

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



A Methodology for Evaluation and Distribution of Patent Applications to INPI-BR Patent Examiners: A Case Study in the Electricity Field

Cesar Vianna Moreira Júnior, Daniel Marques Golodne and Ricardo Carvalho Rodrigues

Abstract

This paper presents the development of a new methodology for evaluation and distribution of patent applications to the examiners at the Brazilian Patent Office considering a specific technological field, represented by classification of the application according to the International Patent Classification (IPC), and the variables corresponding to the volume of data of the application and its complexity for the examination process. After identifying the most relevant variables, such as the Specific Areas of Expertise (ZAE) of the examiners, a mathematical model was developed, including: (a) application of the principal component analysis (PCA) method; (b) calculation of a General Complexity Ratio (IGC); (c) classification into five classes (very light, light, moderate, heavy and very heavy) according to IGC average ranges and standard deviations; (d) implementation of a logic of distribution, compensating very heavy applications with very light ones, and light applications with heavy ones; and (e) calculation of a Distribution Balancing Ratio (IBD), considering the differences between the samples' medians. The model was validated using a sample of patent applications including, in addition to the identified variables, the time for substantive examination by the examiner. Then, a correlation analysis of the variables with time and a comparison of the classifications according to the time and the IGC generated by the model were carried out. The results obtained showed a high correlation of the IGC with time, above 80%, as well as correct IGC classes in more than 80% of applications. The model proposed herein suggests that the three main relevant variables are: total number of pages, total number of claims, and total number of claim pages.

Keywords: Patents, Applications, Evaluation, Distribution, Volume data, IPC, Examiner

1. Introduction

The granting of industrial property (IP) assets should be based on two central and interrelated principles: quality and efficiency. An efficient system to ensure

protection of IP rights is basically bound to the time taken to execute the granting procedures, but also to the clarity and organization of the analysis performed during the technical examination. On its turn, quality is usually bound to the standardization and improvement of administrative proceedings, providing reliability and legal certainty in decisions.

As to the granting procedures and time for granting industrial property rights, the Brazilian National Institute of Industrial Property – INPI-BR has been using its efforts to significantly reduce and adjust the granting time taken by the Institute to the average of major international offices. Among its actions, we note the implementation of the so-called “backlog combat plan” announced still in 2019, which is already in effect and expected to be in place until 2021 [1]. It is also worth highlighting the results and efforts regarding the time taken by the Institute to register trademarks, leading to the inclusion of INPI-BR among the signatories of the Madrid Protocol [2].

Over the last decades, INPI-BR has been incorporating to its Strategic Plan several initiatives related to the quality of its applications/registrations. The Strategic Plan 2018-2021 [3] provided for partnerships, cooperation agreements with international offices, improvement of prioritized examination programs, implementation and review of patent examination guidelines, among other activities. As to the technical cooperation initiatives, we note a recent strategic partnership between INPI-BR and the European Patent Office – EPO that includes training, discussion of best practices, sharing of tools, and exchange of patent databases [4].

It is important to highlight that there are studies [5–7] showing that changes in the workload of examiners can affect the quality of the examination process and its results, suggesting that a factor with high potential of causing instability in the applications/registrations and discrepancies in the decisions of a patent office is the unbalanced workload of patent examiners, i.e., an uneven distribution of patent applications among them. Additionally, it is also relevant to highlight that a specific study [8] using automatically data related to INPI-BR applications showed that the volume of patent applications can be used as a measure of examiner’s workload, and suggests that claim’s pages is probably a key workload indicator. Such facts become even more relevant since INPI-BR currently does not apply a well defined application distribution method, and the decision/responsibility for distribution is on behalf of the team leaders, who are somewhat free to apply their own criteria.

Thus, due to the importance of the subject for INPI-BR itself, specifically for the Patent Directorate of the Brazilian National Institute of Industrial Property – DIRPA, and to the scarcity of studies on methods for analyzing and distributing patent applications to the examiners, aiming at balancing the workload in an optimized manner, the development of a method in this regard would be a great contribution to the efficiency and quality of patent examination, being an important element to fulfill the mission of international Intellectual Property offices in general.

2. Patent document and the variables of interest to evaluate the distribution

2.1 Basic structure of patent applications

The Brazilian Industrial Property Law (LPI) [9], in its chapter III – article 19, defines that a patent application shall include: application; description; claims; figures, if any; abstract; and proof of payment of the application fee. Therefore, disregarding the abstract, which is usually limited to a single page, a patent

application is basically composed of three major parts: description, patent claims, and figures (whenever necessary). We note that these three main parts have specific shape and content characteristics that should be evaluated in more details.

2.1.1 Description

Regarding the description, article 24 of the LPI provides that: “The description shall describe the object clearly and sufficiently, as to enable reproduction by an expert, and indicate, when applicable, the best mode of execution.”.

Thus, the description includes the details of the applicant/inventor’s invention, clearly explaining its practical implementation to third parties and a person skilled in the art. Consequently, it is one of the most relevant parts of the patent application and is expected to have more pages. Therefore, the number of pages of description of a patent application is a variable relevant to its evaluation.

2.1.2 Claims

Regarding claims, article 25 of the LPI provides that: “The claims shall be based on the description, describe the particularities of the application, and provide a clear and accurate definition of the subject matter of the protection”.

We note that, within a same patent claim’s section, the claims may be classified as independent or dependent. Independent claims seek to protect essential and specific technical characteristics of the invention as a whole, while dependent claims are those that, while keeping the unity of invention, include all features of other previous claim(s) and define details of these characteristics and/or additional characteristics not deemed essential to the invention [10].

Therefore, according to the comprehensive content in the description, the claims specify, on a more accurate basis, the object of protection, i.e., define the scope of protection. In a recent study [11], the role of the characterizing term (the expression “characterized by”) in the patent documentation for delimitation of industrial property rights is discussed in more details considering the history of the Brazilian and international legal framework.

In this context, we note that both the extent of the claims (number of patent claim pages) and the number of claims (total number of independent and dependent claims) are highly relevant to the examination of a patent application.

2.1.3 Figures

As provided for in article 19 of the LPI, the submission of drawings/figures in the patent application is optional. While the figures are not mandatory, we note that the majority of applicants and inventors include them in the documentation submitted for examination. Such fact points that figures are an important part of the patent application, and one of the main reasons therefor is that they facilitate reading and understanding of the subject matter under examination. Hence, the number of figures pages can also be relevant when examining a patent application.

