards cross-sectional data, such as machine learning and probabilistic graphical models [7]. However, the former cannot be applied because the actual available data for downtimes only records the date but not the time when it occurs. Machine learning makes excellent predictions, but it does not properly determine the causes or adequately measure the causal effect size [8,9].

Therefore, the objective of this paper is to present a methodology to measure the causal effect size of the total downtime (seconds) of production orders on waste (LM) using production observational data. In addition, we will study its strengths and weaknesses. We believe that this methodology can be extended to other industries.

Addressing this problem is challenging due to the multiple sources of noise that occur in an industry: data distributions are non-normal, relationships are non-linear, and many variables exhibit multicollinearity [10]. Another problem is to find a parametric model that adequately measures the effect size. In addition, one of the challenges is to preprocess the data in such a way as to eliminate as much bias as possible [11].

We will use the matching methods to balance observational data and choose the best matching models that properly measure the effect size [12]. However, traditionally, it has been recommended to use several matching models to choose the best models, this being a manual search process. To our knowledge, no specific methodology exists for this purpose, nor have matching methods been used in the industry. The main contributions of this research are:

- The presentation of the exploratory matching model search algorithm (EMMSA) for causal analysis.
- The exploratory matching algorithm (EMA) for building matching models for each machine and treatment.
- The homologous model selection algorithm (HMSA) to choose the best matching models and select the one that detects the maximum effect.
- The application of meta-analysis to integrate the findings from the different machines and treatments.

To accomplish this, we will use a real dataset from a cardboard factory that is not fully automated and facing organizational difficulties and limitations, which increases the

dataset's variability. For these data, we pose a single research question: What is the effect size of total downtime on the waste generated on a printing line?

This article is organized into five sections. Section 2 summarizes the main causal methods used in the industry, the potential outcome framework, multivariate matching techniques, and meta-analysis models. Section 3 presents a description of the data, and how they were obtained. Section 4 presents the proposed industry application methodology and describes the exploratory matching algorithms and the homologous model selection algorithm. Section 5 presents and discusses the results. Section 6 concludes the study.

2. Literature Review

2.1. Causality Techniques Applied to the Manufacturing Process

Almizahad et al. synthesized the different methods of root cause analysis for both time series and cross-sectional data [7]. For time series, two approaches are available depending on whether they use multivariate Granger causality (MGC) [12] or transfer entropy (TE) methods.

Granger causality (G-causality) is a concept of direct causality used to establish whether the past values of a time series help to improve the prediction of another time series [13]. This method makes several assumptions, such as that the cause occurs before its effect, the time series values are continuous and Gaussian, the causal effect is linear, the time series is stationary [14], there is a discrete sampling frequency, and unmeasured confounding variables are not present [15]. These assumptions are unlikely to be met in practice [15]. MGC extends bivariate Granger causality to multiple variables and its advantage is that it has a low computational cost, is easier to implement, and works well with little data [16]. Principal component analysis is traditionally used as a feature selection method [7]. The main applications of MGC in root cause analysis are indicated in Table A1.

TE measures the amount of information transferred between two time series [17]. The TE methods were developed by establishing probability density functions (PDF) for each variable using kernel estimators because they offer robustness and accuracy [18]. A causal measure is established comparing the transfer entropy between two variables in one and another direction, and the significance level (threshold) of this measure is defined using Monte Carlo simulations of surrogated data [18]. Surrogated data can be generated by randomly shuffled data, or by the interactive amplitude-adjusted Fourier transform [19]. The main applications of MGC in root cause analysis are indicated in Table A2.

TE methods have the advantage of detecting linear or nonlinear causal relationships but have a high computational cost due to the application of their significance test based on Monte Carlo simulations. In addition, they require a large amount of seasonal data. They also require a method to select model parameters due to the use of PDF methods based on kernel estimators. Other disadvantages of TE and MGC are the inability to detect spurious relationships and to determine the existence of hidden (unmeasured) variables.

In cross-sectional data, Bayesian networks have been widely used. A Bayesian network (BN) is a directed acyclic graph (DAG) in which the nodes represent the random variables of the system under study, and the edges indicate the incidence of one node on another [20]. However, they have the disadvantage that they do not allow cycles, which are very present in industrial control processes. They are also not as accurate because the learning algorithms find equivalent Markov classes with multiple linking possibilities, which makes them difficult to interpret [21].

All these methods are data-driven, only detect directionality, and do not measure the causal effect size. To achieve this, they require the elimination of random noise that could affect the measurement of the effect size. In addition, they must control for bias introduced, and none of these approaches have addressed this yet.

As far as we know, no relevant records of matching methods were obtained in the fields of production and logistics. Therefore, it can be deduced that matching techniques have not been used to measure causal effect size in the industrial sector.

2.2. Potential Outcome Framework

To make inferences about the causal effect of a treatment, the potential outcome model was formalized by Donald Rubin [22]. A *causal effect* is the value comparison (difference or ratio) of the potential outcome (Y) when it is exposed to two alternative *causal states or alternative treatments* [22,23]. Alternative treatments can be binary (1: treatment and 0: control) or multi-treatment, and they should be defined into mutually exclusive alternative states [24]. In randomized trials, the way “each individual” is chosen to receive the treatment condition is known as the *allocation mechanism* [25], but in observational studies, it is impossible to control this aspect.

The individual causal effect is defined by the contrast between alternative states (treatment and control) for the same individual in the population of interest. The *aggregate causal effects* (average treatment effect—*ATE*) estimate the mean of the individual effects [26] which are shown in Equations (1) and (2). These causal effects can be defined over any subset of the population. $E[.]$ signifies the expectation operator of probability theory.

$$ATE = E[Y^1] - E[Y^0] = E[Y^1 - Y^0] \tag{1}$$

$$ATE = \frac{1}{N} \sum_{i=1}^N E[Y_i^1 - Y_i^0 | X_i] = \frac{1}{N} \sum_{i=1}^N (\mu_1(Y_i) - \mu_0(Y_i)), \tag{2}$$

where $E[Y^1]$ is the expected value of potential outcome for the treatment group; $E[Y^0]$ is the expected value of the potential outcome for the control group; $E[Y^1 - Y^0]$ is the expected value of the difference of the two expectations equivalent to the mean treatment effect; $E[Y_i^1 - Y_i^0 | X_i]$ is the expected value of the difference in the individual potential outcome, given their covariables X_i ; $\mu_1(Y_i)$ is the mean individual potential outcome of the treatment group; $\mu_0(Y_i)$ is the mean individual potential outcome of the control group; and N is the size of the control or treatment groups.

In observational studies, the potential outcome of an alternative treatment that is not known is called the *counterfactual*. Thus, each observed potential outcome variable only reveals half of the information contained because it is impossible to observe both values for the same individual, either for ethical or practical reasons [23], and so one can never calculate the individual causal effect. This challenge is known as the fundamental problem of causal inference [26].

Because of this impossibility, the *average treatment effect on the treated* (*ATT*) is used more often (Equation (3)), where n is the size of the groups, m is the number of treatments, and T_j is the weighting of each treatment.

$$ATT = \frac{1}{\sum_{j=1}^m T_j} \sum_{i=1}^n T_j E[Y_i^1 - Y_i^0 | X_i] = \frac{1}{\sum_{j=1}^m T_j} \sum_{i=1}^n T_j [\mu_1(Y_i) - \mu_0(Y_i)] \tag{3}$$

When the causal effects are constant in each i , the *ATT* is equal to the *ATE*. However, another group-level causal effect is the *causal hazard rate* (or causal risk ratio), which is often used in epidemiology and health sciences [27].

$$Causal\ Hazard\ Rate = \frac{E[Y^1]}{E[Y^0]} = \frac{P(Y^1 = 1)}{P(Y^0 = 1)} \tag{4}$$

2.3. Multivariable Matching Methods Overview

Estimating the causal effect in observational studies is a critical task as the researcher has no control over the treatment *allocation mechanism*. Therefore, its covariates may differ between treatment and control individuals (unbalanced) [28]. The balance refers to the similarity of the marginal covariate distribution in both groups and thus avoids the *model dependence* or imprecise inference [29].

The *Model dependence* is the variation in causal effect estimation at point x_i among plausible alternative models that fit the data reasonably well [30]. However, when there

is no model dependency, the functional form does not matter because the result will be approximately the same [31].

Thus, observational studies require a *design phase* to resemble randomized experiments, using background information (covariates), create more balanced treatment and control samples, and support more robust and credible inferences [25,29,32].

Multivariate matching is a set of nonparametric methods applied before parametric analysis to eliminate the difference in the covariates (X_i) affecting the causal variable (T_i) without accessing the outcome variable (Y_i) [31]. It improves the balance with a minimum of loose observations in the process, allows for controlling the influence of confounding factors, and obtains more precise estimates of the causal effect with reduced bias and variance [28].

In order not to bias the analysis, the outcomes should not be available, and thus, by comparing several matched designs, the search for a satisfactory design becomes transparent [33].

The fundamental rule to avoid **selection bias** is that our preprocessed data set should include a subset of records from the observed sample whose treatment and control groups are unrelated and which have the same baseline characteristics [31], such that:

$$\tilde{p}(X_i | T = 1) = \tilde{p}(X_i | T = 0) \quad (5)$$

Matching techniques only require that their distributions (treatment and control) be as close as possible, although the unobserved covariate bias is always possible [29]. In this way, T_i is independent of X_i , and the dependence on the functional form in the parametric analysis will be eliminated [31].

A first matching attempt was to use *one-to-one exact matching*, a treated unit with a control unit that contains both identical values of X_i , discarding all units that do not match. *Exact matching* uses one or more control units that match each treated unit, which improves efficiency [28].

According to [30], they suggest removing observations from the control group that are outside the "*convex hull*" of the treatment group. That is, units in the control group that are outside the range of the treatment variable are discarded. For this reason, to avoid extrapolation bias, it is necessary to check the *common support* and verify whether the control units are in the convex hull of the treatment group [31].

Thus, if the number of observations of the matched sample N is not reduced so much, matching will frequently reduce the bias and the variance of the estimates of the parametric analysis [31]. If $N_C \gg N_T$, then more control units can be matched with each treatment unit, which improves the efficiency of estimation, and the bias may be reduced [34]. If $N_T \gg N_C$, then each unit control may be matched with several treated units, which is known as *matching with replacement* [35].

The *propensity score* is a measure that summarizes all X_i variables, and it is defined as the conditional probability of unit i receiving treatment T , given their covariates X_i [36]. The *propensity score* is formulated as $e(X_i)$ in the following equation:

$$e(X_i) = \Pr(T = 1 | X_i) \quad (6)$$

It is commonly calculated using a logit model (or random forests). Some matching techniques are based on a **proximity metric**, either *distance* or *similarity*. It may also use a threshold to ensure that the difference between the values of the metrics of the matched treatment and control units is less than or equal to it [37]. This threshold is known as a *caliper*.

Nearest-neighbor matching calculates the closeness metric of each observation based on the values of its covariates and then it matches the closest control and treatment units to each other [38]. Two versions are available for matching without substitution: *greedy matching* or *optimal matching*. The first matches a treated unit with the control unit that has the most similar propensity score without minimizing the overall distance, and the second includes minimizing the absolute distance between the matched units [39].

However, propensity scores attempt to balance the covariates but two units with the same value score may be very different. So, another approach is to additionally use the *Mahalanobis* distance within the *calibrators* of the propensity score, thus adding a layer of protection, but will not normally generate optimal pair matching [40].

Optimal matching solves an optimal assignment problem to find the closest treatment and control units that meet certain requirements or constraints on the balance of covariates [41]. Optimal matching is often used with the RELAX IV algorithm to solve the minimum cost flow problem in a network [42].

Full matching creates matching subsets containing one treated unit with one or more control units, or one control unit with one or more treated units [43]. *Optimal full matching* minimizes a weighted average of the estimated distance measure between pairs matched within each subclass, and it is available in the *matchit* package [41,44].

Genetic matching implements a more general distance measure based on the Mahalanobis distance, which incorporates a matrix of weights to be optimized over several generations. Optimization is carried out by minimizing a loss function over the matched sample at each iteration, identifying the corresponding weights until it converges asymptotically to the optimal solution [45].

Multiple measures of balance provide a more complete image of balance and the biases [46]. One such measure is the *standardized mean difference (SMD)* between treatment and control groups before and after matching. The *SMD* of each variable can be visualized with an *SMD* chart. If the absolute value of *SMD* is greater than 0.1, it indicates evidence of imbalance between the treatment and control groups. The *SMD* is not dependent on the unit of measure, and it is expressed for continuous covariates with the following equation [47]:

$$SMD = \frac{\bar{X}_1 - \bar{X}_0}{\sqrt{\frac{s_1^2 + s_0^2}{2}}} \tag{7}$$

For binary covariables, it is calculated using the proportions (\hat{p}_i) of their values, both in the treated group and the control group [48].

$$SMD = \frac{\hat{p}_1 - \hat{p}_0}{\sqrt{\frac{[p_1(1-\hat{p}_1) + \hat{p}_0(1-\hat{p}_0)]}{2}}} \tag{8}$$

The *variance ratio (VR)* compares the variances of the two matched groups. If this indicator is equal to 1, it means that the balance is good, and less than 2 is generally acceptable [31]. The hypothesis test and its corresponding *p*-value should be used with caution. The significance test can detect spurious statistically significant differences [28].

The **prognostic score** is defined as the predicted probability of an outcome under control conditions [49]. It has been demonstrated to be an excellent indicator of the degree of bias reduction, and it is more efficient than the covariate mean differences and significance tests [46], even when the prognostic scoring model is misspecified. For this, a regression model is fitted to predict the outcome using the control group covariates. Thus, using the same model, the outcome of both control and treatment group units is predicted. The absolute *SMD (ASMD)* of both groups is then compared [50].

Even if optimally matched records were found, it is possible that they do not coincide in their unobserved variables, and therefore, there may be a bias in the treatment assignment due to these factors [51,52]. *Sensitivity analysis* in observational studies is a mathematical calculation of the amount of unobserved covariate bias needed to change the conclusions of the study [33]. This mathematical calculation varies an underlying assumption to a statistical procedure to determine what magnitude of deviation from that assumption would be necessary to alter the conclusion [33]. *The Rosenbaum sensitivity test for Hodges—Lehmann point estimate* calculates the odds of differential assignment to treatment due to unobserved factors for a range of gamma values. It is an unconfounded estimate that reflects “the uncertainty of the hidden bias due to failure to control for un-

observed covariates” [52]. Better observational study design decisions result in better insensitivity to large unmeasured biases [51].

