JAMCM Journal of Applied Mathematics and Computational Mechanics 2025, 24(3), 76-97
www.amcm.pcz.pl p-ISSN 2299-9965
DOI: 10.17512/jamcm.2025.3.05 e-ISSN 2353-0588

PAIRWISE-COMPARISON-BASED PRIORITIZATION AND THE PARADIGM OF RECIPROCITY

Andrzej Z. Grzybowski

Department of Mathematics, Czestochowa University of Technology
Częstochowa, Poland
andrzej.grzybowski@pcz.pl, azgrzybowski@gmail.com

Received: 12 May 2025; Accepted: 29 August 2025

Abstract. Pairwise comparison-based inference for deriving priority weights has a long-standing history. However, its widespread application in multiple-criteria decision analysis began in the early 1980s with the introduction of the Analytic Hierarchy Process (AHP), a decision-making framework proposed by T. Saaty. A core tenet of the AHP is the reciprocity property, which is widely accepted and often treated as fundamental in constructing pairwise comparison matrices. Nonetheless, only a few studies have questioned whether this artificially imposed reciprocity might, in fact, degrade the quality of prioritization outcomes. This paper critically examines the validity of the reciprocity assumption and its impact on the accuracy of priority weight estimation. Through extensive computer simulations involving various prioritization-quality metrics, we find that enforcing reciprocity in pairwise comparisons often leads to significantly poorer prioritization outcomes. These results cast serious doubt on the usefulness of the reciprocity axiom in practical applications. We also discuss the methodological implications of relaxing this assumption and propose modified AHP methods that support more flexible and potentially more accurate judgment structures in the AHP practice.

MSC 2010: *68U35, 90B50*
Keywords: *pairwise comparison, reciprocity, prioritization quality, final ranking, AHP*

1. Introduction

Pairwise-comparison-based inference about objects' priority weights has a long history; some authors suggest that it perhaps started in the Stone Age, and certainly was known and applied in the XVII century, see e.g. [1]. Although it is a long-standing idea, the rapid growth of its modern applications in multiple-criteria decision analysis began only in the early 1980s, with the emergence of the Analytic Hierarchy Process (AHP) – a decision making methodology proposed by T. Saaty in his seminal work [2]. Since then, the AHP has become one of the most widely used approaches to prioritization [3]. Its range of applications is vast, but it is not the aim of this paper to review studies describing real-world decision making problems solved using AHP. However, to illustrate the breadth of its applicability, we mention just a few recent examples from entirely different fields where the methodology has been applied: outsourcing and engineering [4], transportation [5], consumer preference studies [6], the ranking of industrial technologies [7], and water management [8].

Almost all theoretical and application-oriented papers that adopt the AHP emphasize that the reciprocity of paired judgements is a *fundamental requirement*. Perhaps all of Saaty's papers emphasize the importance of the reciprocity requirement. It is even called an axiom of the AHP. For example, in Satty's paper [9], at its very beginning, it reads that Axiom No. 1 of the AHP is: "… the reciprocal property that is basic in making paired comparisons …". In his other, more recent work [10], it also reads "… thus, for example, if one stone is judged to be five times heavier than another, then the other is automatically one fifth as heavy as the first because it participated in making the first judgment. The comparison matrices that we consider are formed by making paired reciprocal comparisons." Yet another important work, the Encyclopedia of Information Science and Technology [11], which also addresses multiple-criteria decision making in its section on the AHP, contains a similar passage "… To carry out these comparisons matched, AHP is supported in three axioms: Reciprocity, Homogeneity, and Synthesis", and further on in the text: "The comparative relation of two elements or criteria between itself, opposite to his criterion mother *must be reciprocal*. That is to say, if A is k times more important than B, then B must be $1/k$ times more important than A." The reciprocity axiom is treated as an obvious one also in papers [12-14], and the book [15], and a plethora of other pairwise-comparisons-related works published during the years. It proves that the reciprocity axiom has already become a scientific paradigm, i.e., as stated by Thomas Kuhn: a principle that is "taken for granted by the scientific community in a given field" [16].

Is the reciprocal property of the *comparison matrices* so natural and obvious that the paradigm cannot be questioned?

Before embarking on a deeper exploration of this issue, and to clarify our point, let us take a closer look at one of the statements mentioned earlier:

"If A is k times more important than B, then B must be $1/k$ times more important than A."

At first glance, this appears to be a reasonable claim. However, a more thorough analysis of *the context in which it is applied* raises doubts about its validity. It is important to recognize that this property, in the context of the AHP methodology, does not describe the actual nature of the relationship between elements. Instead, it is a condition imposed on the elements of the comparison matrix, which, by definition, is constructed based on human judgments.

Below, we offer several remarks that call into question the self-evidence of the reciprocity requirement.

Remark 1 Let us note that, in practice, the decision maker (DM) must express their judgments using numerical values from a predefined, finite set known as the judgment scale. The most widely used is Saaty's scale (SS), which consists of the integers 1, 2, ..., 9 and their reciprocals.

Now, consider a *thought experiment*: suppose the true value of k is 1.45 (such values are entirely plausible – priority weights in many AHP studies are estimated to several decimal places). Even a perfectly rational DM must select a value from the scale, and it is reasonable to choose the closest available option, which is 1. The reciprocal of 1 is also 1. However, the true reciprocal of 1.45 is approximately 0.69, and the closest value on the scale to this would be 1/2.

Consequently, a perfectly rational decision maker striving to approximate the true priority ratios may inadvertently violate the reciprocity property. Conversely, strictly enforcing reciprocity can lead to the loss of valuable information – potentially degrading the accuracy of the final priority estimates. Similar dilemmas may arise in many practical situations.

Is this trade-off significant for real-world decision making? Do the distortions introduced by enforcing reciprocity materially affect the resulting prioritization? A numerical example illustrating such a case is presented in Section 3.2, where these questions are additionally examined more thoroughly through simulation experiments.

Remark 2 The previously considered statement is a conditional sentence that begins with the clause: "If A is k times more important than B." However, in practice, the decision maker does not know this with certainty – they merely estimate that "A is k times more important than B." In reality, such judgments are often imprecise or prone to error. Given this uncertainty, we should be even less confident in the assertion that "B must be $1/k$ times more important than A." Therefore, requiring the DM to provide a separate judgment approximating $1/k$ could yield additional, valuable information that should not be disregarded.

This expectation is also supported by statistical estimation theory. It is well known that even if a value a is a good estimate of some quantity α, it does not follow that $1/a$ is an equally good estimate of $1/\alpha$.

Consider another example: suppose we know that one object (say, A) is exactly three times longer than another object (say, B), but we do not know their actual lengths. Would it be reasonable to estimate the length of only one of them in order to infer both? According to estimation theory, it is not. Separate measurements provide better estimates, and the same logic applies to pairwise judgments in AHP.

Remark 3 In many – perhaps most – real-world applications of the AHP, decision makers are not asked to express their judgments about the relative importance of alternatives using numerical values, but rather in linguistic terms (see, e.g., [2, 17]). A predefined mapping is then used to assign numerical values from a selected scale to these verbal expressions. This approach is especially common when evaluating quality-related criteria – for instance, in car comparisons involving attributes such as suspension quality, vehicle safety, control equipment, or environmental friendliness.