2.2 Cover sheet: bibliographic data of the patent documentation

Information included in patent documentation (including bibliographic data in the cover sheet) is an important tool for technological research and development [12]. It is important to note that information included in the cover sheet of a patent document is identified by numeric codes, as the Internationally Agreed Numbers

for the Identification of Data (INID), and its specific standards are defined by the World Intellectual Property Organization – WIPO.

Among the bibliographic data included in the cover sheet, we highlight: application number (21), number of the priority document, if any (31); filing date (22); priority date, if any (32); date of publication (43); name of the inventor (72); name of the holder of the rights on the patent (73); International Patent Classification (51); title of the invention (54); and abstract (57). A detailed description of these codes can be found in Appendix 1 of Standard ST.9 of the WIPO Handbook on Industrial Property Information and Documentation [13].

In this context, it is important to highlight that, as we intend to obtain variables relevant for evaluating and distributing patent applications for examination, among these bibliographic data, we are only interested in those already available from the filing until publication of the application, i.e., those that can be obtained before the patent examination process. Thus, in a first analysis, we focus on the identification numbers of the applications, filing date/year, inventors, priorities, and international patent classification (IPC), provided that the latter will be addressed in further details below.

2.3 International patent classification (IPC)

The IPC [14] allowed for standardization of documents from different countries with different languages and technical expressions. Pursuant to WIPO, it has an important role as it serves as: (i) an instrument for the orderly arrangement of patent documents aiming to facilitate access to the technological and legal information contained therein; (ii) a basis for selective dissemination of information to all users of patent information as a reference and/or knowledge; (iii) a basis for investigating the state of the art in given fields of technology; (iv) a basis for the preparation of industrial property statistics which in turn allows the assessment of technological development in various areas [15].

The IPC is divided into eight sections (A, B, C, D, E, F, G, and H). The sections are the highest levels of hierarchy of the classification. Each section is subdivided into classes, representing the second hierarchical level of the classification. Each class comprises one or more subclasses, indicating the third hierarchical level of the classification. Each subclass is broken down into subdivisions, referred to as “groups”, which are either main groups (fourth hierarchical level of the classification) or subgroups (lower hierarchical levels dependent upon the main group level of the classification). We present below an example classification of the electricity area by using all hierarchical levels, from section to subgroup.

Example: H02M 7/48.

Where:

H Section (A, B, [...], H): Electricity;

H02 Class (two digits): Energy Production, Conversion, or Distribution;

H02M Subclass (one letter): Apparatus for Conversion Between DC and AC;

H02M7 Main Group (one or more numeric digits): Conversion of AC power input into DC power output;

H02M7/48 Subgroup – (at least two numeric digits): using discharge tubes with control electrode or semiconductor devices with control electrode.

We highlight that the IPC is revised periodically based on meetings with experts from WIPO member countries, and such revision is published and may be accessed through INPI and WIPO websites. We further highlight that a patent document may fall under more than one IPC symbol. This is due to the fact that it can have claims in more than one category (device and method, for example), be related to one or more areas of application, or even to specific functions.

The study of Harloff and Wagner [16], based on EPO data for modeling the examination time, defined that, among other factors that impact the complexity of a patent application, the total number of claims and the number of IPC subclasses are major direct variables of the application. Therefore, the number of IPC subclasses is also a variable to be considered when evaluating a patent application to be distributed.

2.4 Studies using variables of patent documentation

2.4.1 Miscellaneous studies

As seen previously, a patent application is basically composed of three main parts: description, claims and figures (if any). For description and figures, for example, the number of pages is a relevant variable. As for patent claims, the number of pages, number of claims, whether independent or not, etc. Additionally, bibliographic data present information that also become potentially relevant variables, such as IPC classification, inventors, filing dates and years, etc. In this context, there are some papers that use specific variables of patent documentation to determine specifications like: economic value, time effort applied to the technical evaluation, scope of patent protection, among others.

An economic value approach to European patents using their claims, references, among others as variables, is described in [17].

Other papers [18, 19] are related to methods for analyzing and assessing the scope of protection of a patent based on variables related to the patent claim scope. The first uses the number of pages of the first independent patent claim as a relevant variable. The second uses two variables: the number of words in the shortest independent patent claim and the total number of independent claims.

However, none of the papers mentioned address the use of the parameters of the patent application or data in the examination process to create a method for distributing the applications to examiners. In [20, 21], studies were carried out with evidence that, at the United States Patent and Trademark Office –USPTO, after the application was directed to a large technological area (for example, electricity or chemistry), its subsequent designation to a specific examiner was nearly random. On the other hand, [22] found evidence that, at USPTO, applications of the same applicant, as well as applications with similar abstracts and titles, tend to be distributed to the same examiner. The authors suggest that, although such methods seek to follow the principle of efficiency, a balance with the principle of justice should be pursued as well. On the other hand, they emphasize that a random distribution favors the principle of justice, but fails to follow the principle of efficiency, which counts on the expertise of the examiners in certain examination subclasses.

In view of the foregoing, one has to take into account not only factors contributing to efficiency, such as distribution of application by subclasses pursuant to the examiners' education/interest, but also factors contributing to better balance, i.e., variables including the amount of data and the complexity of the patent applications to be examined. Consequently, justice should not be confused with randomness.

2.4.2 Workload balancing, voluminosity, and complexity of patent applications

As already seen, the workload in each patent office is one of the indicators affecting the quality not only of patent examination, but of patent systems as a whole. It is also worth mentioning that the increased workload of examiners, and

the consequent decrease in time for examining a patent application, may adversely affect the quality of patents granted, i.e., it tends to increase the granting of improper patents. Thus, it is important to seek better workload distribution and balancing to the examiners, as to reconcile working conditions and results.

Papers [23, 24] show that the volume of patent applications at EPO has been increasing throughout the years, and they relate the volume of patent application to an essentially two-dimensional problem, with two predominant variables: total number of pages of the patent application and number of claims. In this context, the total number of pages would be related to the level of description of the invention at stake and the total number of claims to protection scope. The authors also emphasize that these variables would be correlated, in different proportions, to the number of priorities, inventors, and IPC classifications. As limitations to the study, they mention that it would be better to evaluate the amount of independent and dependent claims, although this information is not easily accessible in data sources available.

To assess the hypothesis of technological complexity of a patent application, another study [25] uses a mathematical model considering the number of inventors, the number of IPC classes, and the number of references to previous patents. The authors emphasize that the increased volume of data may be related to both the technical complexity of the invention/application and a strategy matter, i.e., the desire to maintain certain know-how rather than actually protect it. It is important to highlight that, with the results of the model applied, the authors suggest that an increased number of inventors tend to create more complex inventions, requiring a greater number of pages and claims so that such inventions are described in details.