Cohen’s [53] *effect size* measure (d) is defined as the mean difference between the treatment and control groups divided by the population standard deviation. There are other more precise measures with various corrections of the standard deviation depending on the size of the groups and the (non)similarity of variances. However, the square root of the mean of the variances of both groups is frequently used as the population standard deviation. Thus, the effect size is similar to a *Z-score* and its formula is the same as the *SMD*. The *Z-score* is a standardized measure that allows us to make comparisons between different populations. In addition, Cohen defined three values for the value of d for the corresponding small, medium, and large effect sizes (0.2, 0.5, and 0.8, respectively) [53].

The proximity metrics are described in Table A3. The R libraries used in this paper are shown in Table A4.

In practice, many researchers have recommended building different matching models and choosing the one with the best fit of their covariates [33,54,55]. However, we have reviewed the literature (as of 30 September 2023), and, to our knowledge, only individual approaches exist for each specific problem, but we have not found generic methods to identify these models, which represents a knowledge gap.

In the absence of a general methodology to determine the best matching models, we developed in this paper a novel methodology to find them. For this purpose, we have run the most popular matching methods, providing balancing criteria and priorities to facilitate the identification and selection of the best models. The proposed methodology is presented in Section 4.3.

2.4. Statistical Models for Meta-Analysis

Meta-analysis is a statistical technique for integrating the findings from different comparative studies [56].

Two statistical models are most commonly used: the *fixed-effect model* and the *random-effect model*. The first assumes that a common effect size exists for each study and that their differences are due to sampling errors [57]. The second assumes that a distribution of true effects exists with different effect sizes for each study, and it focuses on calculating the parameters of this distribution [57]. The difference among studies can be due to background factors as well as the design or execution of the study. The random effects model nomenclature of DerSimonian-Laird and their formulas are shown in Tables A5 and A6, respectively [56].

3. Data

The production process we are concerned with has been summarized in the Introduction because it is the stochastic process that generates the available data.

The cardboard industry has two main production areas: corrugating and printing. The printing subsystem is composed of several printing lines. Each printing line is made up of several sequential modules or stages, for example, feeder, printer, die cutter, slotter, folder, gluer, stacker, or binder. These printing lines are also referred to as machines. The characteristics of each machine will depend mainly on the maximum number of colors it can print, the production volume, and the post-printing stages that compose it.

The industry works in three rotating shifts, 24 h a day, 7 days a week. The work teams are roughly similar in productivity and efficiency. During the production process of an order, there may be one or more stoppages, starting with the setup stoppage, and then, there may be zero or n stoppages due to scheduled or unscheduled operations. At the end of the production of an order, the wasted units are counted.

Data on production orders and downtime are stored in the organization’s ERP (enterprise resource planning) and alarm monitoring software, respectively.

The process begins when a production order is created in the sales department at the client’s request. Then, during the week before delivery, the scheduler assigns a date, shift, and machine on which the order will be produced, as well as the raw material and

personnel requirements. When the assigned day arrives, the production order enters the queue of the machine, according to the allocated position.

Before the machine starts processing the order, the raw material and personnel are provisioned, and the machine is guaranteed to be operational. The production order starts with the setup stoppage where the cliché of each color is loaded, color shade is adjusted, and print accuracy adjustments are made. Each subsequent stoppage may be related to the modules of the printing line.

We want to determine the effect of the total downtime (seconds) of a production order on its waste (linear meters). Preliminary data that could help answer the causal question were selected by eliminating incomplete orders (due to cancellations), inconsistent or flawed records, and production lines with insufficient records (less than 1000 records).

New variables are calculated from the existing variables. Additionally, aggregate variables are generated by production orders. All available variables that could be associated with the outcome are described in Table 1. The dataset contains 59,676 records of the production orders collected over three years (2019 to 2021).

Table 1. Dataset variables description.

Variable	Description	Datatype/Unit	Operation
MCH	Print subsystems (machine)	Categoric	-
PRICELAM	Sheet price	euros	-
MT2LAM	Sheet area	m ²	-
TEST	Box compression test	N	-
ORDEREDLM	Linear meters of an order	LM	-
SCHEDLM	Planned linear meters	LM	-
FINALLM	Final linear meters to send to client	LM	-
OVERPROCLM	Linear meters of overprocessing	LM	-
ASLEEPTIME	Time of the production order since it was created until it enters the machine queue	minutes	Calculated
QUEUETIME	Queue time of a production order	minutes	Calculated
PRODCETIME	Order production time	minutes	Calculated
MSETUP	Stoppage time for setting the order on the machine	seconds	-
MSUMALL	Sum of the time of all stoppages after setup of a production order (causal variable)	seconds	Calculated and aggregated
TMSUMALL	Treatment variable for causal variable	1, ..., 10	Calculated
WASTELM	Linear meters of waste	LM	Calculated

A more complex data set containing the individual stoppages of the different stages will not be used for the effect size calculations because it increases the number of variables and increases the complexity of the matching. However, we will address this data set in another study, for which we will need the collaboration of an expert in the process.

4. Methodology

To address how to use matching methods to measure the effect of a causal variable on the outcome variable, we propose the following steps outlined in Figure 1, which are described in the subsequent sections.

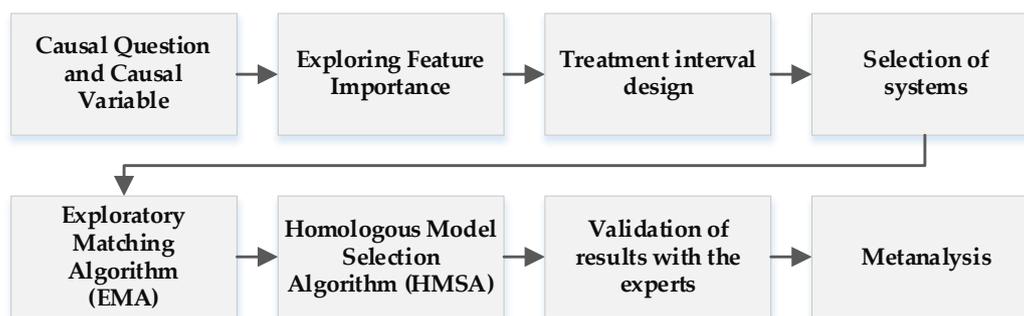


Figure 1. Exploratory matching model search algorithm (EMMSA) for causal inference.

4.1. Causal Questions and Causal Variable

To conduct a causal analysis of observational data, a causal question must first be established to guide the whole process [27]. For this, it is necessary to distinguish between causal and non-causal questions. Causal questions deal with the mechanisms of data generation or predictions of how a system will behave after an intervention [58].

Therefore, for the selection of the causal variable, the phenomenon is first analyzed in search of variables that allow an intervention, to produce the expected value in the outcome variable [23,59].

Then, to formulate the causal question, both the causal and outcome variables must be explicit in the question with the corresponding units of measurement. For example, “What is the effect size of the *causal variable* (units) on the *outcome variable* (units) in the system?”

4.2. Preparation of Dataset for the Matching Algorithms

4.2.1. Exploring Feature Importance

It is advisable to create a regression model (e.g., light gradient boosting [60]) to determine the importance of the predictor variables on the outcome variable. In addition, we could examine the order of importance of variables to check that the chosen causal variable is included. If not, it is advisable to review the causal question again by choosing an important variable on which an intervention is possible [27].

In the industrial environment, production order variables are strongly associated with the output variable (e.g., the higher the number of order units is, the higher the number of waste units is). They are not causal variables, but associative, because they are set by the client, and, therefore, we cannot modify them; i.e., they do not allow interventions.

The causal variable is not always the one with the highest level of association with the outcome variable. Some causal variables may have a low level of association. Likewise, causal variables exist that are key and trigger the increase (or decrease) in the outcome variable, but their level of association could be close to zero, so they might not be considered causal variables. Therefore, the causal variable may or may not be strongly associated with the outcome variable.

Thus, causal analysis is an adjusted analysis, because the covariates, whether causal or associative, are controlled to rule out their effects on the outcome variable so that the size of the causal effect can be measured directly with the minimum possible bias and variance.

Therefore, in observational studies, the causal variable is not always required to be a necessary and sufficient condition to yield the effect on the outcome variable. In most cases, it is not sufficient, or it may be sufficient but not necessary, because other variables may be causing the effect.

4.2.2. Treatment Interval Design

A first approach consists of measuring the effect of a causal variable that groups together many causes affecting an outcome variable. To analyze it, the causal variable is segmented into intervals, depending on the amount of data, to measure the effect produced by each of them. From this result, we could determine the intervals with the greatest impact

on the outcome variable to prioritize the search for its causes in the production process. This is because problems are often multi-causal, which is common in the manufacturing environment.

Each treatment (interval) is numbered from 0 to N , in ascending order of the causal variable. A control treatment is considered when the value of the causal variable is zero, and the treatment number assigned will be 0. Each treatment (from 1 to 10) will be contrasted with the control treatment (0). Then, the effect of treatment n will be the mean of the difference in potential outcomes between the two groups, treatment n and treatment 0.

But how many intervals should be defined? One alternative is to use a scatter diagram of the two variables (cause and outcome) and manually define the intervals of interest. Another alternative is to define the different intervals using percentiles and adjust their values, considering, as far as possible, at least 100 records per treatment to be able to detect the effect size with the minimum possible bias.

4.2.3. Selection of Systems (Machines)

Matching algorithms require sufficient records to function properly. For this reason, we scanned for the number of records for each production line and discarded those with insufficient records; for example, for 10 treatments, at least 1000 records are required for the printing line.

4.3. Exploratory Matching Algorithm

To assemble the samples as randomized experiments using the covariates prior to the causal variable, several popular matching algorithms are used, such as nearest, optimal, full, genetic, and coarsened exact matching (Section 2.3).

Algorithm 1 shows the exploratory matching algorithm (EMA). It begins by defining the formulas for matching and random forest (RF) models, as well as defining the lists of treatments, the matching methods, and the distance measures used. Then, it proceeds to set up four nested loops for the different machines, treatments, methods, and distances to be used in the matching methods. This algorithm is very brief as it does not contain the R-specific programming details but illustrates how to proceed. Before the end of each iteration, different tests are performed to collect the balancing information for subsequent analysis.

All model and test summaries are saved in a text file. Another algorithm in Python extracts the interest variables for its posterior analysis which are saved in a CSV file. These variables are machine; treatment; matching algorithm; distance; binning; control units; treatment units; matched control units; matched treatment units; variance of treatment group; and variance of control group. From the predictions of the RF model, the explained variance and the size effect (the mean difference for each group) are obtained, as are the following: the number of non-balanced variables whose standard mean differences are greater than 0.1; the number of non-balanced variables whose variance ratios are greater than 2; gamma and upper bound of the Rosenbaum sensitivity test when the upper bound is greater than 0.05; the F-statistic, p -value, and ratio of variances of the F-test; the t , p -value, confidence interval, and mean in group C and T of the Welch two-sample t -test; and the unconfounded estimate of the Rosenbaum sensitivity test for Hodges-Lehmann.

Algorithm 1. Exploratory matching algorithm (EMA).

```

Data:  $\delta = \text{ProductionOrders}$ 
 $mf \leftarrow \text{MatchingFormula};$ 
 $rff \leftarrow \text{RFFormula};$ 
 $\tau \leftarrow \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\};$ 
 $S \leftarrow \text{MachineList};$ 
 $M \leftarrow \{\text{Nearest}, \text{Fullmatch}, \text{Optimal}, \text{Genetic}, \text{CEM}\};$ 
 $D \leftarrow \{\text{Bart}, \text{Euclidean}, \text{Gam}, \text{Gbm}, \text{Glm}, \text{Mahalanobis}, \text{Nnet},$ 
   $\text{RobustMahalanobis}, \text{ScaledEuclidean}\};$ 
 $DC \leftarrow \{\text{Binary}, \text{Camberra}, \text{Euclidean}, \text{Mahalanbis}, \text{Maximum},$ 
   $\text{Minkowski}, \text{RobustMahalanbis}, \text{ScaledEuclidean}\};$ 
 $B \leftarrow \{\text{scott}, \text{sturges}, \text{q4}, \text{q5}\};$ 
foreach  $s \in S$  do
  foreach  $t \in \tau$  do
    foreach  $m \in M$  do
      foreach  $d \in D$  do
        if  $m == \text{CEM}$  then
          foreach  $b \in B$  do
             $\text{Matching}(mf, \delta[0], \delta[t], \text{method} = m, \text{distance} = d, \text{bin} = b);$ 
          end
        else
           $\text{Matching}(mf, \delta[0], \delta[t], \text{method} = m, \text{distance} = d);$ 
        end
         $\text{RF\_Model}(rff, \delta[0_M], \delta[t_M]);$ 
         $\text{Calculate Size Effect using RF Model};$ 
         $\text{Perform Rosenbaum Sensitivity Test};$ 
         $\text{Perform F Test to compare variances};$ 
         $\text{Perform Hodges Lehmann Point Estimate};$ 
         $\text{Print test summaries};$ 
      end
    end
  end
end

```

4.4. Homologous Model Selection Algorithm (HMSA)

The CSV file of the matching results contains the relevant variables that we could use to select the matching models that best balance our data for each machine-treatment. Then, HMSA processes this file, calculates the Z-Score, and proceeds to find the register of matching results with the lowest number of unbalanced variables and with the highest number of matched units. To accomplish this, we define two conditions to validate models. These are responsible for assigning balancing priorities to each matching model. A third condition is set to search for models with higher priority in both condition 1 and condition 2, and then, at the machine-treatment level, the mean Z-score is calculated and those models whose Z-score is greater than 95% of the mean are selected. These models selected are considered *homologous* due to their similar characteristics. Next, the selection of the better matching method is chosen from homologous models for each machine-treatment. Algorithm 2 summarizes the HMSA algorithm with its conditions and priorities.

Algorithm 2. Homologous model selection algorithm (HMSA).