In such contexts, the DM does not typically state, *"A is 7 times more important than B"*, but instead provides a verbal judgment such as, *"A is very strongly more important than B"*. In [2], for example, the term "*very strong importance*" is mapped to the numerical value 7.

Now, let us revisit the issue of the presumed obviousness of the reciprocity principle. What is the reciprocal of *"very strong importance"*? Or of any other linguistic term presented to the DM as a possible expression of their judgment? Would it not, in some cases, be more appropriate to ask the DM to also provide a linguistic approximation of the inverse relation? After all, the more insight we gain into the DM's perception of the relative importance of the elements being compared, the more likely it is that they will find the resulting priority vector satisfactory.

In our view, the above considerations make the reciprocity axiom appear far less natural and unquestionable than it might initially seem. However, to avoid misunderstanding, we want to emphasize that we are not arguing against the theoretical principle of reciprocity of the *underlying true ratios*. Rather, our focus is on scrutinizing the assumption that if a DM *estimates* – or merely senses – that a given priority ratio is, say, k, then it is automatically valid to consider $1/k$ as the best estimate of the inverse ratio. This assumption is a core component of the reciprocity paradigm as implemented in standard AHP procedures, yet we believe its validity is far from self-evident.

Indeed, over recent years, there have been several authors who claimed that humans in real-world problems provide paired judgments that are usually non-reciprocal, and suggested that the artificially imposed reciprocity may lead to loss of information about the decision maker's real preferences [1, 18-21]. Our paper focuses on a deeper numerical study of that issue. We propose a simulation framework to examine the impact of the artificially forced reciprocity on the quality of final priority-weight estimates. In our experiments, three different prioritization-quality characteristics are taken into account. It turns out that the simulations prove that the forced reciprocity of paired judgments leads to significantly worse estimation results for all quality criteria. Then, we discuss modifications of the usual AHP approach, which are necessary because of the repudiation of the reciprocity paradigm.

The remainder of the paper is organized as follows. Section 2 presents the necessary formal definitions. Section 3 describes the adopted prioritization-quality criteria and the simulation framework in more detail. Section 4 reports the simulation results and provides a discussion of the findings. In Section 5, we propose modifications to existing inconsistency indices and evaluate their performance through computer simulations. Additionally, we introduce a simple yet statistically well-justified acceptance procedure for pairwise comparison matrices (PCMs). Finally, in the *Final remarks and recommendations* section, we reflect on the issue of inconsistency from the broader perspective of the objectives of prioritization methods, and we highlight our most important conclusions.

2. Basic notions and facts

A vector $\boldsymbol{v} = (v_1, \dots, v_n)^T$, which consists of numbers that reflect the intensity of importance of each alternative concerning a given criterion, is called a priority vector (PV), while its components are called priority weights (or briefly: the *priorities*). The PV is assumed to possess only positive components and is usually normalized to unity. The result of comparisons of each pair of decision alternatives is the so-called pairwise comparison matrix (PCM), $\boldsymbol{A} = [a_{ij}]_{nxn}$ with a_{ij} being the DM's judgments about the ratios v_i/v_j.

Obeying the reciprocity condition, in the AHP the data for the PCM is collected only for the elements in the upper triangle of the matrix $\boldsymbol{A}$, while the rest of its elements are set as the inverses of the corresponding symmetric elements in the upper triangle i.e. $a_{ij} = 1/a_{ji}$ for all $i > j$. Any PCM with elements that satisfy such a condition is said to be a *reciprocal* one.

Any three elements (a_{ij}, a_{jk}, a_{ik}) $\forall i, j, k = 1, \dots, n$, of a PCM are called a *triad* [30].

A PCM is called a *consistent* one if it is reciprocal and for all its triads, the following condition holds:

$$a_{ij}a_{jk} = a_{ik}, \forall i, j, k = 1, \dots, n$$

A PCM that contains the DM's judgments about the priority ratios forms a basis for the priority weights estimation, i.e. for prioritization. In the AHP literature, two prioritization methods can be distinguished as the most widely applied in practice. One of them is the geometric mean method (GM). The estimated priority vector (EPV) in the GM method can be obtained by the following formula:

$$w_i = \left(\prod_{j=1}^{n} a_{ij}\right)^{\frac{1}{n}} \Bigg/ \sum_{k=1}^{n} \left(\prod_{j=1}^{n} a_{kj}\right)^{\frac{1}{n}}$$

Another commonly used prioritization method is based on some specific results from the spectral theory. This one is called the right eigenvector method (REV) [2]. The description of the REV method can be found in the vast literature, including university textbooks devoted to multiple-criteria decision making. Very briefly, the REV method suggests taking as the EPV, the normalized eigenvector associated with the principal eigenvalue of the PCM, i.e., a normalized vector w that satisfies the equation:

$$\boldsymbol{Aw} = \lambda_{\max}\boldsymbol{w}$$

where $\lambda_{\max}$ is the principal eigenvalue of the PCM.

That definition makes sense because the famous Perron's theorem assures the uniqueness of such a normalized $\boldsymbol{w}$ and the positivity of all its components.

Over the last few decades, many papers have been devoted to the comparison of the prioritization methods' performance. In those studies, it is found that the results provided by methods GM and REV differ very little, and it is not known which one is better – both have their supporters and rarely serious opponents, see e.g. [22, 23]. Because it is easier to use and numerically more precise, the GM method is adopted in these studies.

During the analysis of the typical AHP problem, the prioritization methods are used several times (at least for the criteria and for the decision alternatives for each criterion separately). Let n, $n > 2$ be the number of decision alternatives, and k; $k > 2$, the number of criteria.

Let us denote by PCM(Cr) the PCM that was provided by the decision-maker for the criteria, and by the symbol PCM(i) the PCMs obtained for the decision alternatives for the i-th criterion. The PCM(Cr) has order k, and each PCM(i) has order n. Let $\boldsymbol{v}^0$ and $\boldsymbol{w}^0$ be, respectively, the true PV for criteria and the EPV for the criteria that was obtained based on PCM(Cr). Let $\boldsymbol{v}^i$ and $\boldsymbol{w}^i$ be also the true PV and its EPV for the alternatives in regard to the i-th criterion. The final true PV – say $\boldsymbol{v}$ – is given by:

$$\boldsymbol{v} = \sum_{i=1}^{k} v_i^0 \boldsymbol{v}^{\boldsymbol{i}} \tag{1}$$

Analogously, the estimated final PV – say $\boldsymbol{w}$ – is defined as

$$\boldsymbol{w} = \sum_{i=1}^{k} w_i^0 \boldsymbol{w}^{\boldsymbol{i}} \tag{2}$$

Now let us focus on the issue of DM's judgments contained in the PCM. As we already emphasized, conventionally, the DM's opinions about priority ratios are expressed in linguistic terms, and then the terms are mapped into numbers belonging to a given, predefined scale. As a consequence of such a technique, the elements of PCM are hardly believed to give priority ratios precisely. Moreover, other possible judgment errors result from the human brain's natural limitation. Accordingly, the final EPV is also erroneous. The principal goal of any improvement of the prioritization process is to make those estimation errors as small as possible. If the consequences of estimation errors differ, then this process should also take into account those errors' repercussions. Three characteristics of the prioritization-quality that are used in our simulations are defined in the next section.