2.5 Substantive patent examination: main steps and related variables

2.5.1 Step 1: initial technical analysis

When the examiner receives a new patent application for examination, it is clear that the first step, which may takes a longer time, is reading the application. In principle, as the examiner needs to carefully read the entire application and see the figures, further confirming if the claims are based in the description and if the matter is sufficiently described so that a third person skilled in the art can execute it, the relevant variable to be considered in this step is the application's total number of pages.

2.5.2 Step 2: prior art search

Upon the Initial Technical Analysis, the examiner should carry a prior art search aiming at determining the state-of-the-art closer to the matter claimed. For the searches, the examiner should basically consider the patent claim scope, more specifically its independent claims. Additionally, if the claims include several specifications and are quite extensive, an even greater effort shall be employed. Finally, the greater the number of classifications in the application, the greater the tendency to address more than one technological area or borderline regions, which can also cause the search to be even more complex. Thus, the number of independent claims, the number of pages of the patent claims, and the number of subclasses of the application may also be variables to be considered in this step.

2.5.3 Step 3: specific technical analysis

Upon searches and determination of the useful prior art for examination, the third major step is the comparison between the matter claimed and the knowledge

presented in the state-of-the-art, i.e., analysis of the patentability. This third step is hereinafter referred to as Specific Technical Analysis. Upon analysis of patentability, a detailed examination of the independent claims in the application is mandatory. Thus, the first relevant variable in this step will be the number of independent claims. As in the searches, if the claims are quite extensive and have several technical specifications, the effort applied will be greater. So the number of claim pages is also a relevant variable in this step. As in some cases, dependent claims are also analyzed in details, the possibility of using the number of dependent claims is considered, and, in this case, the total number of claims could be used as a variable of interest.

2.6 Selected variables and initial hypotheses

Twelve possible variables were initially identified to be used in this paper: number of pages of description (Variable 1); number of claim pages (Variable 2); number of pages of figures (Variable 3); number of third parties observation pages (Variable 4); number of independent claims (Variable 5); number of dependent claims (Variable 6); number of IPC subclasses (Variable 7); year of filing (Variable 8); number of inventors (Variable 9); number of priorities (Variable 10); number of references to the state-of-the-art in the patent document under examination (Variable 11); number of references to the patent document under examination in other patent documents (Variable 12).

Upon identification of the possible variables of interest, the following additional criteria were established for selecting study variables:

- moment of availability of the variable - as the main objective is to evaluate the document for later distribution for examination, only variables accessible from filing until publication of the patent application will be selected;
- reliability and efficiency - variables failing to provide full reliability in data obtained, which may require the use of more than one platform/database and/or which require a very long time to be obtained, which could greatly increase the complexity of the model or even render its practical application unfeasible, will not be selected.

It should be noted that most of patent applications filed in INPI only include references to the state-of-the-art after a first examination, i.e., after the application has already been distributed, and references to other documents would still require access to more than one database. In this context, these variables are not deemed to fully meet the criteria of “moment of availability” and “reliability and efficiency”. Thus, at first, the variables related to references will be disregarded, hence ten variables will be used to obtain data samples, which are: number of pages of description (Variable 1); number of claim pages (Variable 2); number of pages of figures (Variable 3); number of third parties observation pages (Variable 4); number of independent claims (Variable 5); number of dependent claims (Variable 6); number of IPC subclasses (Variable 7); year of filing (Variable 8); number of inventors (Variable 9); and number of priorities (Variable 10).

Based on the studies carried out with EPO data on voluminosity (volume of data) of a patent application, it appears that the volume of data that the examiner needs to deal with during the examination of a patent application is one of the main constraints on examination effort/time. Additionally, such studies indicate that voluminosity is a problem related to two patent application variables: total number of pages and total number of claims. Thus, a first assumption will consider that the

volume of data of a patent application can be represented by the variables with the greatest positive correlation with the total number of pages of the application and/or with the total number of claims. Initial hypothesis 1 is that there are five variables directly bound to the volume of data: number of pages of description, claim pages, and pages of figures, in addition to the number of dependent and independent claims.

On the other hand, although most of the examination effort is bound to the direct volume of data that the examiner deals with, it is possible to note that there is still the hypothesis of existence of complementary variables, bound to an indirect and more subjective complexity, specifically related to the patent application itself, the applicants' strategy, or even particularities of the examination process. In this context, there are some variables suggested in the studies carried out; however, there is no consensus by the authors or in the studies carried out revealing the exact influence of each of them, if any. Initial hypothesis 2 is that there is, even if reduced, indirect influence of the other abovementioned variables.

3. Methodology

The proposed methodology aims at creating a model capable of evaluating the complexity and volume of patent applications, in addition to a new fair and efficient manner of distributing patent applications to patent examiners. For this, it is necessary to obtain the application data with its variables of interest, evaluate patent applications according to the selected variables, create a specific logic for this distribution, and, finally, evaluate the new proposed logic compared to the original distribution. Thus, the proposed method can be divided into four main parts, which complement each other and will be detailed below.

3.1 Evaluation of patent applications as to volume of data and complexity: initial tests

We obtained data from applications that already went through the first examination step during two years, in the area of electricity, more precisely from May 2015 to May 2017, month in which the research was initiated. By identifying and selecting the variables, the proposal was to tabulate data from all patent applications analyzed, including all relevant variables selected, and identify the patent examiner who received the application for analysis. These data were defined as the Initial Test Sample and, based on it, the IPC for the sample patent applications and, consequently, the specific area of expertise (ZAE) of each examiner were then identified.

As this is a problem with multiple variables of interest, we attempted to find a multivariate analysis method to solve it. The bibliographic review was made to choose a method that meets the following criteria:

- did not limit the number of variables used;
- was based on a pair correlation analysis of each variable in order to enable a specific analysis of their relations;
- in case of a high number of variables, allowed for reduction of the size of the problem, i.e., to reduce the number of variables (n) to other components or variables (x), with $x < n$, without significant loss of data or of information about the problem to be resolved; and

- was mathematically and statistically robust and scientifically tested in many different fields of knowledge.