Data: $\delta = \text{MatchingResults}$
 Calculate : $ZScore$;
 $NVar = \text{Number of variables}$;
 $nbSMD = \text{Number of unbalanced variables by SMD}$;
 $nbVR = \text{Number of unbalanced variables by Variance Ratio}$;
 $uMT = \text{Matched Treatment Units}$;
 $uMC = \text{Matched Control Units}$;
foreach $s \in S$ **do**
 foreach $t \in \tau$ **do**
 foreach $m \in M$ **do**
 $\lambda \leftarrow \delta[s, t, m]$;
 foreach $reg \in \lambda$ **do**
 Condition 1:Unbalanced variables;
 $con1 = 1, \text{if } nbSMD \text{ and } nbVR = 0$;
 $con1 = 2, \text{if } nbSMD \text{ xor } nbVR = 0 \text{ and the another } \leq 0.2NVar$;
 $con1 = 3, \text{if } nbSMD \text{ and } nbVR \text{ are } \leq 0.2NVar$;
 $con1 = 9, \text{if it do not satisfy the above criteria}$;
 Condition 2:Matched Units;
 $con2 = 1, \text{if } uMT \text{ and } uMC \text{ are maximum}$;
 $con2 = 2, \text{if } uMT \text{ xor } uMC \text{ is max and the another } \geq 0.9umax$;
 $con2 = 3, \text{if } uMT \text{ and } uMC \text{ are } 0.9umax$;
 $con2 = 9, \text{if it do not satisfy the above criteria}$;
 end
 end
 Condition 3:Homologous and Selected Cases;
 $\rho \leftarrow \lambda[con1 \leq 3, con2 \leq 3]$;
 $ZScoreMin \leftarrow 0.95(\text{mean}(\text{abs}(\rho[\min(\text{cond1}), \min(\text{cond2})][ZScore])))$;
 foreach $reg \in \rho$ **do**
 if $reg[\text{cond1}] == \min(\rho[\text{cond1}])$ **then**
 if $reg[\text{cond2}] == \min(\rho[\text{cond2}])$ **then**
 $H \leftarrow 1; \text{Homologous if } ZScore \geq ZScoreMin$;
 else
 $H \leftarrow 0$;
 end
 else
 end
 end
 $\gamma \leftarrow \delta[H = 1]$;
 foreach $MachineTreatment \in \gamma$ **do**
 $S \leftarrow 1; \text{The case with max}(ZScore) \text{ of homologous registers.}$;
 $EffSize \leftarrow \text{mean}(\text{effetSize of } MachineTreatment)$;
 end
 end
 end
end

4.4.1. Condition 1—Unbalanced Variables

The first condition has three priorities. Priority 1 is assigned to models that have zero unbalanced variables in both SMD and VR. Otherwise, priority 2 is assigned if zero unbalanced variables exist in SMD and up to 20% unbalanced variables exist in VR, or zero unbalanced variables exist in VR and up to 20% unbalanced variables exist in SMD. If the two previous cases do not occur, priority 3 will be assigned if the models present up to 20% of unbalanced variables both in SMD and VR. If the above three cases do not occur, priority 9 will be assigned to models with an excessive number of unbalanced variables.

4.4.2. Condition 2—Matched Units

The second condition also has three priorities. Priority 1 is assigned to models whose matched units in the treatment and control groups are equal to the maximum number of treatment and control units, respectively, i.e., no treatment units have been lost, and the number of control-matched units is at least the same as the number of treatment units. Otherwise, priority 2 is assigned to models with a maximum number of matched units in the treatment group and with matched units in the control group greater than or equal to 90% of the control units, or vice versa. If the above two cases do not occur, priority 3 is assigned to models with a number of matched units greater than or equal to 90% of units in both the control and treatment groups. If the above three cases do not occur, priority 9 is assigned to indicate the loss of a significant number of treatment or control units.

4.4.3. Condition 3—Homologous and Selected Models

Matching models whose priorities of condition 1 and condition 2 are less than or equal to 3 are found, thus discarding all models that do not present a satisfactory balance. If no records remain after filtering for a machine-treatment, it is assumed that a satisfactory model does not exist. Then, the highest priority models of condition 1 (minimum value) are chosen. From these registers, the highest priority registers of condition 2 (minimum value) are chosen. Then, for each machine-treatment, the models whose Z-score exceeds 95% of the mean Z-score are considered *homologous models* because they have similar characteristics and represent the best models that capture the largest effects. From the homologous models, the model with the maximum Z-score is selected as the best model for the respective machine-treatment.

In addition, two values are obtained as effect size for each machine-treatment, and the first one will be obtained from the selected model (Z-score), and the second one will be the mean of the effect size of the homologous models with their respective confidence intervals.

The results of homologous models, including the selected model, are exported to Excel for visualization and analysis. The EMA is implemented using the R libraries listed in Table A4. The HMSA is implemented in Python. Both algorithms are available for consultation upon reasonable request by any reader. However, it will not be possible to establish comparisons of this methodology with other existing methodologies because such comparisons do not exist, as already indicated in the Literature Review section.

4.5. Validation of Results with the Expert

The results obtained from the quantitative models will be validated with the qualitative knowledge of the expert in this process. In addition, the results are contrasted with a data analysis of the outcome variable.

4.6. Meta-Analysis

Meta-analysis techniques are used to make comparisons of the effect of the causal variable on the outcome variable in a subsystem. For this purpose, the “meta” library in R was used to develop another algorithm that integrates these results and determines, in a subsystem, the contribution of each treatment to the outcome variable. Similarly, for each treatment, a meta-analysis is performed to determine the contribution of the causal effect in each subsystem to the outcome variable. Since these are standardized measures, it allows these comparisons to be made.

5. Results and Discussion

5.1. Causal Questions and Causal Variable

The causal variable chosen in the Introduction to formulate the causal question is the total downtime (MSUMALL) of a production order. This variable can be intervened in since it can be changed by using different strategies. The main causes of these stoppages can be due to inefficient processes, operator errors, lean inventory of materials, or inefficient

machine maintenance. By addressing these causes, downtimes can be reduced, and at the same time, waste can also be reduced as it is likely that the machine will be stopped when it produces defective units. Intuitively, the downtime could be directly proportional to the error severity.

The outcome variable is WASTELM since we are concerned about its behavior and intend to control it through MSUMALL to reduce costs and increase profitability.

For this reason, these two variables are crucial in the analysis of productivity, and it becomes valuable to detect the time intervals in which waste is higher. Thus, with an appropriate intervention according to the error type, it will be possible to change the causal variable to the lowest possible intervals (treatments) where the waste is smaller.

5.2. Exploring the Feature Importance

A regression was performed using the LightGBM library. The most satisfactory model was found using a hyperparameter grid, and its values are shown in Figure 2, which illustrates that among the important features associated with the outcome variable (WASTELM) are (i) OVERPROCLM, which indicates the linear meters of overproduction, either by a small safety percentage to cover possible defective units not observed by the naked eye, or by replacement of non-conforming units; (ii) M2LAM, which is a product specification; (iii) SCHEDLM, which is determined in planning, and its value is slightly higher than the linear meters specified by the client in the production order; (iv) ORDEREDLM, which represents the linear meters of the production order; (v) PRODTIME, which depends on the downtimes of the current order; (vi) ASLEEPTIME, which depends on the order date and its delivery date; (vii) FINALLAM, which represents the linear meters of production (without waste) shipped to the client; (viii) MSUMALL, which depends on the downtime of all stoppages after MSETUP; (ix) MSETUP, which depends on the operator and the order complexity; and (x) QUEUETIME, which depends on the downtime of previous orders. However, OVERPROCLM and FINALLAM do not directly cause waste but are its effects.

Therefore, the causes of waste could somehow be reflected in the total downtime, which will be confirmed by this study. However, it is previously necessary to segment MSUMALL to generate the treatment variables (TMSUMALL). Thus, with the proposed methodology, we will be able to identify for each machine the time interval where the greatest waste is generated in order to subsequently identify with the expert the possible causes that explain it.

Then, the covariates that we should control for matching algorithms are all those that occur prior to the treatment variable (TMSUMALL), such as MT2LAM, ORDEREDLM, SCHEDLM, QUEUETIME, ASLEEPTIME, TEST, and MSETUP. The covariate TEST is included because it is important to control the carton strength, as it is somehow an indicator of the thickness and would help to obtain more similarly matched units. MSETUP is considered because it occurs before the stoppages to be analyzed.

Equation (9) shows the matching formula that will be used to build the models controlling the TMSUMALL variable with the variables that precede it and which are not causes of WASTEML.

$$\text{TMSUMALL} \sim \text{MT2LAM} + \text{ORDEREDLM} + \text{SCHEDLM} + \text{QUEUETIME} + \text{ASLEEPTIME} + \text{TEST} + \text{MSETUP} \quad (9)$$

PRODTIME, OVERPROCLM, PRODLM, and FINALLM are involved a posteriori and will be included in the random forest regression model to measure the effect size of each model. Equation (10) shows the formula for building random forest regression models in which participate all predictors variables. In the end, after selecting the best models, it will be checked whether the effect size is similar between the simple mean of the two groups and the mean of the difference of the predictions of the RF models in both groups.

$$\begin{aligned} \text{WASTELM} \sim & \text{PRICELAM} + \text{MT2LAM} + \text{ORDEREDLM} + \text{SCHEDLM} + \\ & \text{FINALLM} + \text{QUEUETIME} + \text{ASLEEPTIME} + \text{PRODTIME} + \text{MSETUP} \\ & + \text{MSUMALL} + \text{OVERPROCLM} + \text{TEST} \end{aligned} \quad (10)$$

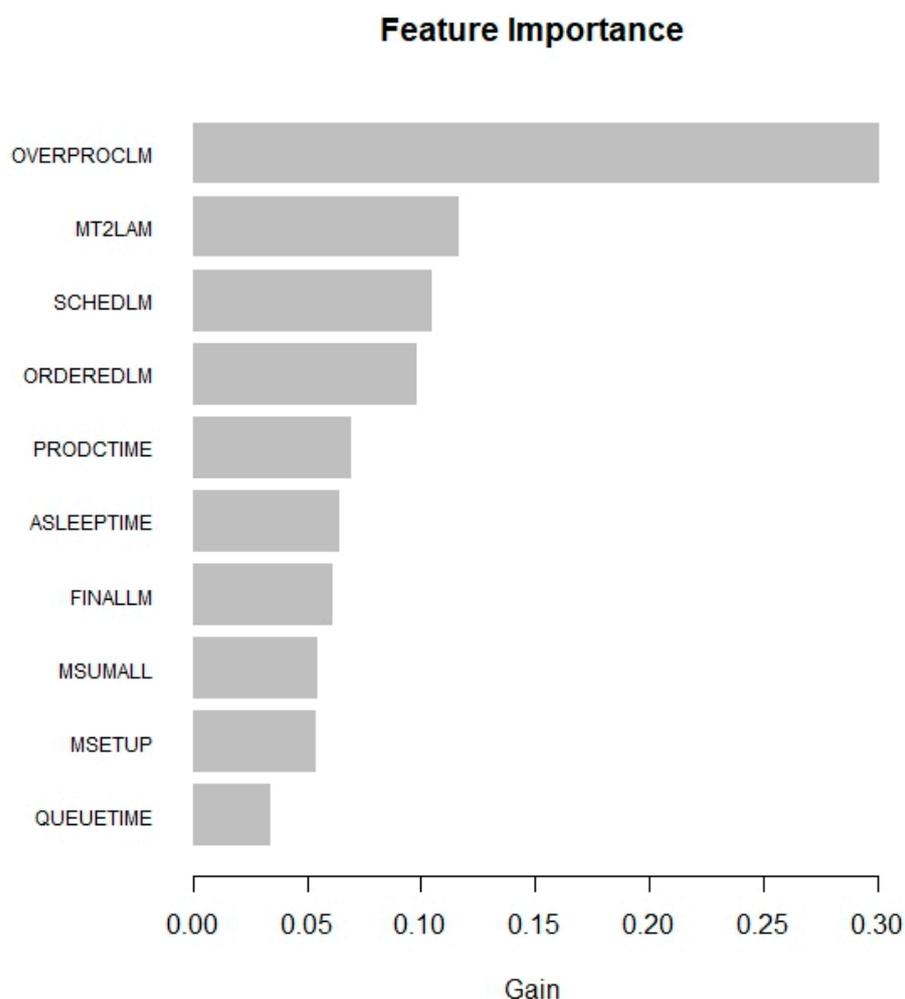


Figure 2. Feature importance for WASTELM. **Note: Parameters for LightGBM Regression.** max_depth = 5, num_leaves = 28, num_iterations = 200, early_stopping_rounds = 40, and learning_rate = 0.46.

5.3. System Selection and Treatment Interval Construction

The treatment variable was generated by splitting the MSUMALL variable, considering all machines with more than one thousand records (machines with an insufficient number of records were excluded). The treatments correspond to each decile, considering all the records for their calculation to obtain the same analysis interval for all the machines. Records whose MSUMALL is zero are considered the treatment group (0). Table A7 shows a summary of the number of records and the treatment interval (in seconds) for each machine. Figure A1 shows the MSUMALL density function for each machine. Figure A2 shows the WASTELM density function for each machine.

5.4. EMA and HMSA Algorithms

Algorithm 1 (EMA), which executes the different matching methods with the various distance metrics selected, was run and 7691 matching models were obtained. Then, Algorithm 2 (HMSA), which performs an exhaustive search for the best models to be considered as homologous, was run, and 2644 models (34%) fulfilling conditions 1 and 2 were obtained, this being the first filter.

The results of the application of Algorithm 2 (HMSA) are summarized in Table 2. It shows a summary of all models found that meet condition 1 (rows) and condition 2 (columns) showing their respective priorities.

Table 2. Model number resulting of the application of HMSA by conditions and priorities.

Condition 1 Condition 3: Homol.	Condition 2: Methods and Priorities									Total Cond1	% Total
	CEM	FullMatch		Genetic		Nearest		Optmatch			
	2	1	2	3	1	1	2	1			
Priority 1	176	80	57	130	839	318	86	189	1875	71%	
0	175	38	57	130	439	158	86	85	1168	62%	
1	1	42			400	160		104	707	38%	
Priority 2	24	72	17	41	183	155	32	100	624	24%	
0	19	69	17	41	179	153	32	98	608	97%	
1	5	3			4	2		2	16	3%	
Priority 3		4	2	4	71	36	8	20	145	5%	
0		4	2	4	71	36	8	20	145	100%	
I. Total by priorities	200	156	76	175	1093	509	126	309	2644	100%	
II. Total by methods	200		407		1093		635	309	2644		
III. % Models by meth.	7.6%		15.4%		41.3%		24.0%	11.7%	100%		
IV. Homol. models	6		45		404		162	106	723		
V. % Homol. models (II)	3.0%		11.1%		37.0%		25.5%	34.3%	27.3%		
VI. % Homol. models	0.8%		6%		55.9%		22%	15%	100.0%		
VII. % Homol. Models of total	0.2%		1.7%		15.3%		6.1%	4.0%	27.3%		
VIII. Selected models	0		15		39		8	8	70		
IX. % Selected models by meth.	0.0%		21.4%		55.7%		11.4%	11.4%	100.0%		

Note: Condition 1 and 2 can have up to three priorities each one. Condition 3 can be 0 (for discarded models) or 1 (for homologous models). Total values are shown in bold, either by conditions, priority, methods, homologous models or selected models.