3. Description of simulation experiments

In this section, we present the simulation algorithm and the overall framework used to investigate the relationship between errors in priority ratio judgments, measures of judgment inconsistency, and errors in the estimation of the priority vector. We begin by introducing the prioritization quality characteristics adopted in our analysis.

3.1. Prioritization-quality-measures

There are two main purposes of the real-world applications of the AHP. The first and still very common goal is to point out the best alternative (the best location for new industrial investment, the best production technology for specific goods, etc.). In such a case, the prioritization errors are not very meaningful as long as the final ranking is correct. Another purpose is to allocate limited resources between several entities (e.g. to allocate constrained financial resources between different scientific projects). In the second case, the parts of resources granted to particular entities are proportional to their priority weights. In such problems, the accuracy of the estimates of the weights is essential. Taking those possible goals into account in our studies, we adopt three prioritization-quality characteristics (PQCs) defined below.

Let $\boldsymbol{v} = (v_1, \dots, v_n)$ be the true PV whilst the $\boldsymbol{w} = (w_1, \dots, w_n)$ its EPV. In various papers (e.g. [22, 24]) it was proposed to measure the PV estimation errors (PVEE) as the average absolute (AE) and relative (RE) errors. The errors are defined as follows:

average absolute error: $$\mathrm{AE}(\boldsymbol{v}, \boldsymbol{w}) = \frac{1}{n}\sum_{i=1}^{N} |v_i - w_i| \tag{3}$$

average relative error: $$\mathrm{RE}(\boldsymbol{v}, \boldsymbol{w}) = \frac{1}{n}\sum_{i=1}^{N} \frac{|v_i - w_i|}{v_i} \tag{4}$$

In [25], it was proposed to take into account also the chances for "significantly incorrect" final EPV (understood as the relative frequencies of such events). They consider the final EPV $\boldsymbol{w}$ to be "significantly incorrect" if the truly best alternative (i.e. the one associated with the greatest component of $\boldsymbol{v}$) is not the best one in the estimated final ranking (i.e. its corresponding component in $\boldsymbol{w}$ is not the greatest one) and the mistake is a "serious one". The final ranking is considered as seriously wrong if the weights of the two best alternatives differ significantly. More precisely: if i is the number of truly best alternative (and v_i is its true weight) and j is the number of the best alternative accordingly to the EPV, $i \neq j$, then we say that EFPV gives *significantly incorrect* PV (SIPV) when the following condition holds (compare [25]):

$$v_i - v_j \geq \mathrm{d1} \ \text{ or } \ w_j - w_i \geq \mathrm{d2} \tag{5}$$

where d1, d2 are given positive numbers.

So, in this concept, even if the estimated-best-alternative is not the true indeed, the mistake is not considered as a serious one when the truly best alternative possesses almost the same final weight. The term "almost the same" is specified by the constants d1 and d2.

The *third* prioritization-quality measure taken into account in our simulation experiments is the *chances of receiving the* SIPV. In our studies, we assume various values of the parameters d1 and d2 (from 0 to 0.025).

3.2. Simulation framework

To analyze the impact of imposed reciprocity on prioritization quality, we employ a simulation framework based on the concept presented in [25]. The motivation for such a simulation-based analysis was discussed in papers [24, 25]. However, to make this paper self-contained, we also provide an additional numerical example here to clarify our approach.

As a thought experiment, let us assume that the true priority vector is $\boldsymbol{v} = (0.45, 0.32, 0.23)^{\mathrm{T}}$. The corresponding true pairwise comparison matrix (TPCM) is then:

$$TPCM = \begin{bmatrix} 1 & 1.406 & 1.957 \\ 0.711 & 1 & 1.391 \\ 0.511 & 0.719 & 1 \end{bmatrix}$$

Suppose that a perfect decision-maker approximates these values using Saaty's discrete scale. The resulting pairwise comparison matrix (RPCM) would be:

$$RPCM = \begin{bmatrix} 1 & 1 & 2 \\ 0.5 & 1 & 1 \\ 0.5 & 0.5 & 1 \end{bmatrix}$$

This is the best the decision-maker can do under the constraint of the scale, but the matrix is **nonreciprocal**. Since the AHP methodology requires **reciprocal** matrices, we must instead work with the following adjusted matrix:

$$RPCM = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 1 & 1 \\ 0.5 & 1 & 1 \end{bmatrix}$$

Now, let us compare the estimated priority vectors obtained in both cases.

In the reciprocal case, we can apply both the GM and REV methods. In this specific case, both yield the same result: $\boldsymbol{rw} = (0.413, 0.327, 0.260)^{\mathrm{T}}$.

In contrast, when using the more accurate but nonreciprocal matrix (NPCM), only the GM method is applicable. The resulting estimate is $\boldsymbol{nw} = (0.469, 0.296, 0.235)^{\mathrm{T}}$. The corresponding estimation errors are:

$$\mathrm{AE}(\boldsymbol{v}, \boldsymbol{rw}) = 0.079, \ \ \mathrm{AE}(\boldsymbol{v}, \boldsymbol{nw}) = 0.047,$$

$$\mathrm{RE}(\boldsymbol{v}, \boldsymbol{rw}) = 0.025, \ \ \mathrm{RE}(\boldsymbol{v}, \boldsymbol{nw}) = 0.016.$$

We observe that, regardless of which error measure is considered, the estimation is significantly better when the nonreciprocal matrix is used. This is an important observation. However, one might argue that this outcome could be a coincidence, resulting from a specific choice of the true priority vector.

To address this concern, we design a simulation experiment in which thousands of such "examples" are generated and analyzed using statistical methods. Let us now describe the simulation framework in more detail.

Within that framework before the simulations start, we set the following: the number of alternatives (n), the number of criteria (k), the number of simulated different AHP problems (N), the probability distribution (PD) of PCM-elements-judgment--errors, and the interval being the support for the "big" judgment errors. In the course of the simulation experiments at the very beginning, $(k+1)$ "true" priority vectors are generated along with the corresponding "true" pairwise comparison matrices. Then the elements of these perfect PCMs are randomly disturbed with the help of multiplicative random value, in literature sometimes called a "perturbation" factor, so that the PCM elements and the true priority weights are related in the following way:

$$a_{ij} = \varepsilon_{ij} \frac{w_i}{w_j} \tag{6}$$

where ε_{ij} is the perturbation factor which is expected to be near 1, e.g. [22-24, 26, 27].

Those estimation errors are independently distributed according to the PD. Apart from those perturbation-factor-errors, randomly selected elements are additionally disturbed by "big" judgment errors that, as suggested in the literature, may result from the questioning procedure, erroneous entering of the data, etc (see [22, 23, 25]). In our simulations, the big error occurs with predefined probability Pr(BE). In such a way we randomly produce PCMs that contain erroneous judgments about priority ratios. Next, two cases are simulated. First, the non-reciprocal case, where elements of those PCMs are rounded to the closest values from the adopted judgment scale. Second, the reciprocal case where additionally, complying with the reciprocity paradigm, all elements in the lower triangle of the matrices are replaced with the inverses of the corresponding symmetric elements from the upper triangles. Based on those rounded PCMs all related priority vectors are estimated with the help of the GM method. Then the estimation errors, AE and RE, as well as an indicator of "significantly incorrect" final PV are stored in one record along with various other characteristics of the generated AHP problem. Finally, all records are returned in one database. The consecutive steps conducted in our simulation experiments are the following:

S1 Randomly generate true priority vectors $\mathbf{v}^i$; i = 0,…,k, along with related true comparison matrix $\boldsymbol{M}^i = [m^i_{j,l}]$, with $m^i_{j,l} = v^i_j / v^i_l$. According to Eq. (1) compute the final true priority vector $\mathbf{v}$.