Given the established criteria, the method of principal component analysis (PCA) was selected as the basic tool for evaluation of patent applications. Such method allows for determination of such principal components of the specific problem according to the share of general variance explained by each of the components. Following identification of these new components, a General Complexity Ratio (IGC) is proposed for the patent applications, which is the ratio of the weighted sum of the most significant components plus their eigenvalues to the sum of the eigenvalues themselves, which were obtained from the correlation matrix of the original variables of the problem. Based on these ratios for each of the applications, these were classified into up to five classes (Very Light, Light, Moderate, Heavy, and Very Heavy) according to the average and the standard deviation of the general ratios obtained. It is important to highlight that the eigenvalues and eigenvectors were determined both manually and with the assistance of the software Matlab and of the Matrix Calculator (Available at <https://matrixcalc.org/pt/vectors.html>), and the other steps were executed using Excel electronic spreadsheets (Microsoft Office 2010).

3.2 Evaluation of patent applications as to volume of data and complexity: validation tests with time

After choosing the method, determining the ratios and the classification of the applications into classes, the next step is a sensitivity analysis/validation of the ratios and classification of the patent applications. For this step, an experimental/empirical research is proposed, aiming to establish a correlation between the ratios and classifications found and the time/effort to exam the patent applications. First, the substantive process of patent examination by INPI and the standard examination report were analyzed in order to identify the main examination steps and directly or indirectly related variables. Based on these main examination steps, a form was prepared to survey the time for examination, to be filled up by the examiners, in which information about the time to execute each step is inserted. Thus, a list of applications is determined, hereinafter referred to as Standard Sample, with tabulation of data, including all the variables of interest in addition to the time for examination. The PCA method will then be applied to this new sample, and the IGC ratios for each patent application will be calculated, in addition to their classification into classes. In this context, the correlation between the ratios obtained and the time for examination will be verified, and the PCA method will be applied, including several simulations with variations in the sample size and in the number of variables. This procedure aims at showing the variables with direct impact on the time for examination and the representativeness of the IGC regarding these variables, evaluating the minimum necessary sample size, and also testing the applicability of the PCA method.

3.3 Distribution of patent applications

In this 3rd part of the proposed method, a specific logic for distribution of applications was built based on the previously obtained classification into the five classes, and also on the classification of the application according to the IPC. Data obtained from the patent applications were separated by main IPC subclasses, and so, the main subclasses examined by each examiner were identified, as well as their ZAE was determined. This area was obtained considering the subclasses of patent

applications with occurrences above 5% of the total examinations by each examiner evaluated based on the largest sample obtained, the Initial Test Sample. Thus, a new logic of distribution of patent applications is proposed according to the classification of the general ratios and to the ZAE, considering the following criteria:

- Very heavy and heavy applications shall be equally distributed among the examiners;
- Very heavy and heavy applications shall be compensated, respectively, with distribution of very light and light applications;
- The remaining moderate applications shall be distributed;
- In all previous steps, patent applications shall be distributed considering the ZAE determined for each examiner.

3.4 Evaluation of the new distribution logic

In the fourth and last part of the proposed method, first a new sample of more recent patent applications was obtained, hereinafter referred to as the Final Redistribution Sample. Patent application data were obtained from the same examiners in the field of electricity that made up the Initial Test Sample, however, with the first examinations carried out between May and July 2020. It is important to note that, after the backlog combat plan was implemented by INPI, the examination process has somewhat changed for most of the patent applications in the area of electricity. Hence, obtaining this redistribution sample was necessary to harmonize the examination process carried out by the examiners evaluated therein with the process implemented by the examiner, from which the Standard Sample with time was obtained. This harmonization was made using standard samples with time and redistribution samples containing the same type of patent applications, in other words, patent applications that may be examined using data from previous searches by international offices. It is important to note that this type of application covers an average of 88% of the total stock of patent applications filed until 2016 in the field of electricity.

Based on this new sample, we apply the proposed model and logic for distribution, and then calculate a Distribution Balancing Ratio (IBD) both for the original distribution and for the new distribution, a ratio that ranges from zero to one, and considers the differences between the medians of the variables of each examiner's samples and the general medians of the division's variables. Within this context, it should be noted that the closer the medians of the variables of the examiners' individual samples are to the general medians of the division's variables, the larger the amount of the IDB is and, consequently, the better balanced the distribution is. The breakdown of the IBD, including explanations and analyses about its formula, maximum and minimum limits, will be presented in item 4 – Development.

4. Development

4.1 Principal component analysis (PCA)

The principal component analysis is a multivariate statistics technique focused on explaining the variance–covariance structure of a set of data, and its main objectives are the reduction of the dimensionality of the problem and the better

interpretation of data [26–28]. Still according to [27], the PCA usually reveals relations that would not have been previously identified only with an analysis of the original set of data and variables, enabling a more comprehensive interpretation of the study phenomenon.

To examine a patent application, several measured variables of each patent application of the study population should be considered. The proposal of the principal component analysis method is to apply a transformation to such variables, so that the new components obtained enable a better breakdown and analysis of the elements of such population. In [29], it is shown that this new look has great value when it comes to creating a typology for the population, classifying the elements according to certain criteria, etc. According to Vicini [30] “In practice, the algorithm is based on the variance-covariance matrix, or on the correlation matrix, from which the eigenvalues and eigenvectors are extracted” and “finally, writing the linear combinations, which will be the new variables, referred to as principal components”.

It is important to note that the PCA is widely employed, evidencing its efficiency and robustness in applications in several fields of knowledge, such as agronomy, zootechnics, medicine, among others [31–34]. Examples of practical applications of the PCA for evaluation of public services in Brazilian states, assessment of the regional development of cities in the Brazilian state of Santa Catarina, as well as analysis of crime statistics in U.S. states can be found, respectively, in the papers [35–37].

The following steps are necessary for determining the principal components:

1. create an original data matrix X_{ij} of size $n \times p$ in which the columns ($j = 1, 2, \dots, p$) are the original variables of the problem and the rows ($i = 1, 2, \dots, n$) are individuals in the population (in our case, each patent application);
2. standardize the original variables so all of them have mean equal to zero and standard deviation equal to one, avoiding the influence of different orders of magnitude and obtaining a new data matrix Z_{ij} ;
3. calculate the variance–covariance matrix (S) and the correlation matrix (R) which, in case of standardized variables, will be equal;
4. find the eigenvalues of the matrices and their corresponding eigenvectors;
5. select the components calculating linear combinations of original variables with the eigenvectors of the correlation matrix.

Eqs. (1) and (2) below show, respectively, the calculation for standardization of the variables and matrix Z of standardized variables.

$$Z_{ij} = \frac{X_{ij} - \mu_p}{\delta_p} \quad (1)$$

$$Z = \begin{bmatrix} Z_{11} & Z_{12} & \dots & Z_{1p} \\ Z_{21} & Z_{22} & \dots & Z_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ Z_{n1} & Z_{n2} & \dots & Z_{np} \end{bmatrix} \quad (2)$$

Where: X_{ij} is the element of the original data matrix; μ_p is the average of variable p , being $p = j$; and δ_p is the standard deviation of variable p , being $p = j$.