In addition, for each priority of condition 1, there can be two rows to indicate the fulfillment of condition 3. If it is 1, it indicates the number of models that fulfill condition 3, which we will call *homologous models*, or 0 indicates the number of models that do not fulfill this condition, and therefore, they are consequently discarded. Condition 3 acts as the second filter.

For example, the HMSA algorithm found 839 models from the genetic method that met condition 1—priority 1 and condition 2—priority 1, of which 400 models were considered homologous because they also met condition 3, and 439 were discarded. In addition, it found 183 genetic models that fulfilled condition 1—priority 2 and condition 2—priority 1, of which 4 models were considered homologous because they also met condition 3 and 179 were discarded.

The bottom part of Table 2 shows a summary of models filtered by HMSA. For example, it found 1093 genetic models that met conditions 1 and 2 (II), and they represent 41.3% (III) of the 2644 models that met both conditions (1 and 2) for any priorities. In addition, it found 404 homologous models (IV), and they represent 37% of the total genetic models (V), 56% of the models found (VI) from all homologous models (723), and 15.3% of the total models (VII) fulfilling conditions 1 and 2. Finally, it selects for each machine–treatment 39 genetic models that represent 56% of the 70 models selected from all methods.

Therefore, 723 homologous matching models fulfilling conditions 1, 2, and 3 were obtained (IV), which corresponds to 27% of the models (V) that meet conditions 1 and 2 (2644).

In other words, only 9% of all generated models (7691) fulfill the three expected conditions of having almost all variables balanced (condition 1), the maximum number of treatment and control units (condition 2), and an effect size greater than 95% of the mean (condition 3). The remaining percentage of models is discarded because the models have a priority value greater than 1 in conditions 1 and 2 or because they do not meet condition 3, which acts as a second filter.

The results of the application of Algorithm 2 (HMSA) are plotted by methods (Figure 3) and distances (Figure 4). Thus, Figure 3a and Table 2 (V and VI) show that most of the homologous matching models are selected by the genetic (56%), nearest (22%), and optimal (15%) methods. Figure 3a shows that the most selected distances in the homologous models are scaled Euclidean (33%), GAM (15%), GBM (15%), Mahalanobis (12%), and GLM (10%).

Methods and distances by machine: Figure 3b shows that the genetic, nearest, and optimal methods find homologous models on all machines. Fullmatch does not find homologous models in M3, and CEM only finds them in M1. More models are obtained in M4 because it has a larger number of records, which could facilitate matching.

Figure 4b shows that the genetic method obtains an almost similar number of models for each distance measure, whether it uses a caliper or not. Figure 4c shows that the scaled Euclidean and Mahalanobis distances have most of the homologous models across all machines.

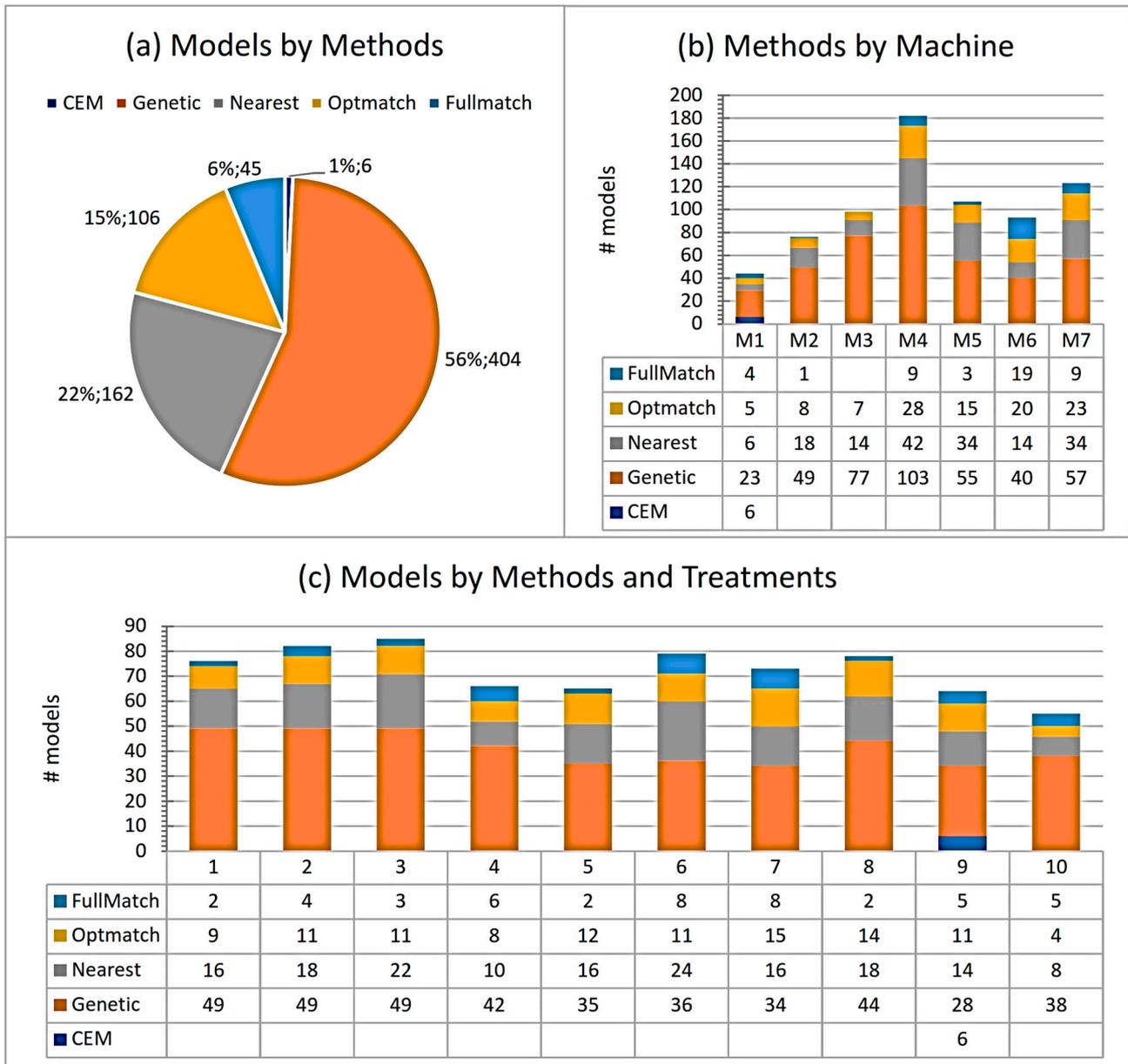


Figure 3. Homologous models by methods, machine, and treatment.

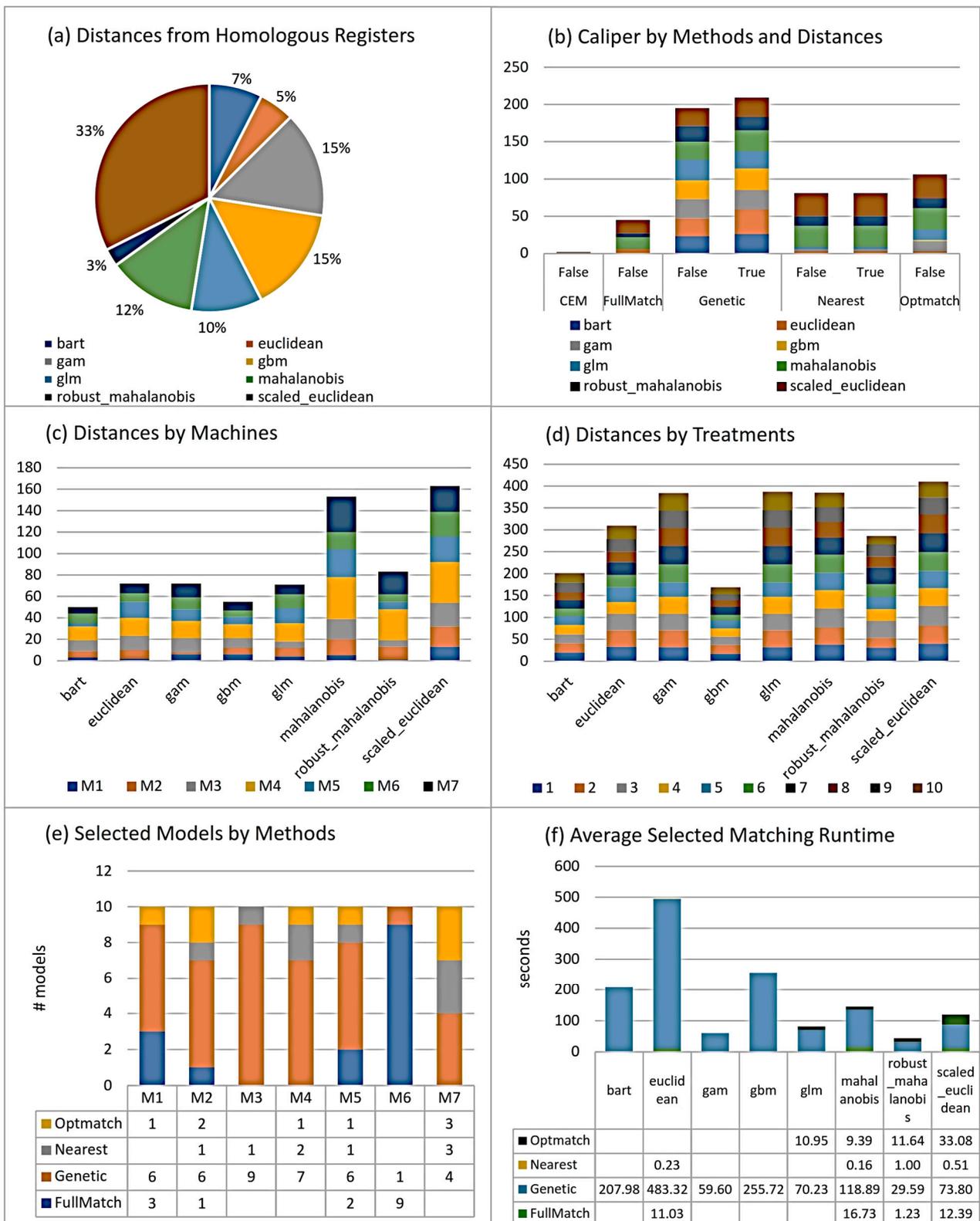


Figure 4. Homologous models by distances, methods, machine, and treatment.

Therefore, the genetic method is the one that generates more homologous models due to its iterative process to automatically evaluate the balance and optimize the search for a better solution. For this, it uses a generalized version of the Mahalanobis distance that incorporates a matrix of weights to be adjusted in each iteration to minimize the loss

function. At each iteration, it produces a new generation and converges asymptotically [45]. However, the disadvantage of the genetic method is the mean execution time with any distance, as shown in Figure 4f, and the Euclidean, GBM, and Bart distances are the most time-consuming.

Methods and distances by treatment: Figure 3c shows that genetic, nearest, optimal, and fullmatch methods find homologous models on all treatments. Figure 4d shows that the scaled Euclidean, GAM, GLM, and Mahalanobis distances have the most homologous models across all treatments.

Therefore, the genetic method with Mahalanobis and scaled Euclidean distances ensures finding homologous models for all machines and all treatments, with or without caliper, as shown in Figure 3b,c and Figure 4b–d.

The run time was approximately 140 h on an Intel Core I7 Fourth Generation computer with 12 Gb RAM (1600 Mhz) and an SSD hard disk (560 Mb/s). The matching library used is *matchit*. For the genetic method, the maximum number of generations is 100, the maximum unchanging generational is 4, and the convergence tolerance is 0.001. More details about this algorithm can be found in [45].

5.5. Effect Size

Finally, HMSA chooses one model among homologous models for each machine-treatment. The model chosen is the one with the highest Z-score to indicate the largest effect captured from the corresponding homologous models. Then, HMSA calculates the mean effect size of these homologous models and their confidence interval (CI).

The results of the selected models for each machine-treatment are shown in Table A8, and their effects are plotted in Figure 5a,b. Almost all selected models obtained priority 1 in conditions 1 and 2, indicating that all variables were balanced, with no loss of treatment units, and with the maximum number of control units (or at least equal to the number of treatment units). The exceptions appear in machine M1, treatments 8, 9, and 10, in which there was an unbalanced variable. Furthermore, it is observed in Table A8 that the mean of the effects calculated as difference of means (EffMD), and the values of the differences of the means of the predictions using RF models (EffRF) are approximately similar, which confirms the low dependence of the models.

In addition, the fullmatch method detects the largest effects because it incorporates almost all of the control units, increasing the variability of mean differences and capturing larger Z-scores. However, it does not find balanced models in all treatments, which are covered by the models of the other methods.

Therefore, the genetic, fullmatch, optimal, and nearest methods guarantee to find models with the highest Z-score on each machine and treatment, as also is shown in Table 2 (VIII y IX). In addition, the scaled Euclidean, Mahalanobis, GAM, robust Mahalanobis, and Euclidean distances are used in 80% of the selected models, as shown in Figure 6.

Table 3 summarizes the Z-score greater than 0.30 and the mean effect of homologous models greater than 20 LM. Then, we can determine that, in standardized terms, the three higher Z-scores are in M4-T10, M5-T10, and M6-T10. In linear meters, wastes greater than 20 linear meters are generated in M1-T10, M4-T10, M5-T10, and M6-T10. Therefore, treatment 10 is the one that generates the most waste on machines M1, M4, M5, and M6.