S2 According to Eq. (6) modify independently the elements of the matrices $\boldsymbol{M}^i$, $i = 0, \dots, k$ with the help of perturbation factor that possesses the distribution *PD.* The new matrices denote as $\boldsymbol{MM}^i = [mm^i_{j,l}]$.

S3 In matrices generated in step S2, with probability Pr(BE) replace elements $mm^i_{j,l}$ with $B \cdot m^i_{j,l}$, where *B* (the "big" error) is randomly drawn from the predefined interval [LB, RB]. The matrices received after steps S2 and S3 denote as $\boldsymbol{DM}^i$.

S4 Compute matrices $\boldsymbol{NRM}^i$, i = 0,…,k, by replacing all elements in the matrices $\boldsymbol{DM}^i$ with the closest values from the adopted judgment scale.

S5 Compute matrices $\boldsymbol{RM}^i$, $i = 0,\ldots,k$, modifying matrices $\boldsymbol{NRM}^i$, by replacing all elements in their lower triangles with the inverses of the corresponding symmetric elements in the upper triangles i.e. $rm^i_{j,l} = 1/nrm^i_{l,j}$, $j > l$.

S6 On the basis of the reciprocal matrices $\boldsymbol{RM}^i$, $i = 0,\ldots,k$, compute the estimated priority vectors $\mathbf{rw}^i$ with the help of the GM method.

S7 On the basis of the non-reciprocal matrices $\boldsymbol{NRM}^i$, $i = 0,\ldots,k$, compute the estimated priority vectors $\mathbf{nrw}^i$ with the help of the GM method.

S8 According Eq. (2) compute the final estimated priority vectors **rw** and **nrw**, for reciprocal and nonreciprocal cases, respectively.

S9 According the formulae (3) and (4), compute the estimation errors $AE(\mathbf{v}, \mathbf{rw})$, $RE(\mathbf{v}, \mathbf{rw})$ and $AE(\mathbf{v}, \mathbf{nrw})$, $RE(\mathbf{v}, \mathbf{nrw})$. Additionally, record which one of the estimated PVs, **rw** or **nrw** is better (with regards to each of the errors separately).

S10 Set indicator $rsi = 1$, if the final estimated priority vector **rw** is significantly incorrect, otherwise set $rsi = 0$. Set indicator $nrsi = 1$, if the final estimated priority vector **nrw** is significantly incorrect, otherwise set $nrsi = 0$.

S11 Write down values of all prioritization quality characteristics computed and/or set in steps S9 and S10, along with other problem characteristics as one record.

S12 N times repeat Steps S1 to S12.

S13 Return all records in one database file.

In our experiments described here, the number of decision alternatives varies from 3 to 8, while the number of criteria k takes on values from 3 to 6. In step S1, the coefficients of the true PVs are generated independently according to the uniform distribution on the interval (0, 1), next the PVs are normalized to 1. In step S2 the probability distributions (PD) of the multiplicative perturbation factor involve lognormal, gamma, uniform, and truncated normal ones. Such distributions of the disturbances are often considered in the literature. The perturbation factor distributions have expected values equal to 1, while their support is a subset of the interval DS = [0.5, 1.9]. In our experiments, any matrix can be also disturbed by big multiplicative error B (step S3), and the probability of its occurrence during any single judgment is Pr(BE) = 0.05, while its support is the interval [2, 5]. In step S10, to determine the significantly wrong ranking, we adopt the condition (5) with d1 = d2 = 0.001. The latter parameters' values say that if the weights of the best and second-best alternatives are the same up to three decimal digits, then the differences between them are insignificant for us.

4. Results of the simulation experiments

The first remark in the Introduction concerns the fact, that the non-reciprocity of paired comparisons may arise solely due to the rounding errors, i.e. even perfect DM in some cases has to produce the non-reciprocal PCM if he/she wants to be as precise as possible. In such a case, interesting questions related to that observation can be asked: whether such errors, originated solely due to the rounding procedure are meaningful for the prioritization outcomes, and, whether the imposed reciprocity has any significant impact on the magnitude of those errors. In order to answer the questions we conduct our simulations with steps S2 and S3 being omitted, so that the only source of judgment errors is the rounding procedure in step S4 (in the non-reciprocal case), and possibly the forced reciprocity in step S5 (in the reciprocal case). Selected statistics of the considered types of errors as well as the frequency (i.e. the estimated probability) of the SIPV in the considered simulation setup are presented in Table 1. For both types of errors, the AE and RE, the mean values are presented as well as the quantiles of order 0.9, denoted $Q_{0.9}$. Because the type of an judgment scale adopted directly affects rounding errors, in addition to the Saaty scale (SS) already introduced here, our studies also make use of other scales suggested in the literature: the so-called Extended Saaty scale (ESS$[N]$) and the geometric scale (GS$[c]$). The ESS$[N]$ contains integers from 1 to N ($N > 9$), along with their reciprocals. Note that the SS is the ESS[9], so the ESS[17] that is used in our simulations is a much richer one. The geometric scale GS$[c]$ contains numbers s of the form $s = c^{i/2}$, $i \in \mathbf{I}$, with $\mathbf{I}$ being a predefined set of integers. In our studies, we assume $c = 2$ and $\mathbf{I} = \{0.1, 2, \ldots, 8\}$ so that we have the same number of possible different judgments as in the SS. It is worth emphasizing once again that, in practice, the DM's judgments are initially expressed in linguistic terms, after which the corresponding numbers from the adopted scale are entered into the PCM. For the analysis of various aspects of the scales usage see e.g. [17].

Table 1. Selected statistics of the PQC for the Reciprocal Case (RC) and Non-reciprocal Case (NRC). Rounding-Errors-Only Case. Results based on the WDB1

	Absolute Errors		Relative Errors		SIPV
	Mean	$Q_{0.9}$	Mean	$Q_{0.9}$	Pr
Saaty's scale					
RC	0.0079	0.0124	0.0397	0.0602	0.1103
NRC	0.0062	0.0097	0.0317	0.0482	0.0825
Extended Saaty's scale ESS[17]					
RC	0.0077	0.0120	0.0382	0.0579	0.1054
NRC	0.0059	0.0093	0.0299	0.0451	0.0776
Geometric scale GS[2]					
RC	0.0042	0.0066	0.0211	0.0323	0.0527
NRC	0.0038	0.0059	0.0190	0.0291	0.0450

For each pair (n, k) the number of simulated AHP problems is 10000. We denote those sub-databases as $\mathrm{DB}(n, k)$, $n = 3, \ldots, 8$, $k = 3, \ldots, 6$. Consequently, for each number of alternatives n we have 40000 records, while the whole database $\bigcup_{n=3}^{8} \bigcup_{k=3}^{6} \mathrm{DB}(n, k)$ contains 240000 records related to different AHP setups. This database is denoted WDB1. Table 1 is based on those records. From that table, we learn that independently of the adopted judgment scale, all statistical characteristics of the PQCs are much better in the non-reciprocal case. Under Saaty's scale, the average absolute error AE of the final estimated PVs obtained in our experiments as well as the 90 %-quantile of these errors' values are about 27 % greater in the reciprocal case. Similarly, both the mean value and the 90 %-quantile of the relative errors RE are about 25 % greater in the reciprocal case. The harm that results from the forced reciprocity is even more impressive when we consider the frequency of occurrence of the SIPV, it amounts to 34 %. Very similar, even slightly greater losses are observed when we look at the results obtained for the ESS[17]. The errors that originate solely from the rounding procedure are the smallest ones in the case of Geometric scale GS[2]. But also for this scale the statistics confirm that the prioritization quality characteristics are better for the non-reciprocal case.