Eqs. (3)-(5) below show, respectively, the calculation of the variances, covariances, and matrix S.

$$\text{VAR}[Z_j] = \frac{1}{n} \sum_{k=1}^n (Z_{jk} - \mu_j)^2 \quad (3)$$

$$\text{COV}[Z_j; Z_{j'}] = \frac{1}{n} \sum_{k=1}^n [(Z_{jk} - \mu_j)(Z_{j'k} - \mu_{j'})] \quad (4)$$

$$S = \begin{bmatrix} \text{VAR}[Z_1; Z_1] & \text{COV}[Z_1; Z_2] & \dots & \text{COV}[Z_1; Z_p] \\ \text{COV}[Z_2; Z_1] & \text{VAR}[Z_2; Z_2] & \dots & \text{COV}[Z_2; Z_p] \\ \vdots & \vdots & \ddots & \vdots \\ \text{COV}[Z_p; Z_1] & \text{COV}[Z_p; Z_2] & \dots & \text{VAR}[Z_p; Z_p] \end{bmatrix}$$

$$= \begin{bmatrix} 1 & \text{COV}[Z_1; Z_2] & \dots & \text{COV}[Z_1; Z_p] \\ \text{COV}[Z_2; Z_1] & 1 & \dots & \text{COV}[Z_2; Z_p] \\ \vdots & \vdots & \ddots & \vdots \\ \text{COV}[Z_p; Z_1] & \text{COV}[Z_p; Z_2] & \dots & 1 \end{bmatrix} \quad (5)$$

Where: $\text{VAR}[Z_j]$ is the variance of the standardized variable Z_j ; $\text{COV}[Z_j; Z_{j'}]$ is the covariance of the standardized variables Z_j and $Z_{j'}$; n is the number of individuals; μ_j and $\mu_{j'}$ are, respectively, the average of the standardized variables Z_j and $Z_{j'}$; and Z_{jk} are the data matrix elements.

Eqs. (6) and (7) below allow for determination of the eigenvalues and eigenvectors of matrix S and, so, matrix V of eigenvectors of S is determined according to Eq. (8).

$$|S - \lambda.I| = 0 \quad (6)$$

$$S.V = \lambda.V \quad (7)$$

$$V = [V_1 \ V_2 \ \dots \ V_p] = \begin{bmatrix} v_{11} & v_{12} & \dots & v_{1p} \\ v_{21} & v_{22} & \dots & v_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ v_{p1} & v_{p2} & \dots & v_{pp} \end{bmatrix} \quad (8)$$

Where: S is the Variance–Covariance matrix of $p \times p$ dimension of standardized data; λ is one of the p eigenvalues of matrix S; I is the identity matrix of order p ; p is the total number of standardized variables; and V is an eigenvector of S with dimension $p \times 1$;

Upon acquisition of the eigenvectors associated with eigenvalues in a descending order, the principal components (Y_1, Y_2, \dots, Y_p) for each of the n individuals under analysis is determined through a linear combination between the standardized variables and the eigenvalues calculated. Therefore, we can then write the components of individual n in the form of the following equation:

$$Y_{i(n)} = Z_{i1}.v_{1i} + Z_{i2}.v_{2i} + \dots + Z_{ip}.v_{pi} \quad (9)$$

Where: i ranges from 1 to p ; $Y_{i(n)}$ is component i of individual n ; Z_{ip} are the elements of matrix Z of standardized variables; and v_{pi} are the elements of the eigenvectors calculated.

It is important to highlight that in the PCA, the contribution of each principal component (Y_1, Y_2, \dots, Y_p) is measured in terms of its variance. Thus, it is possible to calculate that contribution considering the relation between the variance of the component under analysis and the sum of the variances of all components, resulting in the proportion (or percentage) of total explained variance to each of the components. Eq. (10) below shows how to calculate each contribution C_i (C_1, C_2, \dots, C_p).

$$C_i = \frac{VAR(Y_i)}{\sum_{i=1}^p VAR(Y_i)} \cdot 100 = \frac{\lambda_i}{\sum_{i=1}^p \lambda_i} \cdot 100 \quad (10)$$

Where: C_i is the contribution or total % variance explained by component Y_i ; $VAR(Y_i)$ is the variance of principal component Y_i ; λ_i is one of p eigenvalues of matrix S ; p is the total number of standardized variables.

It should be noted that further details about the PCA formal mathematical statements and its properties, as well as about all linear algebra used, may be consulted throughout the already mentioned works [26–37].

4.2 Selection criteria for principal components

Kaiser's [38] is the most used criterion to date. According to such criterion, only components associated with eigenvalues with ranges wider than the unit are considered principal components, i.e.: $\lambda_i > 1$. However, some practical cases show that only this selection criterion may not properly represent the majority of the total data variance.

A second option is to perform a graphic analysis and verify greater differences among the consecutive eigenvalues. The Cattell [39] criterion, for example, suggests a graphic representation of the range of eigenvalues based on the number of eigenvalues, arranged in an ascending order. The number of components would be selected based on the breaking point of the graph. This breaking point occurs when there is a slump in the range of eigenvalues [40].

A third possible criterion, also quite disseminated, is to use a reference value for the proportion of variance explained by the principal components. Following this logic, the principal components whose cumulative percentage of explained variance exceeds such reference value shall be selected. It is important to highlight that there is no consensus among researchers about which percentage should be used, and there are several practical examples. A great part of the applications uses the limit of 70%. In [40], the problem was ranked in levels of acceptance, and amounts between 62 and 80% were considered reasonable or “partially good”.

Although each criterion has advantages and disadvantages, in this paper a combination of the three above-described criteria was adopted. As a reference value for the third criterion, we believe a percentage of explained variance starting at 60% is a suitable amount for selecting the most representative principal components.

4.3 General complexity ratio (IGC)

Hongyu et al. [31] also states that “In order to establish a ratio that enables us to order a set of n objects, according to a criterion defined by a set of m suitable variables, it is necessary to choose the weights of the variables so that they translate information contained in them,” provided that, to create a ratio as a linear combination of variables, “it is desirable that this ratio includes the maximum possible

information of the set of variables selected for study”. According to Sandanielo [41] (*apud* HONGYU, 2015), “a method that creates linear combinations with maximum variance is the principal component analysis”.