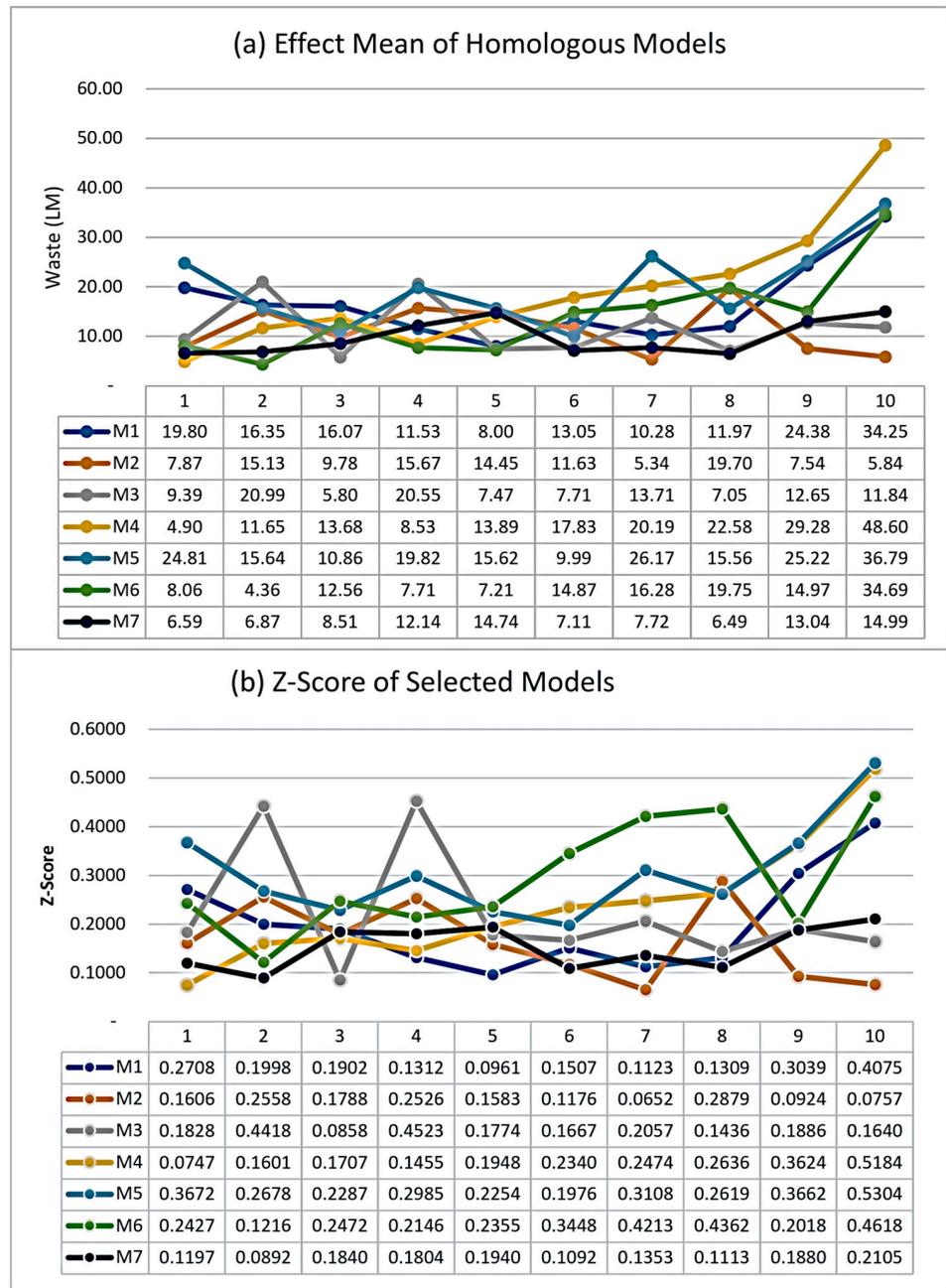


Figure 5. Mean effect size of homologous models and Z-scores of selected models.

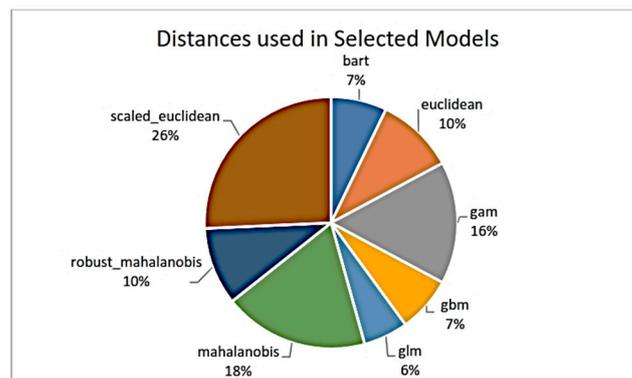


Figure 6. Distances used in selected models.

Table 3. Selected models: Z-scores and EffMeans greater than 0.30 and 20 LM, respectively.

Selected Models (Z-Score > 0.30)				Selected Models (EffMean > 20 LM)			
MCH	Treat	EffMD	Z-Score	MCH	Treat	EffMD	Z-Score
M1	T9	24.38	0.30	M1	T9	24.38	0.30
M1	T10	34.25	0.41	M1	T10	34.25	0.41
M3	T2	20.99	0.44	M3	T2	20.99	0.44
M3	T4	20.55	0.45	M3	T4	20.55	0.45
-	-	-	-	M4	T7	20.19	0.25
-	-	-	-	M4	T8	22.58	0.26
M4	T9	29.28	0.44	M4	T9	29.28	0.44
M4	T10	48.60	0.52	M4	T10	48.60	0.52
M5	T1	24.81	0.37	M5	T1	24.81	0.37
M5	T7	26.17	0.31	M5	T7	26.17	0.31
M5	T9	25.22	0.37	M5	T9	25.22	0.37
M5	T10	36.79	0.53	M5	T10	36.79	0.53
M6	T6	14.87	0.34	-	-	-	-
M6	T7	16.28	0.42	-	-	-	-
M6	T8	19.75	0.44	-	-	-	-
M6	T10	34.69	0.46	M6	T10	34.69	0.46

5.6. Overall Effect Size

To find out which machine generates more waste in statistical terms, we integrate the different studies of each machine and treatment using the meta-analysis techniques described in Sections 2.4 and 4.6. For this purpose, the meta-library has been used and, for the sake of brevity, we only present the M6 machine and treatment 10, which is where more waste is generated in standardized terms, as shown in Figure 7a,b.

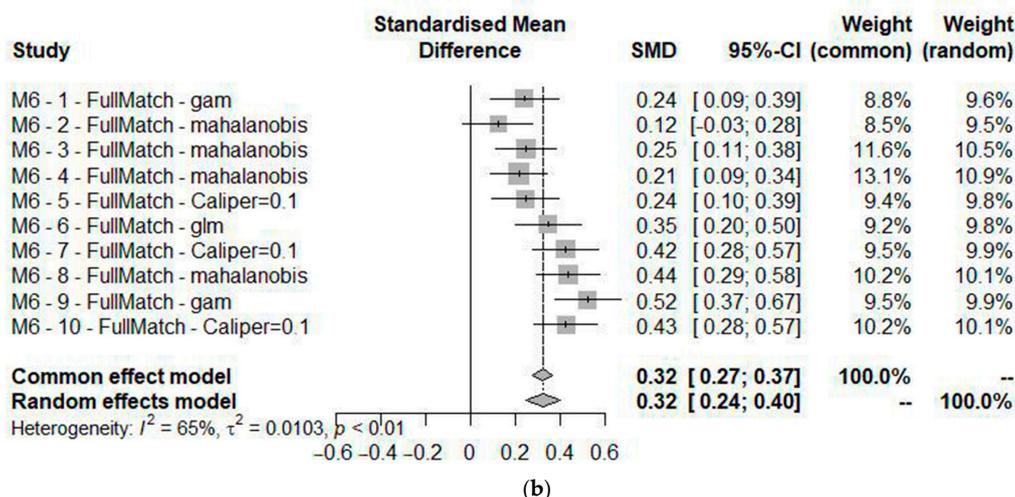
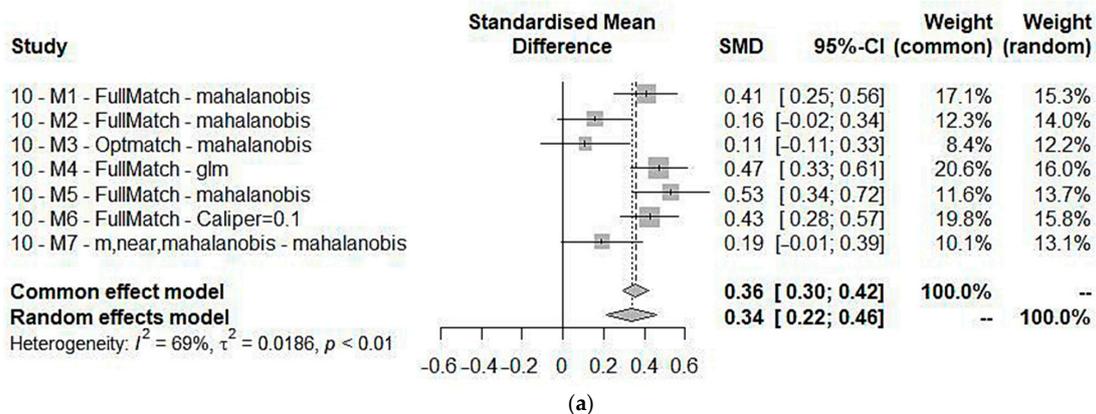


Figure 7. Meta-analysis for (a) machine 6 and (b) treatment 10.

The results of the meta-analysis are summarized in Table 4. It shows the random effect (RandEff) among studies, as well as their maximum effects found. Columns I^2 and τ^2 indicate the among-study heterogeneity, measured as the percentage of the total variance (I^2) that is explained by the between-study variance (τ^2). The “Max” column indicates, for each machine (treatment), the treatment (machine), method, and distance that produce the maximum mean effect measured in SMD or EffAvg LM, to which their respective confidence intervals are appended (95%-CI and 95%-CI LM).

Table 4. Meta-analysis for each machine and treatment with its maximum effect.

Meta-Analysis for Each Machine					Maximum Effect for Each Machine				
MCH	RandEff	95%-CI	I^2	τ^2	Max (Treat)	SMD	95%-CI	EffAvg LM	95%-CI LM
M1	0.18	[0.09;0.26]	37%	0.0067	T10-Full-Mah	0.41	[0.25;0.56]	34.25	[12.53;55.97]
M2	0.11	[0.03;0.19]	0%	0.0001	T8-Near-Mah	0.26	[0.02;0.50]	19.70	[18.18;21.22]
M3	0.07	[−0.01;0.16]	0%	0.0000	T3-Full-Calip	0.30	[−0.02;0.61]	5.80	[5.99;5.61]
M4	0.20	[0.13;0.28]	76%	0.0108	T10-Full-Glm	0.47	[0.33;0.61]	48.60	[14.48;82.72]
M5	0.23	[0.10;0.36]	70%	0.0271	T10-Full-Mah	0.53	[0.34;0.72]	36.79	[13.19;60.39]
M6	0.32	[0.24;0.40]	65%	0.0103	T9-Full-Gam	0.52	[0.37;0.67]	14.97	[4.32;25.62]
M7	0.07	[−0.01;0.15]	44%	0.0072	T5-Near-Mah	0.19	[0.04;0.35]	14.74	[12.41;17.07]
All Mchs.						0.38		24.98	[13.40;36.23]
Meta-Analysis for Each Treatment					Maximum Effect for Each Treatment				
Treat	RandEff	95%-CI	I^2	τ^2	Max	SMD	95%-CI	EffAvg LM	95%-CI LM
1	0.11	[0.01;0.21]	40%	0.0055	M1-Opt-Mah	0.27	[0.05;0.49]	19.80	[16.13;23.47]
2	0.13	[0.07;0.20]	0%	0.0000	M3-Full-Calip	0.30	[−0.02;0.61]	20.99	[21.69;20.29]
3	0.16	[0.09;0.22]	5%	0.0001	M6-Full-Mah	0.25	[0.11;0.38]	12.56	[6.53;18.59]
4	0.14	[0.07;0.20]	0%	0.0001	M6-Full-Mah	0.21	[0.09;0.34]	7.71	[4.77;10.65]
5	0.16	[0.10;0.22]	10%	0.0001	M6-Full-Calip	0.24	[0.10;0.39]	7.21	[4.51;9.91]
6	0.13	[0.01;0.26]	67%	0.0181	M6-Full-Glm	0.35	[0.20;0.50]	14.87	[6.37;23.37]
7	0.18	[0.03;0.33]	79%	0.0313	M6-Full-Calip	0.42	[0.28;0.57]	16.28	[5.81;26.75]
8	0.21	[0.10;0.31]	60%	0.0120	M6-Full-Mah	0.44	[0.29;0.58]	19.75	[6.28;33.22]
9	0.22	[0.02;0.42]	87%	0.0544	M6-Full-Gam	0.52	[0.37;0.67]	14.97	[4.32;25.62]
10	0.34	[0.22;0.46]	69%	0.0186	M5-Full-Mah	0.53	[0.34;0.72]	34.69	[12.44;56.94]
All Treat.								16.88	[7.99;25.77]

τ^2 is the variance of the random effects and has a very relevant value because it indicates the importance of each machine for the differences in waste. For example, 0.0271 is an important variance concerning the other variances. This indicates great variability exists between machines, and some machines have an inherent effect on waste. Furthermore, we did not eliminate or correct for this variability in the matching process. Therefore, there must be other variables in the M5 machine that have more influence on the waste.

In Table 4, the column “EffAvg LM” is the mean of the maximum effects in linear meters corresponding to the SMD value. The “All Mchs” row is weighted average of the effects of the treatments of each machine in which the maximum total wastage is reached. Thus, the maximum average waste for each production order that has stoppages after MSETUP is 24.98 LM with a confidence interval of [13.40;36.23].

In addition, we observe that the selected models of machines M2 and M3 present similarity since their I^2 is 0%. The same occurs with treatments 2 and 4. In these cases, their mean random effects are approximately equal in normalized terms.

In contrast, it is observed that the selected models of machines M4, M5, and M6 present greater differences, which indicates the heterogeneity of the studies, and therefore, their means and CIs are different. The same is true for treatments 6 to 10. In addition, it is observed that the fullmatch method detects the largest effects in 71% of the machines and 90% of the treatments.

5.7. Advantages and Disadvantages of This Methodology

An advantage of this methodology is that it measures the causal effect of downtime on waste. In addition, this methodology could be used to study with the experts the causes that reduce productivity and overall equipment effectiveness (OEE). Therefore, it is important

to address the downtimes to increase machine availability, avoid their deterioration and improve the performance, efficiency, and quality of the printing line. This way of studying the impact of downtime on waste is justified because it addresses two crucial issues at the same time: waste as a function of downtime, which will help reduce production costs and improve the profitability of the organization.

The process followed to obtain the models to which the required strong restrictions apply means that, in percentage terms, this number of models is small, although in absolute numbers it is still a very large number of resulting models.

For this reason, the three HMSA conditions are so restrictive that models could be obtained whose causal effect sizes, for the same machine-treatment, have slight variations.

Condition 3 (effect size greater than 95% of the mean) selects models that capture large effects, so it is possible that other more moderate effects are being neglected in this situation.

However, the choice to stay with models that meet all the constraints and are therefore strongly consolidated gives value to these models as it allows us to detect the variability caused by waste due to downtime in very extreme situations.