So we see, that even in such an ideal situation where only rounding errors are imposed on the perfect PCMs, the gain in prioritization quality due to rejection of the reciprocity property is very significant.

Table 2. Selected statistics of the PQC for the Reciprocal Case (RC) and Non-reciprocal Case (NRC). Simulated-Judgment-Errors Case. Results based on the WDB2

	Absolute Errors		Relative Errors		SIPV
	Mean	$Q_{0.9}$	Mean	$Q_{0.9}$	Pr
Saaty's scale					
RC	0.0264	0.0441	0.1406	0.2298	0.2034
NRC	0.0216	0.0357	0.1142	0.1840	0.1579
Extended Saaty's scale ESS[17]					
RC	0.0281	0.0475	0.1525	0.2547	0.2107
NRC	0.0227	0.0377	0.1213	0.1979	0.1633
Geometric scale GS[2]					
RC	0.0269	0.0458	0.1452	0.2447	0.2016
NRC	0.0221	0.0369	0.1180	0.1934	0.1595

Now, let us turn our attention to the "Simulated-Judgment-Errors Case", i.e. the case where the perfect PCMs, generated within our simulation experiments, are disturbed by random "judgment-errors", as described in steps S2 and S3, in the previous section. The new database – denoted WTB2 – obtained in this situation also consists of 240000 records and has got the same structure as in the "rounding-errors-only" case. Table 2 is based on that database. We see that the results are in agreement with the previously presented ones. All the considered prioritization quality characteristics are worse in the case with imposed reciprocity. This time, the loss in quality caused by the reciprocity axiom amounts to 22 %-29 % dependently on the PQC, regardless of the adopted judgment scale. During the simulation experiments for each simulated AHP problem, at step S9, we also check in which case, reciprocal or non-reciprocal, the estimated final PV is better in terms of the errors AE and RE.

The results are the following: for the Saaty's scale, regarding the AE, the PV estimated on the basis of the reciprocal PCMs are better in 32 % of all problems, while the ones estimated based on non-reciprocal PCMs are better in 63 % of cases. Analogous numbers related to the RE are 33 % and 66 %, respectively. Note that these percentages do not sum up to 100 %, because the comparisons were done with the precision up to 3 decimal places, and with that accuracy, some PV-estimation-errors are the same for both cases. As we can see in Table 2, the results obtained for the scales ESS[17] and GS[2] are very similar.

As we have indicated earlier, Tables 1 and 2 are based on two databases that comprise all sub-databases $\mathrm{DB}(n,k)$ obtained for particular numbers of criteria and alternatives (n,k). At this point, it should be emphasized that for each of those sub-databases, similar results concerning the impact of the imposed reciprocity are obtained. For example in the reciprocal case, based on the sub-database $\bigcup_{k=3}^{6}\mathrm{DB}(6,k)$, for Saaty's scale, we obtain the mean values of AE and RE equal to 0.0186 and 0.1168, respectively, and the frequency of the SIPV is 0.3190. Analogous values related to the non-reciprocal case are as follows: 0.0147, 0.0913, and 0.2572, respectively. The complete comparison results obtained separately for databases $\mathrm{DB}(n,k)$, $n = 3,\dots,8$, $k = 3,\dots,6$, are presented in Tables A1-A3 in the Appendix. As the results for different scales are very similar in spirit, these tables contain only the ones received for Saaty's scale.

Given the presented simulation results, it seems undeniable that adoption of the reciprocity property is very costly in terms of the quality of the prioritization results. In the next section, we discuss the modifications of the prioritization techniques that would be a necessity in consequence of the repudiation of the reciprocity paradigm.

5. Consequences of the reciprocity-axiom rejection

One may wonder why the reciprocity axiom is so popular. After all, there are even more natural requirements (such as the ordinal transitivity condition) that are not explicitly incorporated into prioritization techniques, whereas reciprocity is. In our opinion, the main reason for making the reciprocity requirement compulsory is Saaty's *consistency index*.

Apart from the prioritization methods that allow us to estimate the final PV based on the PCMs, another crucial problem in the pairwise-comparison based inference is the so-called *consistency analysis*. Due to the rounding procedure and natural human limitations, it is quite obvious, that some level of judgments' incorrectness must be accepted. On the other hand, if the DM's judgment-errors are really serious, then the pieces of information contained in the PCM may be very misleading. The principal goal of the consistency analysis is to help the DM to distinguish between useful PCMs and the useless ones. What is the link between the two: the reciprocity paradigm and the consistency analysis? The link is fundamental, as the historically first and still most widely used approach to consistency analysis requires PCM reciprocity. The method is due to T. Saaty's [2]. To "measure" the degree of the PCM-inconsistency, he proposed to use a certain PCM-characteristic, called the "consistency index". That index denoted here as SI is closely related to the REV and is defined as follows:

$$SI(n) = \frac{\lambda_{max} - n}{n - 1}$$

where n is, as usual, the number of decision-alternatives (or other stimuli that one wants to rank).

It was a brilliant idea on Saaty's part, and the fact that the AHP provided the decision-makers with the method for the consistency analysis was its very important advantage and perhaps a major reason for its growing popularity.

Now, the crucial fact is, that the index SI works properly only if the PCM is the reciprocal one. In such a case, it can be proved that $SI(n) \geq 0$ and that $\underline{SI}(n) = 0$ if and only if the PCM is consistent, e.g. [2, 28]. However, if the PCM is not reciprocal, then the SI is meaningless, as it can be negative. In consequence, in this approach, the PCMs *have to* be reciprocal.

It is worth noting, in one of his papers, Saaty admitted that his index is called a "consistency index" just by misuse of language. In fact, it is an indicator of inconsistency, so the term "inconsistency index" seems to be more adequate, and such a term will be used in our paper.

It results from what we noted that inference based on the nonreciprocal PCMs excludes the usage of the SI. On the other hand, undeniably, the consistency analysis is very important in pairwise comparison based inference. But the question is whether this specific inconsistency index, the SI, is irreplaceable? Is it really necessary to obtain correct prioritization results? In our opinion, it is not. Over the last decades, literature has provided us with many competitive proposals for such indices e.g. [24, 29-31]. Most of them are easier to calculate and in contrast to the *SI*, they have a good underlying interpretation (as a matter of fact, the SI has no meaningful interpretation at all). Moreover, many of them are better related with the prioritization results, see e.g. [24, 25]. What is most important from our perspective, contrary to the SI, the definitions of many competitive indices can be easily adapted for the non-reciprocal case. Below, we present modifications of the two inconsistency indices most frequently used as competitors to SI. The first one, for the reciprocal case, was proposed in [29]. It is the so-called geometric mean index (GI), and it possesses very good properties [23, 25]. Its version for the non-reciprocal PCMs proposed in our studies is the following:

$$GI(n) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} \log^2(a_{ij} w_i / w_j)$$

where $\boldsymbol{w} = (w_1, \dots, w_n)$ is the EPV that was obtained by applying the GM method.