In this context, this paper intends to, in a first evaluation, use a ratio in one dimension based on the most significant principal components (carefully selected using the selection criteria addressed in item 4.2), weighted by their corresponding eigenvalues. Hence, a General Complexity Ratio (IGC) is defined for the patent applications to be evaluated, according to the following equation:

$$IGC_n = \frac{\sum_{i=1}^k Y_i \lambda_i}{\sum_{i=1}^k \lambda_i} \tag{11}$$

Where: Y_i are the principal components calculated, λ_i are the eigenvalues calculated, k is the number of principal components selected and n is the patent application evaluated.

4.4 Classification of the patent applications into classes

To group data according to the IGC ratio calculated, the first step was to standardize the original values using Eq. (1) and, therefore, the standard deviation of the IGC sample will always be equal to one. Based on IGC data for all patent applications under examination, classification ranges were then defined as shown in Figure 1.

4.5 Mathematical model: diagram and evaluation

After calculating the IGC ratio and classifying the applications, the complete model for evaluating patent applications is created. A diagram of the model is presented in Figure 2.

Classes	Ranges
Very Heavy	IGC >3
Heavy	1 < IGC <= 3
Moderate	-1 <= IGC <= 1
Light	-3 <= IGC < -1
Very light	IGC < -3

Figure 1.
Classification of the patent applications into classes.

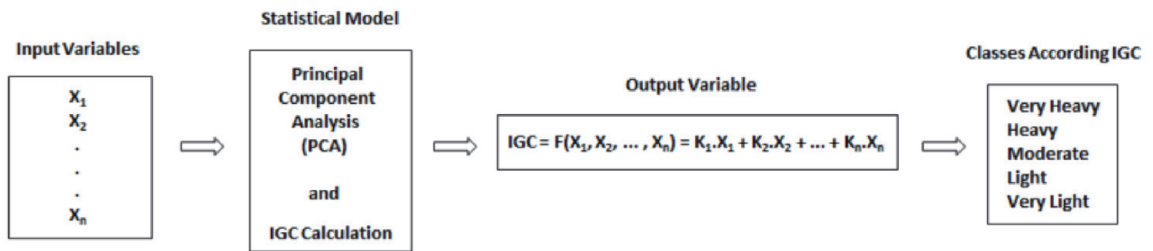


Figure 2.
Diagram of the model.

Thus, the next step is to enable its evaluation through a sensitivity analysis and through the correlations with time for examination/analysis of the application. This evaluation will happen through preparation of a form regarding the time for examination in order to obtain a new Standard Sample of patent applications with time, in addition to several simulations considering different numbers of variables and sample sizes. **Figure 3** shows the template form developed.

4.6 Iterative logic for individual distribution of applications

After classifying all applications to be distributed during the selected period, an iterative sequence of steps for their distribution is then proposed, seeking to prioritize the choice of applications based, whenever possible, on the IGC classification of each application being distributed, in addition to the ZAE of each examiner.

- 1. Verify the Main IPC Subclass in which the Current Application falls:
 - i. If the Current Application pertains to the Current Examiner’s ZAE, then select the Current Application and proceed to Step 2;
 - ii. If there is no application pertaining to the Current Examiner’s ZAE, look for the next application in line that does not pertain to any other Examiner’s ZAE and, only after that, proceed to Step 2.
- 2. Check the Current Application Classification (Very Light, Light, Moderate, Heavy, or Very Heavy):
 - i. If an Application with the same Classification has not yet been distributed to the Current Examiner, then distribute the Current Application and proceed to Step 3;
 - ii. Otherwise, search for the next available Application and return to Step 1.
- 3. Repeat Steps 1 and 2 until an Application with each Classification is distributed to the Current Examiner.
- 4. Go to the next Examiner on the list and perform Steps 1 to 3.
- 5. Repeat Steps 1 to 4 until there are enough Applications in each of the Classifications:

Form - information about the time to execute each step of substantive patent exam												
Step 1: Initial Technical Analysis						Step 2: Prior Art Search			Step 3: Specific Technical Analysis			
Reading and understanding of the application						Search execution			Analysis of the Patentability			
Application number	Type	Description (minutes)	Claims (minutes)	Figures Analysis (minutes)	Total Step 1 (hours)	Formulation of the Search Strategy (minutes)	Execution of the Search Strategy (minutes)	Total Step 2 (hours)	Claims technical Analysis (minutes)	Prior Art Technical Analysis (minutes)	Comparison and Technical Justifications (minutes)	Total Step 3 (hours)
1					0			0				0
2					0			0				0
3					0			0				0
4					0			0				0
5					0			0				0
6					0			0				0
7					0			0				0
8					0			0				0
9					0			0				0
10					0			0				0

Figure 3.
Time for examination form.

- i. If there are no more applications with any of the Classifications, repeat Steps 1 through 5 for the remaining Classifications until there are no more Applications available for distribution.

4.7 Final redistribution sample and calculation of the distribution balancing ratio (IBD)

In order to evaluate the distribution of the patent applications, a new data sample was obtained, referred to as Final Redistribution Sample. As the Standard Sample (with time) used to validate the model was obtained using examination data that, on their turn, were also based on international searches, this new redistribution sample was necessary to harmonize the examination process carried out by the examiners evaluated therein with the process implemented by the examiner, from which the Standard Sample with time was obtained. Based on this new sample, the proposed model and distribution logic are applied, and a Distribution Balancing Ratio (IBD) is then calculated, both for the original distribution and the new one, according to Eq. (12):

$$IBD = \frac{\sqrt{\sum_i \sum_j \left(\frac{m_{edij}}{M_{edj}} \right)^2}}{\sqrt{\sum_i \sum_j \left(\frac{m_{edij}}{M_{edj}} \right)^2 + \sum_i \sum_j \left(\frac{m_{edij} - M_{edj}}{M_{edj}} \right)^2}} \quad (12)$$

Where: m_{edij} is the median of variable j of patent applications distributed to examiner i ; and M_{edj} is the median of variable j for all applications in the final redistribution sample.

In a first analysis of the IBD equation, it can be verified that the ratio seeks to capture and measure the influence of the differences between the medians of the variables of each examiner's samples and the general medians of the division's variables (complete sample). It is important to note that all medians of the variables composing the IBD are normalized (divided) by the general values of the respective medians of the complete sample. With that, we seek to avoid further distortions caused by different orders of magnitude of certain variables.

Additionally, as the several medians calculated can be higher or lower than the respective median of the division, the differences in these values can be positive or negative. Hence, as we wish to obtain an accumulated measurement of all differences in medians with no loss of information and without having a negative deviation in a certain variable compensating a positive deviation in another, we choose to square the differences and then add and extract the square root.