If, with these restrictions, homologous models are not found in some machine-treatment, it means that no reliable models were found to measure the corresponding effect size.

The main limitation of this study is that it does not measure the effect of the stoppages of each stage of the production line of each machine, which will be addressed in a future study, but it serves the purpose of an exploratory search for the main causes using the matching models.

Another disadvantage is that the effect size is calculated assuming that the standard deviations are different, and the group sizes are similar (Equation (7)). For this reason, in cases where the sample size and the standard deviation of each group are markedly different, it will be necessary to correct the standard deviation of the population. This may occur mainly with fullmatch, which considers more control units.

In any case, the differences detected when the standard deviation correction is performed would produce effects analogous or slightly underestimated to the uncorrected effect measurements. Equation (7) has more power to measure the effect because it has more type 1 errors. Therefore, the measured effects of the fullmatch models, using Equation (7), provide sufficient reliability to be considered.

6. Conclusions

In this work, the EMMSA methodology has been developed to measure, for each machine, the size of the causal effect of downtime on waste. As mentioned in the Introduction, being able to measure this effect could contribute to studying, quantitatively and causally, one of the main problems of the manufacturing industry, since it affects productivity and profitability.

This methodology allows a rigorous comparison between the different models generated by the matching methods used to perform the causal analysis.

For this purpose, the EMMSA is carried out with two developed algorithms: Algorithms 1 (EMA) and 2 (HMSA). Both have the advantage of performing an exploratory search for the best models to ensure, as far as possible, comparative groups with all covariates balanced, the maximum number of treatment and control units, and an effect size greater than 95% of the mean.

In the applied case, the machines and treatments where most waste is generated are shown in Table 3. Using meta-analysis methods, it was determined that M5 and M6 machines and treatment 10 generate the most waste in terms of standardized measurements. As such, the meta-analysis method helps us to prioritize, at the machine and treatment levels, where waste should be primarily addressed.

In future studies, the EMMSA methodology can be extended to more than ten covariates, include binary variables, and use variables with very different probability distributions. These extensions may require more criteria or priors to improve the model selection process.

The EMMSA methodology could be adapted for applications in other industries involving production processes in which waste is a fundamental part of the cost. It could also be extended to other fields of knowledge.

Author Contributions: Choice of article topic: R.A.-L. and J.d.D.L.d.C.; Development and review of model comparison methodology: R.A.-L. and J.d.D.L.d.C.; Data collection: R.A.-L.; Editing, cleaning, and adjustment of data: R.A.-L.; Development and review of algorithm application software: R.A.-L.; Analysis of results: R.A.-L., J.d.D.L.d.C. and M.Á.M.-A.; Drafting and review of article: R.A.-L., J.d.D.L.d.C. and M.Á.M.-A.; Final editing and review of article: R.A.-L. and J.d.D.L.d.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request.

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

Appendix A

Appendix A.1. Causality Methods in Time Series

Table A1. Multivariable Granger causality methods for time series.

Method	Relations	Data	Reference
Auto Regressive Model	Linear	Stationary	[16]
Gaussian Process Regression	Non-linear	Non-stationary data	[16]
Spectral Granger Causality	Non-linear	Oscillating variables: Fourier transformation	[61]
Grouping Multivariate Granger Causality	Non-linear	Oscillating variables: Multivariate non-linear chirp mode decomposition	[62]

Table A2. Granger causality methods for time series.

Method	Description	Reference
Direct Transfer Entropy (DTE)	Detects spurious causalities and direct (indirect) paths between univariate time series. Requires stationary data.	[19]
(Direct) Transfer Zero Entropy (T0E)	Like DTE but does not assume well-defined PDF. Detects direct or indirect causalities for multivariate data. Does not require stationary data or a large sample.	[63]
Trend Transfer Entropy (TTE)	Causal analysis based on the time-series trends reducing computational burden.	[64]
Symbolic dynamic based normalized Transfer Entropy (SDNTE)	Fast and efficient to root cause diagnostic in real time using xD-Markov machine. Accepts only stationary variables.	[65]
Normalized (Direct) Transfer Entropy	It assumes a random delay in the occurrence of the alarms and mutual independence between them.	[66]

Appendix A.2. Distances and R Libraries

Table A3. Proximity metrics.

Distance	Description	Reference
Canberra	$d(p, q) = \sum_{i=1}^n \frac{ p_i - q_i }{ p_i + q_i }$	[67]
Euclidean and Scaled Euclidean	$d(p, q) = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}$	
GAM	Generalized additive model distance is a form of parametric logistic regression. It belongs to the family of generalized linear models. It assumes that outcome variables depends linearly on unknown smooth functions of its predictor variables.	[68]
GLM	The generalized linear model distance is a generalization of ordinary linear regression using a link function between predictors and outcome and for each variable a variance function of its predictors. It assumes that each predictor is generated by a probability distribution family (normal, binomial, exponential, poisson, etc.).	[69]
Mahalanobis and Robust Mahalanobis	$d(p, q, \theta) = \sqrt{(p - q)^T S^{-1} (p - q)}$ Where, p and q are vectors of size n , θ is the probability distribution over R^n , and S is the positive definite covariance matrix.	[70]
Manhattan (Taxicab)	$d(p, q) = p_i - q_i $ Hermann Minkowski (1864–1909)	
Maximum	$d(p, q) = \max_i p_i - q_i $ Pafnuty Chebyshev (1821–1894)	
Minkowski	$d(p, q) = \sqrt[s]{\sum_{i=1}^n (p_i - q_i)^s}$ $s = 1$, Manhattan distance; $s = 2$, Euclidean distance; $s = \infty$, Chebyshev distance. Hermann Minkowski (1864–1909)	

Table A4. R Packages used.

Package	Description	Reference
matchIt	It implements a wide variety of matching algorithms such as nearest Mahalanobis, optimal matching, fullmatch, genetic, and coarsened exact matching.	[44]
optmatch	Finds the optimal match and support variable number of controls and full matching. It uses auction algorithm implemented with RELAX IV Fortran code [42] to solve the minimum cost flow problem. It has academic license.	[41]
sensitivitymult	Sensitivity analysis for normal distributions using M-Test.	[51]
car	Sensitivity analysis for no normal distributions using Levene test.	[71]
cobalt	Generates balance tables and plots for covariates of matched groups	[72]
rbounds	Implements sensitivity analysis methods documented in [29] and calculates point estimates from Hodges-Lehmann and Wilcoxon tests.	[73]
tableone	It summarizes the variables (continuous or categorical) of the matching results, thus facilitating their reading. In addition, it performs statistical tests and calculates the standardized mean difference.	[74]

Table A4. *Cont.*

Package	Description	Reference
lightgbm	Optimized machine learning algorithm using gradient boosting decision tree (GBDT) implemented with gradient-based one-side sampling (GOSS) and exclusive feature bundling (EFB)	[60]
randomForest	A classification and regression algorithm that combines randomized decision trees and aggregates their predictions by averages.	[75]
meta	Implements methods for meta-analysis of random-effect models and common-effect models.	[76]

Table A5. DerSimonian–Laird nomenclature.

Variable	Nomenclature of Random Effect Model for DerSimonian–Laird
k	Number of comparative studies
i	$1, \dots, k$
n_{Ti}, n_{Ci}	Number of records in treatment and control groups
r_{Ti}, r_{Ci}	Proportion of records with some event in treatment and control groups
Θ_i	True treatment effect in the i -th study
e_i	Sampling error in the i -th study
s_i	Standard deviation of the i -th study
Observed treatment effect	$Y_i = \text{True treatment effect} + \text{sampling error}$ $Y_i = \Theta_i + e_i$
Variance of e_i	σ_i^2 The sampling variance capturing intra-study variance and the sampling size. It is usually unknown, but it is estimated from study data (s_i^2).
Θ_i	$\Theta_i = \mu + \delta_i$
μ	Overall treatment effect of study population
δ_i	$\delta_i = \Theta_i - \mu$ It is the deviation of the study from overall treatment effect.
Variance of δ_i	τ^2 It is the between-study variance and constitutes both the degree of variation in the true treatment effect across studies and the degree of variation in biased assessments of the treatment effect of individual studies. It is estimated from study data (t^2).
$\tau^2 = 0$	Homogeneity among the true treatment effects. Therefore, $\Theta_i = \mu$ $\tau^2 > 0$ reflects treatment effects heterogeneity.
y_i	The observed treatment effect for i -th study. It is a random variable. y_1, \dots, y_k are values of random variable from a distribution with overall means μ ; true treatment effect Θ_i ; and variances $\tau^2 + \sigma_1^2, \dots, \tau^2 + \sigma_k^2$; $\tau^2 > 0$; and $\sigma_1^2 > 0$.
Variance of each observed treatment effect	$\tau^2 + \sigma_i^2$ Inter-study variance = τ^2 . Intra-study variance = σ_i^2 ; it is approximately equal to s_i^2 .

Table A6. DerSimonian–Laird Formulas.

Formulas for the Estimation of the Overall Population Treatment Effect μ	Equation
Estimate of the initial weighting w_{i0}	$w_{i0} = \frac{1}{s_i^2}$ (A1)
Estimate of initial $y_w(0)$	$y_w(0) = \frac{\sum_{i=1}^k w_{i0} y_i}{\sum_{i=1}^k w_{i0}}$ (A2)
Estimate Q statistics	$Q = \left[\sum_{i=1}^k w_{i0} (y_i - y_w(0))^2 \right]$ (A3)
Estimate t^2	$t^2 = \frac{Q - (k-1)}{\left[\sum_{i=1}^k w_{i0} - \frac{\sum_{i=1}^k w_{i0}^2}{\sum_{i=1}^k w_{i0}} \right]}$ (A4)
Estimate second step weighting, w_i	$w_i = \frac{1}{t^2 + s_i^2}$ (A5)
Estimate m_w	$m_w = \frac{\sum_{i=1}^k w_i y_i}{\sum_{i=1}^k w_i}$ (A6)
Estimate τ^2 , weighted method of moments	$\tau^2 \approx \frac{\left[\sum_{i=1}^k w_i (y_i - m_w)^2 \right] - \left[\sum_{i=1}^k w_i s_i^2 - \frac{\sum_{i=1}^k w_i^2 s_i^2}{\sum_{i=1}^k w_i} \right]}{\left[\sum_{i=1}^k w_i - \frac{\sum_{i=1}^k w_i^2}{\sum_{i=1}^k w_i} \right]}$ (A7)
Estimate final weighting W_i	$W_i = \frac{1}{\tau^2 + s_i^2}$ (A8)
Estimate μ	$M_W = \frac{\sum_{i=1}^k W_i y_i}{\sum_{i=1}^k W_i}$ (A9)
Estimate Standard error μ_W	Standar error of $\mu_W = SE(M_W) = \frac{1}{\left(\sum_{i=1}^k W_i \right)^{\frac{1}{2}}}$ (A10)
Confidence Interval μ	$CI = M_W \pm 1.96 SE(M_W)$ (A11)

Appendix B

Treatment Intervals, Density Functions, and Selected Models

Table A7. Downtime intervals of machines for each treatment (MSUMALL).

Trat	Min	Max	M1	M2	M3	M4	M5	M6	M7	Total
0	-	-	3521	978	3008	7318	1831	3087	4039	23,782
1	59	291	166	62	54	990	50	331	173	1826
2	292	441	150	68	80	897	138	317	210	1860
3	442	580	142	76	109	779	129	436	209	1880
4	581	722	152	77	85	715	127	493	233	1882
5	723	908	187	91	96	693	133	355	308	1863
6	909	1160	219	112	103	673	171	349	243	1870
7	1161	1497	222	106	107	661	165	370	240	1871
8	1498	2010	252	136	119	605	178	389	202	1881
9	2011	3005	274	154	124	531	200	371	214	1868
10	3006	18,113	334	237	162	406	231	407	195	1972
	Total		5619	2097	4047	14,268	3353	6905	6266	42,555
	% Control units		62.7%	46.6%	74.3%	51.3%	54.6%	44.7%	64.5%	55.9%
	% Treatment units		37.3%	53.4%	25.7%	48.7%	45.4%	55.3%	35.5%	44.1%

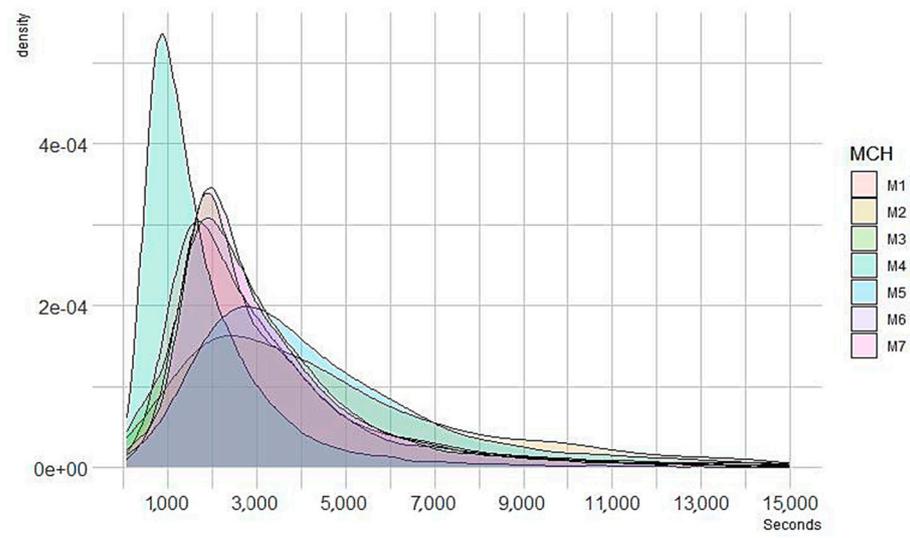


Figure A1. MSUMALL density function for each machine.