Another interesting index (KI) was proposed by Koczkodaj, who proposed characterizing the triad's inconsistency by the number that is given by the following simple formula, equivalent to the original one proposed in [30]

$$TI(a_{ik}, a_{kj}, a_{ij}) = \frac{|a_{ik}a_{kj} - a_{ij}|}{\max(a_{ik}a_{kj}, a_{ij})}$$

Koczkodaj's index KI of any reciprocal PCM is given as follows:

$$KI = \max\{TI(a_{ik}, a_{kj}, a_{ij})\},$$

where the maximum is taken over all triads in the upper triangle of the PCM.

That definition can be very easily extended to the non-reciprocal case. For such cases, it is enough to assume the value of KI as the maximum taken over *all* triads of the PCM.

To verify the usefulness of those indices in the non-reciprocal case, we study their relationship with the quality of the prioritization results, i.e., with PQCs considered here. Any good inconsistency index has to be well correlated with those characteristics; most preferably, those relationships should be monotonic ones. Following the approach presented in [24, 25], to study those relationships, the whole database WDB2 is arranged in ascending order concerning the values of a given inconsistency index and is next split into a number (say K) of separate classes – let us denote those classes as InC_i ($i = 1, \ldots, K$). For each class, InC_i the mean value of the considered index is computed as well as, for the corresponding PCMs, the mean values of errors (AE or RE) and the observed frequency (Fr) of SIPV. A more detailed description of those computations can be found in [20]. Table 3 presents the Spearman and Pearson correlation coefficients between the class mean values of the particular PQCs and indicated indices. In each case, the number of classes is $K = 20$. Although in the case of nonreciprocal PCMs Saaty's index SI is uninterpretable, its values can nevertheless be computed. So we put the correlation coefficients for the SI into Table 3 as well, just to let the reader know how the index behaves for such matrices.

Table 3. Spearman's and Pearson's correlation coefficients between the inconsistency indices SI, GI, KI and the PQCs: AE, RE and Fr. The non-reciprocal case, Saaty's scale

	Spearman's Correlation Coefficients		
	SI	GI	KI
AE	0.561	0.993	0.984
RE	0.559	0.994	0.984
Fr	0.386	0.960	0.924
	Pearson's Correlation Coefficients		
	SI	GI	KI
AE	0.830	0.986	0.969
RE	0.832	0.986	0.968
Fr	0.526	0.904	0.876

As we see, both indices GI and KI perform really well. The monotonic character of their association with the PQCs is close to perfect, as the values of the Spearman's correlation coefficients are close to 1. What is more, Pearson's coefficients are also very high, especially in the case of the GI. Thus we can propose the linear regression models that relate the GI with the PQCs. As presented in Table 3 the average coefficients' values are means of coefficients computed for each database DB(n, k) separately, the regression models are also developed for different numbers of decision-alternatives separately, while the number of criteria k is a second explanatory variable (apart from the GI). Note that in the models, the variable GI (the index value) is the mean of the indices GI computed in step S7 for all matrices $\boldsymbol{NRM}^i$, $i = 0,\ldots,k$.

Table 4 provides us with the estimated models for the frequency of the SIPV'c occurrence Fr, as well as for the errors AE and RE. Apart from the models' coefficients a, b, c, we also present two fundamental quality characteristics of the models: the coefficient of determination R^2 and the model-standard-error MSE. As we see, the models are very good, especially the ones for the absolute and relative errors.

Table 4. Models for Frequency of SIPV (Fr), the Absolute Error (AE), and the Relative Error (RE) in dependence of the inconsistency index GI and the number of criteria k

	$\mathrm{Fr} = a + b \cdot GI + c \cdot k$				
	a	b	c	R^2	MSE
$n = 3$	0.0602	0.2155	–0.00001	0.84	0.00029
$n = 4$	0.1154	0.2291	–0.00819	0.88	0.00044
$n = 5$	0.1435	0.2478	–0.01559	0.91	0.00040
$n = 6$	0.1418	0.2597	–0.01864	0.94	0.00029
$n = 7$	0.1334	0.2585	–0.02092	0.94	0.00029
$n = 8$	0.1261	0.2324	–0.02090	0.92	0.00030
	$\mathrm{AE} = a + b \cdot GI + c \cdot k$				
	a	b	c	R^2	MSE
$n = 3$	0.0343	0.0420	–0.00315	0.96	0.000003
$n = 4$	0.0297	0.0278	–0.00286	0.96	0.000002
$n = 5$	0.0250	0.0220	–0.00259	0.96	0.000002
$n = 6$	0.0200	0.0183	–0.00209	0.97	0.000001
$n = 7$	0.0166	0.0153	–0.00173	0.97	0.000001
$n = 8$	0.0142	0.0130	–0.00149	0.97	0.000001
	$\mathrm{RE} = a + b \cdot GI + c \cdot k$				
	a	b	c	R^2	MSE
$n = 3$	0.1171	0.1361	–0.0117	0.96	0.000034
$n = 4$	0.1313	0.1216	–0.0136	0.96	0.000051
$n = 5$	0.1364	0.1190	–0.0151	0.96	0.000056
$n = 6$	0.1289	0.1179	–0.0142	0.97	0.000037
$n = 7$	0.1252	0.1146	–0.0139	0.97	0.000035
$n = 8$	0.1214	0.1110	–0.0134	0.97	0.000031

Once we have those models built, one can use them in the PCMs acceptance procedure. If the DM is primarily concerned with the detection of the best alternative, then the models for the Fr should be used. If the correct estimation of each priority weight is the most important goal, then the models for the RE or AE can be adopted. Given the value of the index GI and the number of criteria, the associated value of the considered PQC can be easily approximated. The obtained value should then be compared with the level of risk deemed acceptable by the DM (i.e., the acceptable levels of Fr, AE, and/or RE). If the value is less than that level, then the PCMs defining the AHP problem should be accepted, otherwise, the DM should try to improve their consistencies. All those models are valid for the observed range of the explanatory variables, i.e. when the index GI is within the interval [0.1, 1.1] and $k = 3,\ldots,6$.