More specifically, when the numerator and the denominator of the IBD equation are analyzed, it can be noticed that both have the same first term, which is the sum of the squares of the normalized medians of all variables of interest from the examiners' samples. However, the denominator has a second additional term presenting the sum of the squares of the normalized differences between the medians of each examiner's individual samples and the medians of the corresponding variables from the complete sample.

It is important to highlight that, in an ideal distribution, the medians of the variables from all examiners' samples would be equal to the medians of the general variables of the division, i.e., the sum of the squares of the differences of the medians (second term of the IBD denominator) would be zero and, consequently, the IBD would be equal 1. On the other hand, a random distribution in which the examiners' samples have great differences in median values, when compared to the

general division medians, would lead to very high denominator values, thus greatly reducing the IBD and, ultimately, making it tend to zero. Thereafter, we have that, the closer the medians of the variables from the examiners' individual samples are to the general medians of the division's variables, the greater and closer to 1 the IBD value will be and, consequently, better balanced the distribution will be.

5. Results

5.1 Determination of the examiners' ZAE and initial tests

Data were obtained from a total of eleven (11) Examiners from the Electricity Division, to be fully analyzed. For each patent application of such examiners with at least one examination already carried out, all variables of interest were obtained, totaling eight hundred and fourteen (814) patent applications to be evaluated and making up the initial test sample. Data from the Initial Test Sample were standardized according to Eq. (1). **Figure 4** shows the structure of data from the initial test sample with standardized data.

In the initial test sample, a total number of 95 main subclasses was found over all 814 patent applications analyzed. However, when considering only the examiners' ZAE (subclasses including 5% or more applications for each examiner), 25 main subclasses were responsible for 636 applications, i.e., about 80% of the total evaluated. It is important to highlight that, given that 3 of such 25 subclasses had very low occurrences, 22 subclasses were used in the set of interest for evaluation, equivalent to 619 applications (76% of total). **Figure 5** shows each one of the 11 examiners' areas of expertise by IPC subclasses. Note that the gray area corresponds to the examiners' Specific Areas of Expertise (ZAE), while the white area corresponds to

Order	Examiner	Application	Main Subclass	Pages of Description	Pages of Claims	Pages of Figures	Pages of Third Party Information	Independent Claims	Dependent Claims	Subclasses	Year of Filing	Inventors	Priorities
1	1	PI0101329-7	H01C	24	3	5	0	1	13	1	1	7	1
2	1	PI0111250-3	H01H	7	3	2	0	1	7	1	1	2	1
3	1	PI0400832-4	H02K	12	2	2	0	3	3	1	4	1	0
4	1	PI0403974-2	H01B	40	11	21	0	9	44	2	4	5	0
5	1	PI0107166-1	H01H	16	2	1	0	2	10	1	1	4	1
6	1	PI0303675-8	H02G	7	2	4	0	1	12	1	3	3	1
.
.
.
814	11	PI0601239-6	H05B	18	2	10	0	4	3	1	6	2	0

Figure 4.
Structure of the initial test sample with standardized data.

Examiners' Areas of Expertise by IPC Subclasses													
1	2	3	4	5	6	7	8	9	10	11			
Subclass	Subclass	Subclass	Subclass	Subclass	Subclass	Subclass	Subclass	Subclass	Subclass	Subclass	Subclass	Subclass	Subclass
H01H 63,08%	H01R 33,33%	H05B 18,42%	H02K 32,65%	H02J 19,54%	H02P 12,26%	H02K 56,36%	B66B 31,58%	G07F 17,09%	H01R 52,17%	B66B 12,09%			
H02G 15,38%	H01H 23,33%	G01R 18,42%	H01M 14,29%	H02M 16,09%	B60L 9,43%	H01B 9,09%	B60T 14,04%	G08B 11,11%	H02G 24,64%	G01R 10,99%			
H01B 6,15%	H01B 7,78%	H02P 15,79%	B60L 6,12%	G01R 13,79%	H02J 6,60%	H01R 7,27%	B60K 5,26%	B66B 7,69%	H02B 4,35%	H02B 8,79%			
H01F 3,08%	G01R 6,67%	H01F 10,53%	G01R 4,08%	F02D 10,34%	F02D 6,60%	H01F 5,45%	F02D 5,26%	H05B 5,13%	H01H 2,90%	H01H 6,59%			
H02K 1,54%	H05B 4,44%	H02H 7,89%	H02J 2,04%	B60K 4,60%	H01R 5,66%	F21V 3,64%	H02P 3,51%	F21V 5,13%	H02H 1,45%	H05B 4,40%			
H02P 1,54%	H02H 3,33%	H01M 7,89%	F21V 2,04%	H02H 4,60%	B60K 3,77%	G01R 1,82%	H02G 1,75%	G01R 3,42%	H02K 1,45%	F21V 3,30%			
H02B 1,54%	H02J 3,33%	H01H 5,26%	H02H 2,04%	H01M 4,60%	H02K 3,77%	H02G 1,82%	H02H 1,75%	H02H 2,56%		B60L 2,20%			
	H02K 2,22%	H02G 2,63%	H02P 2,04%	H02G 1,15%	H01F 3,77%	H02B 1,82%		H01H 2,56%		H01R 1,10%			
	H02G 2,22%	B60K 2,63%	H01H 2,04%	H01F 1,15%	H01M 3,77%	H01M 1,82%		H02J 1,71%		H01M 1,10%			
		H02K 2,63%	H02B 2,04%	B60L 1,15%	H05B 3,77%			H01M 1,71%		H02J 1,10%			
			H02G 2,04%	H01H 1,15%	H02H 2,83%			H02K 0,85%		H02H 1,10%			
				H02P 1,15%	G01R 1,89%			H01F 0,85%		B60T 1,10%			
					F21V 0,94%			B60L 0,85%		H02P 1,10%			
					H01H 0,94%								

Figure 5.
Examiners' areas of expertise by IPC subclass.

subclasses that, despite not being part of the examiner’s ZAE, are part of the ZAE of some of the other examiners evaluated.

Figures 6 and **7** show the results obtained in the PCA.

By analyzing **Figure 6**, when using only the criterion considering eigenvalues higher than one, only the first four (4) components would be selected. However, these would be responsible for about 65% of total variance. It can also be verified that the range sharply drops when we get to the eigenvalue of component 6 (0.73). Additionally, the first five components explain a total variance of 75.05%. Hence, these first five components were selected for the next steps.