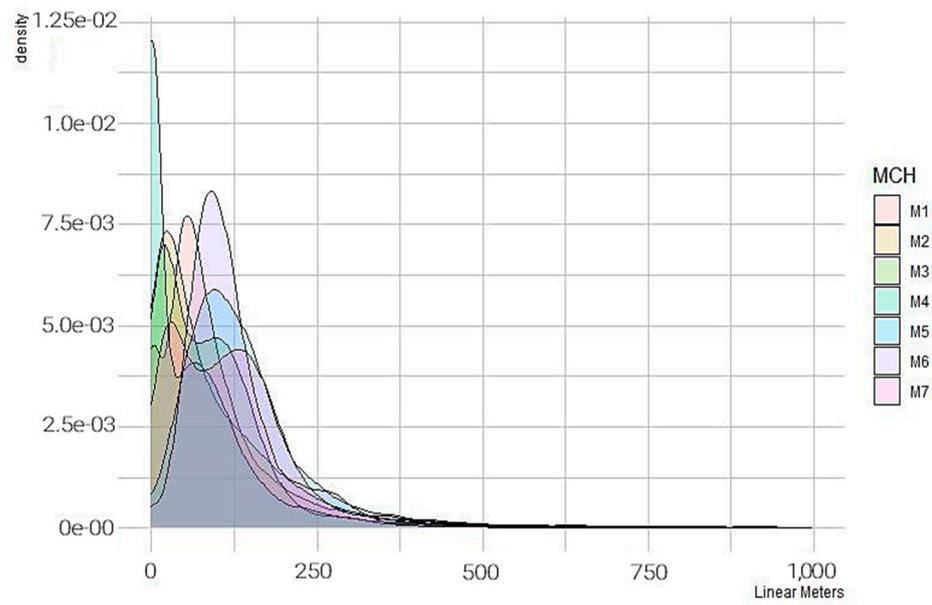


Figure A2. WASTELM density function for each machine.

Table A8. Selected models: balance information and causal effect size.

MCH	Treat	Method	Distance	C1	C2	CU	TU	NSMD	NRV	CMU	TMU	RV	Wtpval	EffMD	EffRF	CIhigh	CIlow	ZScore
M1	1	Optmatch	mahalanobis	1	1	3521	166	0	0	166	166	0.84	0.01	19.80	19.42	21.73	17.88	0.27
M1	2	FullMatch	scaled_euclidean	1	1	3521	150	0	0	3521	150	0.82	0.02	16.35	16.19	17.09	15.62	0.20
M1	3	Genetic	bart	1	1	3521	142	0	0	142	142	0.64	0.11	16.07	15.60	17.61	14.54	0.19
M1	4	Genetic	gam	1	1	3521	152	0	0	152	152	0.91	0.25	11.53	11.27			0.13
M1	5	Genetic	gam	1	1	3521	187	0	0	187	187	0.92	0.35	8.00	7.26	10.00	5.99	0.10
M1	6	Genetic	gam	1	1	3521	219	0	0	219	219	1.02	0.12	13.05	13.78	14.06	12.03	0.15
M1	7	Genetic	gam	1	1	3521	222	0	0	222	222	1.42	0.24	10.28	9.87			0.11
M1	8	Genetic	gbm	2	1	3521	252	1	0	252	252	0.76	0.14	11.97	12.87	12.89	11.05	0.13
M1	9	FullMatch	scaled_euclidean	2	1	3521	274	1	0	3521	274	0.61	0.00	24.38	24.52	26.21	22.55	0.30
M1	10	FullMatch	mahalanobis	2	1	3521	334	1	0	3521	334	0.58	0.00	34.25	34.11	37.57	30.94	0.41
M2	1	Genetic	robust_mahalanobis	1	1	978	62	0	0	62	62	1.82	0.37	7.87	8.38	14.34	1.41	0.16
M2	2	Optmatch	mahalanobis	1	1	978	68	0	0	68	68	0.48	0.14	15.13	15.56	17.67	12.59	0.26
M2	3	FullMatch	robust_mahalanobis	1	1	978	76	0	0	978	76	1.86	0.08	9.78	10.39	11.40	8.15	0.18
M2	4	Genetic	gbm	1	1	978	77	0	0	77	77	1.84	0.12	15.67	16.55	19.42	11.91	0.25
M2	5	Genetic	gbm	1	1	978	91	0	0	91	91	1.42	0.29	14.45	14.16	27.97	0.93	0.16
M2	6	Genetic	gbm	1	1	978	112	0	0	112	112	0.69	0.38	11.63	12.07	11.63	11.63	0.12
M2	7	Nearest	scaled_euclidean	1	1	978	106	0	0	106	106	0.70	0.64	5.34	5.82	9.62	1.06	0.07
M2	8	Genetic	scaled_euclidean	1	1	978	136	0	0	136	136	0.54	0.02	19.70	20.57	21.22	18.19	0.29
M2	9	Optmatch	scaled_euclidean	1	1	978	154	0	0	154	154	1.14	0.42	7.54	6.37	8.58	6.49	0.09
M2	10	Genetic	scaled_euclidean	1	1	978	237	0	0	237	237	0.77	0.41	5.84	5.68	11.22	0.45	0.08
M3	1	Genetic	gam	1	1	3007	54	0	0	54	54	1.31	0.34	9.39	9.35	15.91	2.87	0.18
M3	2	Genetic	mahalanobis	1	1	3007	80	0	0	80	80	0.41	0.01	20.99	22.14	23.46	18.52	0.44
M3	3	Genetic	scaled_euclidean	1	1	3007	109	0	0	109	109	1.39	0.53	5.80	5.07	9.67	1.93	0.09
M3	4	Genetic	bart	1	1	3007	85	0	0	85	85	0.45	0.00	20.55	20.69	23.42	17.67	0.45
M3	5	Genetic	gam	1	1	3007	95	0	0	95	95	0.73	0.22	7.47	7.32	8.53	6.40	0.18
M3	6	Genetic	gam	1	1	3007	103	0	0	103	103	1.88	0.23	7.71	7.95	9.14	6.28	0.17
M3	7	Genetic	gam	1	1	3007	106	0	0	106	106	0.57	0.14	13.71	13.10	15.38	12.05	0.21
M3	8	Genetic	gam	1	1	3007	119	0	0	119	119	0.61	0.27	7.05	6.27	8.28	5.82	0.14

Table A8. Cont.

MCH	Treat	Method	Distance	C1	C2	CU	TU	NSMD	NRV	CMU	TMU	RV	Wtpval	EffMD	EffRF	CIhigh	CIlow	ZScore
M3	9	Nearest	scaled_euclidean	1	1	3007	124	0	0	124	124	0.66	0.14	12.65	11.93	13.46	11.83	0.19
M3	10	Genetic	glm	1	1	3007	162	0	0	162	162	0.71	0.14	11.84	12.34	13.04	10.63	0.16
M4	1	Genetic	euclidean	1	1	7318	990	0	0	990	990	0.65	0.10	4.90	5.09	5.29	4.52	0.07
M4	2	Nearest	scaled_euclidean	1	1	7318	897	0	0	897	897	0.66	0.00	11.65	11.75	12.42	10.89	0.16
M4	3	Genetic	mahalanobis	1	1	7318	779	0	0	779	779	0.78	0.00	13.68	13.78	13.98	13.37	0.17
M4	4	Genetic	gbm	1	1	7318	715	0	0	715	715	0.85	0.01	8.53	8.67	9.05	8.00	0.15
M4	5	Optmatch	scaled_euclidean	1	1	7318	693	0	0	693	693	0.62	0.00	13.89	14.03	14.46	13.33	0.19
M4	6	Genetic	bart	1	1	7318	673	0	0	673	673	0.61	0.00	17.83	18.29	18.55	17.11	0.23
M4	7	Nearest	robust_mahalanobis	1	1	7318	661	0	0	661	661	0.65	0.00	20.19	20.38	21.17	19.20	0.25
M4	8	Genetic	scaled_euclidean	1	1	7318	605	0	0	605	605	0.49	0.00	22.58	22.69	23.28	21.89	0.26
M4	9	Genetic	euclidean	1	1	7318	531	0	0	531	531	0.57	0.00	29.28	29.01	30.12	28.44	0.36
M4	10	Genetic	euclidean	1	1	7318	406	0	0	406	406	0.49	0.00	48.60	49.27	50.23	46.98	0.52
M5	1	Genetic	robust_mahalanobis	1	1	1831	50	0	0	50	50	1.64	0.07	24.81	25.65	28.36	21.26	0.37
M5	2	Nearest	mahalanobis	1	1	1831	138	0	0	138	138	0.70	0.03	15.64	15.60	17.55	13.74	0.27
M5	3	Genetic	mahalanobis	1	1	1831	129	0	0	129	129	0.76	0.07	10.86	10.44	12.14	9.58	0.23
M5	4	Genetic	glm	1	1	1831	127	0	0	127	127	0.39	0.02	19.82	20.09	21.41	18.23	0.30
M5	5	Genetic	robust_mahalanobis	1	1	1831	133	0	0	133	133	0.79	0.07	15.62	15.96	17.33	13.91	0.23
M5	6	Genetic	mahalanobis	1	1	1831	171	0	0	171	171	0.74	0.07	9.99	9.84	14.86	5.12	0.20
M5	7	Optmatch	glm	1	1	1831	165	0	0	165	165	0.63	0.01	26.17	26.32	26.77	25.56	0.31
M5	8	Genetic	gam	1	1	1831	178	0	0	178	178	0.70	0.01	15.56	15.69	17.61	13.50	0.26
M5	9	FullMatch	euclidean	1	1	1831	200	0	0	1831	200	0.82	0.00	25.22	24.96	27.38	23.06	0.37
M5	10	FullMatch	scaled_euclidean	1	1	1831	231	0	0	1831	231	0.48	0.00	36.79	37.42	45.70	27.88	0.53
M6	1	FullMatch	scaled_euclidean	1	1	3087	331	0	0	3087	331	0.84	0.00	8.06	8.11	9.41	6.72	0.24
M6	2	FullMatch	scaled_euclidean	1	1	3087	317	0	0	3087	317	0.84	0.05	4.36	4.24	7.23	1.48	0.12
M6	3	FullMatch	scaled_euclidean	1	1	3087	436	0	0	3087	436	0.61	0.00	12.56	12.77	13.43	11.69	0.25
M6	4	FullMatch	mahalanobis	1	1	3087	493	0	0	3087	493	0.70	0.00	7.71	7.75	12.80	2.62	0.21
M6	5	FullMatch	scaled_euclidean	1	1	3087	355	0	0	3087	355	0.77	0.00	7.21	7.23	9.52	4.91	0.24

Table A8. Cont.

MCH	Treat	Method	Distance	C1	C2	CU	TU	NSMD	NRV	CMU	TMU	RV	Wtpval	EffMD	EffRF	CIhigh	CIlow	ZScore
M6	6	FullMatch	euclidean	1	1	3087	349	0	0	3087	349	0.58	0.00	14.87	14.93	17.80	11.94	0.34
M6	7	FullMatch	scaled_euclidean	1	1	3087	370	0	0	3087	370	0.70	0.00	16.28	16.48	23.38	9.19	0.42
M6	8	FullMatch	mahalanobis	1	1	3087	389	0	0	3087	389	0.51	0.00	19.75	19.96	25.79	13.72	0.44
M6	9	Genetic	bart	1	1	3087	371	0	0	371	371	0.74	0.01	14.97	14.87	16.07	13.88	0.20
M6	10	FullMatch	euclidean	1	1	3087	407	0	0	3087	407	0.46	0.00	34.69	35.61	34.69	34.69	0.46
M7	1	Genetic	robust_mahalanobis	1	1	4039	173	0	0	173	173	0.79	0.27	6.59	6.48	9.09	4.10	0.12
M7	2	Nearest	mahalanobis	1	1	4039	210	0	0	210	210	0.98	0.36	6.87	7.05	8.15	5.59	0.09
M7	3	Nearest	euclidean	1	1	4039	209	0	0	209	209	0.82	0.06	8.51	8.58	11.46	5.57	0.18
M7	4	Optmatch	mahalanobis	1	1	4039	233	0	0	233	233	0.72	0.05	12.14	12.74	14.06	10.23	0.18
M7	5	Nearest	mahalanobis	1	1	4039	308	0	0	308	308	0.65	0.02	14.74	14.47	15.85	13.64	0.19
M7	6	Optmatch	glm	1	1	4039	243	0	0	243	243	0.68	0.23	7.11	7.71	7.94	6.29	0.11
M7	7	Genetic	scaled_euclidean	1	1	4039	240	0	0	240	240	1.16	0.14	7.72	7.81	8.78	6.66	0.14
M7	8	Optmatch	robust_mahalanobis	1	1	4039	202	0	0	202	202	0.59	0.26	6.49	6.56	9.95	3.02	0.11
M7	9	Genetic	gam	1	1	4039	214	0	0	214	214	1.18	0.05	13.04	13.03	14.64	11.43	0.19
M7	10	Genetic	bart	1	1	4039	195	0	0	195	195	0.62	0.04	14.99	15.77	16.68	13.30	0.21

C1:Priority of condition 1; C2: Priority of condition 2; CU: Control units; TU: Treatment units; NSMD: Non-balanced variables by SMS; NRV: Non-balanced variables by RV; CMU: Control-matched units; TMU: Treatment-matched units; RV: Variance ratio; Wtpval: *p*-value of Welch test; EffMD: Mean effect size of treatment of homologous models; EffRF: Effect size of RF model of treatment selected; CIhigh: Confidence interval high value for machine-treatment of homologous models; CIlow: Confidence interval low value for machine-treatment of homologous models; HLUCE: Hodges-Lehmann unconfounded estimate; ZScore: Z-score of selected model.