6. Final remarks and recommendations

We discuss here the results for the number of alternatives $n = 3,\ldots,8$. However with the help of our simulation framework, we have also performed studies for greater numbers of alternatives and/or criteria. All our results confirm the same observation: always considered prioritization quality characteristics are worse in the cases with imposed reciprocity. However, before concluding the paper, we would like to stress once again that we do not suggest abandoning such a legitimate theoretical property as the reciprocity of the underlying true priority ratios. We also do not challenge the idea of PCM consistency analysis – consistency is a very important and desired feature of any judgments. Our arguments are aimed solely against *artificially forced* consistency, that is the consistency that is assured by the adopted method of gathering information about priorities. Consistency is very welcome if it is a result of DMs' real opinions, unfortunately this occurs very rarely in real-world praxis. Obviously, it is reasonable to perform a consistency analysis in order to draw the DMs' attention to particularly inconsistent judgments and allow them to verify whether these "suspicious" opinions are correct. Thus, some specific consistency-improvement-oriented methods are reasonable. But forcing the reciprocity just by the method of collecting DMs' assessments about priority ratios is simply ignoring the pieces of information that would be provided in the additional judgments. It is important to remember that achieving consistency in the pairwise comparison matrix is not the primary goal of decision making analysis. However, many theoretically oriented studies seem to treat it as such. Is this truly the objective we seek? While a reciprocally consistent matrix may appear attractive, it can be more misleading than a nonreciprocal matrix that better reflects the true, often inconsistent, judgments of the decision maker. The judgments (which often stem from the decision maker's subjective feelings about relative importance) are inherently inconsistent. The goal should not be to enforce consistency artificially solely for the sake of making the matrix appear neat. Instead, we should aim to obtain priority weight estimates that most accurately represent the decision maker's (possibly inconsistent) perceptions.

This important aspect was also addressed in a number of papers, e.g. [24, 32], and recently in [33].

Saaty's idea of the consistency analysis was brilliant, and it started a new era in decision making analysis. However, from what we show, it results that his *specific* consistency index should be rejected along with the reciprocity axiom. We argue that this can be done without causing any harm to the prioritization results. On the contrary, the newly proposed models provide more effective tools for supporting consistency analysis and decisions regarding the acceptance of PCMs. They are very reasonable alternatives to the classical AHP approach. The decisions aided by those models are based on the classical statistical idea of taking into account the risk of making wrong decisions, a well-established concept in uncertain decision making. In contrast, the approach still supported by the classical AHP is to make use of a given "consistency threshold". Then the decision about the PCM acceptance is made without any participation of the DMs (e.g. their attitude towards risk is completely ignored). Such an approach in general, and particularly the specific choice of the thresholds were criticized in literature, see e.g. [24, 33, 35, 36].

However, while the procedure for PCMs' acceptance is an important component of the AHP methodology, it is not the primary focus of this paper. By developing these models, we demonstrate that rejecting the reciprocity requirement, along with the index *SI*, does not substantially harm the AHP decision-making process. The *SI* can be successfully replaced by other competitive indices.

To summarize the remarks about the reciprocity axiom: in our opinion, the popularity of the consistency index *SI*, the very first inconsistency indicator in literature, is the real reason why the reciprocity property became a paradigm of the AHP. And perhaps the only serious consequence of the repudiation of the paradigm of reciprocity is that the *SI* has to be rejected as well. However, *it can be done without any loss in the quality of the prioritization results*. On the contrary, if we use the non-reciprocal PCMs, those results are likely to be better.

In [16], Thomas Kuhn wrote that the idea of the repudiation of a paradigm, when proposed, is initially rejected in various forms by scientists in the given field. Since the methodology without the paradigm is initially neither well-developed nor well-understood, the position of its proponents is weak. However, in regard to the reciprocity paradigm, we think that it is worth trying to bring the attention of the decision makers' community to this problem.

References

[1] Fülöp, J., Koczkodaj, W.W., & Szarek, S.J. (2012). On some convexity properties of the least squares method for pairwise comparisons matrices without the reciprocity condition. *J. Global Optim.*, *54*(4), 689706.

[2] Saaty, T.L. (1980). *The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation* (*Decision Making Series*). McGraw-Hill.

[3] Zyoud, S.H., & Fuchs-Hanusch, D. (2017). A bibliometric-based survey on AHP and TOPSIS techniques. *Expert Syst. Appl.*, *78*, 158-181.

[4] Khan, A.W., Khan, S.U., Alwageed, H.S., Khan, F., Khan, J., & Lee, Y. (2022). AHP-based systematic approach to analyzing and evaluating critical success factors and practices for component--based outsourcing software development. *Mathematics*, *10*(21), 3982. DOI: 10.3390/math10213982.

[5] Mohamed, A., Stone, S., & Oloufa, A.A. (2024). Application of the analytical hierarchy process for autonomous truck strategies of commercial vehicles. *Appl. Sci.*, *14*(21), 9702. DOI: 10.3390/app142197.

[6] Chen, M.-S., & Liao, Y.-C. (2025). Applying the analytical hierarchy process to exploring demand and technology preferences in InsurTech: Focusing on consumer concerns. *Eng. Proc.*, *98*(1), 6. DOI: 10.3390/engproc2025098006.

[7] Abdel-Basset, M., Mohamed, R., & Chang, V. (2025). A multi-criteria decision-making framework to evaluate the impact of industry 5.0 technologies: Case study, lessons learned, challenges and future directions. *Inf. Syst. Front*, *27*, 791-821. DOI: 10.1007/s10796-024-10472-3.

[8] Hassoun Nedjar, N., Djebbar, Y., & Djemili, L. (2023). Application of the analytical hierarchy process for planning the rehabilitation of water distribution networks. *Arab. Gulf J. Sci. Res.*, *41*, 4, 518-538. DOI: 10.1108/AGJSR-07-2022-0110.

[9] Saaty, T.L. (1986). Axiomatic foundation of the analytic hierarchy process. *Manag. Sci.*, *32*, 7, 841-855.

[10] Saaty, T., & Kulakowski, K. (2016). Axioms of the Analytic Hierarchy Process (AHP) and its Generalization to Dependence and Feedback: The Analytic Network Process (ANP), https://arxiv.org/abs/1605.05777.

[11] Khosrow-Pour, M. (ed.) (2015). Encyclopedia of Information Science and Technology. 3 rd ed., IGI Global, ISBN 978-1-4666-5888-2.

[12] Bortot, S., Brunelli, M., Fedrizzi, M., & Pereira, R. (2023). A novel perspective on the inconsistency indices of reciprocal relations and pairwise comparison matrices. *Fuzzy Sets Syst.*, *454*, 74-99.

[13] Koczkodaj, W.W., & Urban, R. (2018). Axiomatization of inconsistency indicators for pairwise comparisons. *IJAR*, *94*, 18-29. DOI: 10.1016/j.ijar.2017.12.001.

[14] Brunelli M., & Fedrizzi, M., (2024). Inconsistency indices for pairwise comparisons and the Pareto dominance principle. *Eur. J. Oper. Res.*, *312*, 273-282.

[15] Kulakowski, K. (2022). *Understanding the Analytic Hierarchy Process*. Chapman & Hall/CRC Series in Operations Research. Chapman & Hall.

[16] Kuhn, T.S. (2012). *The Structure of Scientific Revolutions*. Chicago: Univ. of Chicago Press.

[17] Dong, Y., Xu, Y., Li, H., & Dai, M.(2008). A comparative study of the numerical scales and the prioritization methods in AHP. *Eur. J. Oper. Res.*, *186*, 229-242.

[18] Hovanov, N.V., Kolari, J.W., & Sokolov, M.V. (2008). Deriving weights from general pairwise comparison matrices. *Math. Soc. Sci.*, *55*, 205220.

[19] Diaz-Balteiro, L., Gonzlez-Pachn, J., & Romero, C. (2009). Forest management with multiple criteria and multiple stakeholders: An application to two public forests in Spain. *Scand. J. For. Res.*, *24*(1), 8793.

[20] Linares, P.(2009). Are inconsistent decisions better? An experiment with pairwise comparisons. *Eur. J. Oper. Res.*, *193*, 492498.

[21] Linares, P., Lumbreras, S., Santamara, A., & Veiga, A. (2016). How relevant is the lack of reciprocity in pairwise comparisons? An experiment with AHP. *Ann. Oper. Res.*, *245*, 227244.

[22] Choo, E.U., & Wedley, W.C. (2004). A common framework for deriving preference values from pairwise comparison matrices. *Comp. Oper. Res.*, *31*, 893-908.

[23] Lin, C-C. (2007). A revised framework for deriving preference values from pairwise comparison matrices. *Eur. J. Oper. Res.*, *176*, 1145-1150.

[24] Grzybowski, A.Z. (2016). New results on inconsistency indices and their relationship with the quality of priority vector estimation. *Expert Syst. Appl.*, *43*, 197-212.

[25] Grzybowski, A.Z., Starczewski, T. (2020). New look at the inconsistency analysis in the pairwise--comparisons-based prioritization problems. *Expert Syst. Appl.*, *159*, 113549. DOI: 10.1016/j.eswa.2020.113549, ISSN 0957-4174.

[26] Zahedi, F. (1986). A simulation study of estimation methods in the analytic hierarchy process. *Socio-Econ. Plann. Sci.*, *20*, 347-354.

[27] Saaty, T.L. (2003). Decision making with the AHP: Why is the principal eigenvector necessary. *Eur. J. Oper. Res.*, *145*, 85-91.

[28] Saaty, T.L. (2004). Decision making - the analytic hierarchy and network processes (AHP/ANP). *J. Syst. Sci. Syst. Eng.*, *13*, 1, 1-35.

[29] Crawford, G., & Williams, C.A. (1985). A note on the analysis of subjective judgment matrices. *J. Math. Psychol.*, *29*, 387-405.

[30] Koczkodaj, W.W. (1993). A new definition of consistency of pairwise comparisons. *Math. Comput. Model.*, *18*(7), 79-84.

[31] Salo, A., & Hämäläinen, R. (1997). On the measurement of preferences in the analytic hierarchy process. *J. Multi-Criteria Decis. Anal.*, *6*(6), 309-319.

[32] Temesi, J. (2011). Pairwise comparison matrices and the error-free property of the decision maker. *CEJOR*, *19*, 239-249.

[33] Kazibudzki, P.T. (2023). The uncertainty related to the inexactitude of prioritization based on consistent pairwise comparisons. *PLoS ONE*, *18*(9).

[34] Bana e Costa, C.A., & Vansnick, J.-C. (2008). A critical analysis of the eigenvalue method used to derive priorities in AHP. *Eur. J. Oper. Res.*, *187*, 1422-1428.

[35] Kazibudzki, P.T. (2021). On the statistical discrepancy and affinity of priority vector heuristics in pairwise-comparison-based methods. *Entropy*, *23*(9), 1150.

[36] Aguaron, J., & Moreno-Jimenez, J.M. (2003). The geometric consistency index: Approximated thresholds. *Eur. J. Oper. Res.*, *147*, 137-145.

Appendix

Tables presented in this Appendix show the results of the comparison of the prioritization quality characteristics obtained for the reciprocal and non-reciprocal PCMs. They are based on databases $\mathrm{DB}(n,k)$, $n = 3,\ldots,8$, $k = 3,\ldots,6$, obtained with the help of the simulation framework for the *judgment error case* (steps S1-S13) and described in Section 3. In the tables, the acronyms RC and NRC indicate the Reciprocal and Non-reciprocal Cases, respectively. The prioritization-quality characteristics AE, RE, and SIPV are introduced in Section 3.1.

Because the results for different judgment scales are very similar in spirit, these tables contain only the ones received for Saaty's scale.

Table A1. Comparison of the mean values of the average absolute error AE, defined by Eq. (3) for the Reciprocal and Non-reciprocal Cases. Results presented separately for the databases $DB(n, k)$

Average absolute error AE					
		$k = 3$	$k = 4$	$k = 5$	$k = 6$
$n = 3$	RC	0.0414	0.039	0.0361	0.0327
	NRC	0.0334	0.0316	0.0295	0.0267
$n = 4$	RC	0.0378	0.0356	0.0329	0.0305
	NRC	0.0309	0.0290	0.0266	0.0244
$n = 5$	RC	0.0314	0.0292	0.0273	0.0254
	NRC	0.0261	0.0241	0.0221	0.0204
$n = 6$	RC	0.0255	0.0238	0.0222	0.0206
	NRC	0.0214	0.0198	0.0182	0.0167
$n = 7$	RC	0.0214	0.0199	0.0184	0.0172
	NRC	0.018	0.0167	0.0154	0.0141
$n = 8$	RC	0.0181	0.0171	0.0158	0.0147
	NRC	0.0154	0.0143	0.0131	0.0121

Table A2. Comparison of the mean values of the relative absolute error RE, defined by Eq. (4) for the Reciprocal and Non-reciprocal Cases. Results presented separately for the databases $DB(n, k)$

Relative absolute error RE					
		$k = 3$	$k = 4$	$k = 5$	$k = 6$
$n = 3$	RC	0.1377	0.1272	0.1157	0.1040
	NRC	0.1096	0.1015	0.0934	0.0839
$n = 4$	RC	0.1649	0.1532	0.1397	0.1282
	NRC	0.1328	0.1224	0.1113	0.1016
$n = 5$	RC	0.1698	0.1559	0.1437	0.1328
	NRC	0.1388	0.1266	0.1149	0.1054
$n = 6$	RC	0.1634	0.1510	0.1395	0.1290
	NRC	0.1354	0.1244	0.1134	0.1033
$n = 7$	RC	0.1595	0.1469	0.1349	0.1251
	NRC	0.1326	0.1217	0.1111	0.1017
$n = 8$	RC	0.1539	0.1439	0.1321	0.1222
	NRC	0.1296	0.1192	0.1085	0.0998

Table A3. Comparison of the frequencies of SIPV, see Section 3.1, for the Reciprocal and Non-reciprocal Cases. Results presented separately for the databases DB(n, k)

Frequency of SIPV					
		$k = 3$	$k = 4$	$k = 5$	$k = 6$
$n = 3$	RC	0.1436	0.1486	0.1517	0.1434
	NRC	0.1093	0.1128	0.1168	0.1104
$n = 4$	RC	0.2186	0.2145	0.2114	0.2068
	NRC	0.1696	0.1713	0.1693	0.1596
$n = 5$	RC	0.2432	0.2315	0.2317	0.2210
	NRC	0.2017	0.1885	0.1776	0.1682
$n = 6$	RC	0.2372	0.2341	0.2220	0.2103
	NRC	0.1926	0.1832	0.1755	0.1554
$n = 7$	RC	0.229	0.2142	0.2012	0.1932
	NRC	0.1858	0.1694	0.1553	0.1389
$n = 8$	RC	0.2159	0.2022	0.1872	0.1693
	NRC	0.172	0.1514	0.1362	0.1202