References

1. Alfieri, A. Workload Simulation and Optimisation in Multi-Criteria Hybrid Flowshop Scheduling: A Case Study. *Int. J. Prod. Res.* **2009**, *47*, 5129–5145. [CrossRef]
2. Li, X.; Wang, J.; Huang, C.; Gao, D.; Lu, G.; Lu, L.; Wang, Z. Mathematical Models for Predicting the Quasi-Static Stress Characteristics of Corrugated Paperboard with Sinusoidal Core along the Longitudinal Compression. *Int. J. Mech. Sci.* **2018**, *149*, 136–149. [CrossRef]
3. Research and Markets. In *Paperboard Packaging Market—Growth, Trends, COVID-19 Impact, and Forecasts (2023–2028)*; Mordor Intelligence: Dublin, Ireland, 2023; Available online: <https://www.researchandmarkets.com/reports/4536057/paperboard-packaging-market-growth-trends> (accessed on 22 September 2023).
4. Farghaly, A.; Roux, S.L.; Peu, P.; Dabert, P.; Tawfik, A. Effect of Starvation Period on Microbial Community Producing Hydrogen from Paperboard Mill Wastewater Using Anaerobic Baffled Reactor. *Environ. Technol.* **2019**, *40*, 2389–2399. [CrossRef] [PubMed]
5. Kirwan, M.J. *Handbook of Paper and Paperboard Packaging Technology*; Wiley-Blackwell: Chichester, UK, 2013; ISBN 9780470670668.
6. Kot, S.; Grondys, K. Total Productive Maintenance in Enterprise Operations Support Processes. *Appl. Mech. Mater.* **2013**, *309*, 324–331. [CrossRef]
7. Seyed Alinezhad, H.; Roohi, M.H.; Chen, T. A Review of Alarm Root Cause Analysis in Process Industries: Common Methods, Recent Research Status and Challenges. *Chem. Eng. Res. Des.* **2022**, *188*, 846–860. [CrossRef]
8. Vuković, M.; Thalmann, S. Causal Discovery in Manufacturing: A Structured Literature Review. *J. Manuf. Mater. Process.* **2022**, *6*, 10. [CrossRef]
9. Kaddour, J.; Lynch, A.; Liu, Q.; Kusner, M.J.; Silva, R. Causal Machine Learning: A Survey and Open Problems. *arXiv* **2022**, arXiv:2206.15475.
10. Hagedorn, C.; Huegle, J.; Schlosser, R. Understanding Unforeseen Production Downtimes in Manufacturing Processes Using Log Data-Driven Causal Reasoning. *J. Intell. Manuf.* **2022**, *33*, 2027–2043. [CrossRef]
11. Choudhury, M.A.A.S. Plantwide Oscillations Diagnosis-Current State and Future Directions. *Asia-Pac. J. Chem. Eng.* **2011**, *6*, 484–496. [CrossRef]
12. Stuart, E.A. Matching Methods for Causal Inference: A Review and a Look Forward. *Stat. Sci.* **2010**, *25*, 1–21. [CrossRef] [PubMed]
13. Granger, C.W.J. Investigating Causal Relations by Econometric Models and Cross-Spectral Methods. *Econometrica* **1969**, *37*, 424–438. [CrossRef]
14. Barnett, L.; Barrett, A.B.; Seth, A.K. Granger Causality and Transfer Entropy Are Equivalent for Gaussian Variables. *Phys. Rev. Lett.* **2009**, *103*, 238701. [CrossRef] [PubMed]
15. Shojaie, A.; Fox, E.B. Annual Review of Statistics and Its Application Granger Causality: A Review and Recent Advances. *Annu. Rev. Stat. Appl.* **2022**, *9*, 289–319. [CrossRef] [PubMed]
16. Chen, H.S.; Yan, Z.; Yao, Y.; Huang, T.B.; Wong, Y.S. Systematic Procedure for Granger-Causality-Based Root Cause Diagnosis of Chemical Process Faults. *Ind. Eng. Chem. Res.* **2018**, *57*, 9500–9512. [CrossRef]
17. Schreiber, T. Measuring Information Transfer. *Phys. Rev. Lett.* **2000**, *85*, 461. [CrossRef]
18. Bauer, M.; Cox, J.W.; Caveness, M.H.; Downs, J.J.; Thornhill, N.F. Finding the Direction of Disturbance Propagation in a Chemical Process Using Transfer Entropy. *IEEE Trans. Control Syst. Technol.* **2007**, *15*, 12–21. [CrossRef]
19. Duan, P.; Yang, F.; Chen, T.; Shah, S.L. Direct Causality Detection via the Transfer Entropy Approach. *IEEE Trans. Control Syst. Technol.* **2013**, *21*, 2052–2066. [CrossRef]
20. Pearl, J.; Russell, S. Bayesian Networks. *Handb. Brain Theory Neural Netw.* **2003**, *2*, 157–160. [CrossRef]
21. Glymour, C.; Zhang, K.; Spirtes, P. Review of Causal Discovery Methods Based on Graphical Models. *Front. Genet.* **2019**, *10*, 524. [CrossRef]
22. Rubin, D.B. Formal Modes of Statistical Inference for Causal Effects. *J. Stat. Plan. Inference* **1990**, *25*, 279–292. [CrossRef]
23. Rubin, D.B. Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *J. Educ. Psychol. Am. Psychol. Assoc.* **1974**, *66*, 688. [CrossRef]
24. Morgan, S.L.; Winship, C. Counterfactuals and Causal Inference. In *Methods and Principles for Social Research*, 2nd ed.; Cambridge University Press: New York, NY, USA, 2015; ISBN 9781107065079.
25. Imbens, G.W.; Rubin, D.B. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*; Cambridge University Press: New York, NY, USA, 2015; ISBN 0521885884/9780521885881.
26. Holland, P.W. Statistics and Causal Inference. *J. Am. Stat. Assoc.* **1986**, *81*, 945–960. [CrossRef]
27. Hernán, M.A.; Robins, J.M. Estimating Causal Effects from Epidemiological Data. *J. Epidemiol. Community Health* **2006**, *60*, 578–586. [CrossRef]
28. Imai, K.; King, G.; Stuart, E.A. Misunderstandings between Experimentalists and Observationalists about Causal Inference. *J. R. Statist. Soc. A* **2008**, *171*, 481–502. [CrossRef]
29. Rosenbaum, P.R. *Design of Observational Studies*; Springer: New York, NY, 2010; ISBN 0387947256.
30. King, G.; Zeng, L. The Dangers of Extreme Counterfactuals. *Political Anal.* **2006**, *14*, 131–159. [CrossRef]
31. Ho, D.E.; Imai, K.; King, G.; Stuart, E.A. Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference. *Political Anal.* **2007**, *15*, 199–236. [CrossRef]
32. Morgan, S.L. *Handbook of Causal Analysis for Social Research*; Springer: Dordrecht, The Netherlands, 2013; ISBN 9789400760943/99789400760943.

33. Rosenbaum, P.R. Modern Algorithms for Matching in Observational Studies. *Annu. Rev. Stat. Appl.* **2020**, *7*, 143–176. [[CrossRef](#)]
34. Smith, H.L. Matching with Multiple Controls to Estimate Treatment Effect in Observational Studies. *Sociol. Methodol.* **1997**, 325–353. [[CrossRef](#)]
35. Dehejia, R.H.; Wahba, S. Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs. *J. Am. Stat. Assoc.* **1999**, *94*, 1053–1062. [[CrossRef](#)]
36. Rosenbaum, P.R.; Rubin, D.B. The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika* **1983**, *70*, 41–55. [[CrossRef](#)]
37. Cochran, W.G.; Rubin, D.B. Controlling Bias in Observational Studies: A Review. In *Matched Sampling for Causal Effects*; Cambridge University Press: New York, NY, USA, 2006; pp. 30–58. ISBN 9780511810725.
38. Muja, M.; Lowe, D.G. Fast Approximate Nearest Neighbors with Automatic Algorithm Configuration. *VISAPP* **2009**, *1*, 331–340.
39. Gu, X.S.; Rosenbaum, P.R. Comparison of Multivariate Matching Methods: Structures, Distances, and Algorithms. *J. Comput. Graph. Stat.* **1993**, *2*, 405–420. [[CrossRef](#)]
40. Rosenbaum, P.R.; Rubin, D.B. Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score. *Am. Stat.* **1985**, *39*, 33–38. [[CrossRef](#)]
41. Hansen, B.B.; Olsen Klopfer, S. Optimal Full Matching and Related Designs via Network Flows. *J. Comput. Graph. Stat.* **2006**, *15*, 609–627. [[CrossRef](#)]
42. Bertsekas, D.P.; Tseng, P. RELAXATION METHODS FOR MINIMUM COST ORDINARY AND GENERALIZED NETWORK FLOW PROBLEMS. *Oper. Res.* **1988**, *36*, 93–114. [[CrossRef](#)]
43. Hansen, B.B. Full Matching in an Observational Study of Coaching for the SAT. *J. Am. Stat. Assoc.* **2004**, *99*, 609–618. [[CrossRef](#)]
44. Stuart, E.A.; King, G.; Imai, K.; Ho, D. MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. *J. Stat. Softw.* **2011**, *42*, 1–28. [[CrossRef](#)]
45. Diamond, A.; Sekhon, J.S. Genetic Matching for Estimating Causal Effects: A General Multivariate Matching Method for Achieving Balance in Observational Studies. *Rev. Econ. Stat.* **2013**, *95*, 932–945. [[CrossRef](#)]
46. Stuart, E.A.; Lee, B.K.; Leacy, F.P. Prognostic Score-Based Balance Measures Can Be a Useful Diagnostic for Propensity Score Methods in Comparative Effectiveness Research. *J. Clin. Epidemiol.* **2013**, *66*, S84–S90.e1. [[CrossRef](#)]
47. Flury, B.K.; Riedwyl, H. Standard Distance in Univariate and Multivariate Analysis. *Am. Stat.* **1986**, *40*, 249–251. [[CrossRef](#)]
48. Austin, P.C. Balance Diagnostics for Comparing the Distribution of Baseline Covariates between Treatment Groups in Propensity-Score Matched Samples. *Stat. Med.* **2009**, *28*, 3083–3107. [[CrossRef](#)] [[PubMed](#)]
49. Hansen, B.B. The Prognostic Analogue of the Propensity Score. *Biometrika* **2008**, *95*, 481–488. [[CrossRef](#)]
50. Zhang, Z.; Kim, H.J.; Lonjon, G.; Zhu, Y. Balance Diagnostics after Propensity Score Matching. *Ann. Transl. Med.* **2019**, *7*, 16. [[CrossRef](#)] [[PubMed](#)]
51. Rosenbaum, P.R. Sensitivity Analyses Informed by Tests for Bias in Observational Studies. *Biometrics* **2021**, *79*, 475–487. [[CrossRef](#)]
52. Rosenbaum, P.R. Hodges–Lehmann Point Estimates of Treatment Effect in Observational Studies. *J. Am. Stat. Assoc.* **1993**, *88*, 1250–1253. [[CrossRef](#)]
53. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed.; Lawrence Erlbaum Associates: New York, NY, USA, 1988; ISBN 0805802835.
54. Kallus, N. Generalized Optimal Matching Methods for Causal Inference. *J. Mach. Learn. Res.* **2020**, *21*, 1–54. Available online: <https://jmlr.org/papers/v21/19-120.html> (accessed on 22 September 2023).
55. Zhao, S.; van Dyk, D.A.; Imai, K. Propensity Score-Based Methods for Causal Inference in Observational Studies with Non-Binary Treatments. *Stat. Methods Med. Res.* **2020**, *29*, 709–727. [[CrossRef](#)]
56. DerSimonian, R.; Kacker, R. Random-Effects Model for Meta-Analysis of Clinical Trials: An Update. *Contemp. Clin. Trials* **2007**, *28*, 105–114. [[CrossRef](#)]
57. Borenstein, M.; Hedges, L.V.; Higgins, J.P.T.; Rothstein, H.R. A Basic Introduction to Fixed-Effect and Random-Effects Models for Meta-Analysis. *Res. Synth. Methods* **2010**, *1*, 97–111. [[CrossRef](#)]
58. Maathuis, M.H.; Nandy, P. A Review of Some Recent Advances in Causal Inference. *arXiv* **2015**. [[CrossRef](#)]
59. Pearl, J. *Causality: Models, Reasoning and Inference*, 2nd ed.; Cambridge University Press: New York, NY, USA, 2009; ISBN 052189560X/9780521895606.
60. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Ye, Q.; Liu, T.-Y. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. In Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017.
61. Yuan, T.; Qin, S.J. Root Cause Diagnosis of Plant-Wide Oscillations Using Granger Causality. *J. Process. Control.* **2014**, *24*, 450–459. [[CrossRef](#)]
62. Chen, Q.; Lang, X.; Lu, S.; ur Rehman, N.; Xie, L.; Su, H. Detection and Root Cause Analysis of Multiple Plant-Wide Oscillations Using Multivariate Nonlinear Chirp Mode Decomposition and Multivariate Granger Causality. *Comput. Chem. Eng.* **2021**, *147*, 107231. [[CrossRef](#)]
63. Duan, P.; Yang, F.; Shah, S.L.; Chen, T. Transfer Zero-Entropy and Its Application for Capturing Cause and Effect Relationship Between Variables. *IEEE Trans. Control. Syst. Technol.* **2015**, *23*, 855–867. [[CrossRef](#)]
64. Guo, C.; Yang, F.; Yu, W. A Causality Capturing Method for Diagnosis based on Transfer Entropy by Analyzing Trends of Time Series. *IFAC-PapersOnLine* **2015**, *48*, 778–783. [[CrossRef](#)]

65. Rashidi, B.; Singh, D.S.; Zhao, Q. Data-Driven Root-Cause Fault Diagnosis for Multivariate Non-Linear Processes. *Control. Eng. Pract.* **2018**, *70*, 134–147. [[CrossRef](#)]
66. Hu, W.; Wang, J.; Chen, T.; Shah, S.L. Cause-Effect Analysis of Industrial Alarm Variables Using Transfer Entropies. *Control. Eng. Pract.* **2017**, *64*, 205–214. [[CrossRef](#)]
67. Lance, G.N.; Williams, W.T. Mixed-Data Classificatory Programs I—Agglomerative Systems. *Aust. Comput. J.* **1967**, *1*, 15–20.
68. Hastie, T.; Tibshirani, R. Generalized Additive Models; Some Applications. *J. Am. Stat. Assoc.* **1987**, *82*, 371–386. [[CrossRef](#)]
69. Nelder, J.A.; Wedderburn, R.W.M. Generalized Linear Models. *J. R. Stat. Soc. Ser. A* **1972**, *135*, 370–384. [[CrossRef](#)]
70. Mahalanobis, P.C. On the Generalized Distance in Statistics. *Sankhyā Indian J. Stat. Ser. A* **2018**, *80-A*, S1–S7. Available online: <https://www.jstor.org/stable/e48513082> (accessed on 22 September 2023).
71. Fox, J.; Weisberg, S. *An R Companion to Applied Regression*, 3rd ed.; SAGE Publications, Inc.: Thousand Oaks, CA, USA, 2019; ISBN 9781544336473.
72. Greifer, N. Covariate Balance Tables and Plots: A Guide to the Cobalt Package. 2020. Available online: <https://cran.r-project.org/web/packages/cobalt/index.html> (accessed on 22 September 2023).
73. Keele, L.J. Perform Rosenbaum Bounds Sensitivity Tests for Matched and Unmatched Data. R Package ‘rbounds’. 2022. Available online: <https://cran.r-project.org/web/packages/rbounds/rbounds.pdf> (accessed on 22 September 2023).
74. Yoshida, K.; Bohn, J.; Yoshida, M. Package ‘tableone’. R Foundation for Statistical Computing. 2020. Available online: <https://github.com/kaz-yos/tableone> (accessed on 22 September 2023).
75. Liaw, A.; Wiener, M. Classification and Regression by randomForest. *R News* **2002**, *2*, 18–22. Available online: <https://journal.r-project.org/articles/RN-2002-022/RN-2002-022.pdf> (accessed on 22 September 2023).
76. Cohen, Y.; Cohen, J.Y. *Statistics and Data with R. An Applied Approach through Examples*; John Wiley & Sons: Chichester, UK, 2008; ISBN 0470721898.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.