In **Figure 7**, the significant factors to each variable (very close to or above 0.4) were hatched in gray. In a first analysis, it can be noticed that component Y1 is quite detached from the others. In addition to explaining virtually 30% of all variance, the component is associated with variables directly related to the volume of data (pages of description, of claims, and of figures, in addition to the number of independent and dependent claims). Such fact is consistent with the initial hypothesis 1, related to the volume variables. As for components Y2 and Y3, although they may be slightly related to the volume of data, they basically represent the influences of variables, year of filing, number of inventors, and priorities. These components appear to be associated with development strategies, management of the applicants, and maturity of the technology involved. On the other hand, components Y4 and Y5

Principal Component	Eigenvalues	Variance	Accumulated variance
Y1	2,95	29,53%	29,53%
Y2	1,48	14,82%	44,36%
Y3	1,12	11,18%	55,53%
Y4	1,03	10,29%	65,82%
Y5	0,92	9,23%	75,05%
Y6	0,73	7,33%	82,39%
Y7	0,71	7,12%	89,51%
Y8	0,53	5,25%	94,77%
Y9	0,34	3,43%	98,19%
Y10	0,18	1,81%	100,00%

Figure 6.
Eigenvalues and variances.

Variable	Weighting Coefficients (Scores)				
	Y1	Y2	Y3	Y4	Y5
Pages of Description	0,4143	-0,316	-0,1136	-0,0525	0,0835
Pages of Claims	0,4978	0,1897	0,2442	-0,0058	-0,0719
Pages of Figures	0,3317	-0,487	-0,2236	-0,1627	-0,0517
Pages of Third Party Info	-0,0594	-0,0591	0,3375	-0,6677	0,6526
Independent Claims	0,3932	0,2189	0,2599	0,0367	-0,0534
Dependent Claims	0,4574	0,25	0,2467	-0,0404	-0,1294
Subclassess	0,0322	-0,201	0,3019	0,7121	0,5457
Year of Filling	-0,0876	-0,456	0,5003	0,0443	-0,2269
Inventors	0,2168	-0,4437	-0,2729	-0,0036	0,0224
Priorities	0,218	0,2606	-0,4682	0,1138	0,4359

Figure 7.
Weighting coefficients.

complement the others by being associated with the variables, number of sub-classes, and pages of third party observations. Such components appear to represent specific influences of the technological area of the patent applications. This result is consistent with the initial hypothesis 2, related to the variables with complementary or indirect influences.

By applying the proposed redistribution logic, a new configuration of samples by examiner was obtained. **Figure 8** shows a comparison between the percentage of applications distributed to each examiner within their ZAE.

By analyzing **Figure 8**, it is possible to note that, for ten out of the eleven examiners, there was a significant increase in the number of applications distributed to them and pertaining to their own ZAE. Only examiner 9 had a small decrease (that could even be corrected with a fine adjustment), due to the fact that his data sample was significantly larger than that of the others. Thus, this new configuration seems to contribute to examiners to work within their specific fields of expertise and knowledge.

Finally, the IBD ratios of the original sample distribution and its redistribution were calculated. Through Eq. (12), an IBD equal to 0.86 for the original case and an IBD equal to 0.88 for the redistribution were obtained, i.e., there was an increase in the IBD with the new distribution, evidencing that the medians of the examiners' applications after redistribution are closer to the general division's median. Such fact corroborates the fact that with the new distribution, we have a tendency towards greater balance regarding volume of data/complexity of applications distributed to the examiners.

5.2 Model validation: simulations using standard sample with time

Similarly to the procedure carried out for the initial test sample, data of patent applications of the technological area regarding electrical engineering were obtained. However, as, in this case, we intend to obtain a standard sample with time to serve as a reference, all data were obtained from the time for examination form filled by a single examiner. For each patent application of such examiner, all variables of interest were obtained, for a total amount of fifty (50) patent applications, with all first actions already published. Data were collected between January and July 2020, with all sample applications using data from previous searches by international offices.

We note that from the ten possible variables to be analyzed, the only one that was not considered in this case was the number of pages of third parties observations, given that there is no application with such document available in the sample.

For the sensitivity analysis of the model, dozens of simulations were performed considering all cases: from the most complete one, with nine variables, to the simplest ones, with three variables. For all cases, simulations for each ten sample applications were performed, i.e., for each set of cases with three to nine variables, tests were performed considering 10, 20, 30, 40, and 50 patent applications of the standard sample. A minimum of ten test applications was chosen, as it is recommended that the sample should have a population at least larger than the

Examiner	1	2	3	4	5	6	7	8	9	10	11
New ZAE	90,63%	76,14%	97,37%	87,50%	85,06%	46,73%	83,64%	81,03%	41,88%	91,30%	76,92%
Original ZAE	84,62%	71,11%	84,21%	53,06%	59,77%	40,57%	78,18%	56,14%	46,15%	76,81%	38,46%

Figure 8.
Percentage of applications within the ZAE before and after the new distribution.

number of variables in order to apply the PCA method, provided that the larger the sample, in theory, the best for the model.

When executing the simulations, the correlations of all variables and of the IGC ratios with time were verified, and the IGC was calculated both using the criterion of 70% of the variance, being referred to as $IGC_{70\%}$, and using the criterion of eigenvalues higher than one, being referred to as $IGC_{\lambda>1}$. It is important to note that, when $IGC_{70\%}$ and $IGC_{\lambda>1}$ are equal, we will refer to it simply as IGC. Finally, an IGC related only to the principal component (Y1) of the cases, the most significant component in terms of variance, being referred to as IGC_{Y1} , was also calculated.

After executing all the simulations, and having a gamut of results for dozens of cases, the cases of more relevance and interest in terms of analyzed variables and their correlations with time were selected. Namely:

- Case 1– Case 9 Var, complete with the nine variables;
- Case 2 – Case 5 Var, including only the five variables of direct volume;
- Case 3 – Case 4 Var, similar to case 2, but excluding the variable “number of pages of figures” (given that this variable presented lower scores in the PCA and the lowest individual correlation with time among the five of volume);
- Case 4 – Case 3 Var, similar to case 3, but aggregating the amount of dependent and independent claims in only one variable (i.e., considering pages of description, claim pages, and total claims); and
- Case 5 – Case 3 Var (2), similar to case 4, but replacing the pages of description with the total of pages. Hence, the following variables were considered: total number of pages, claim pages and total claims.

Figures 9 and 10 show the results of eigenvalues and cumulative variances for all the five described Cases.

By analyzing **Figures 9 and 10**, it is possible to note that:

- Case 9 Var: simulations using 10 and 20 applications deviate from the others, and the cases using 30 to 50 applications are almost coincident, i.e., the

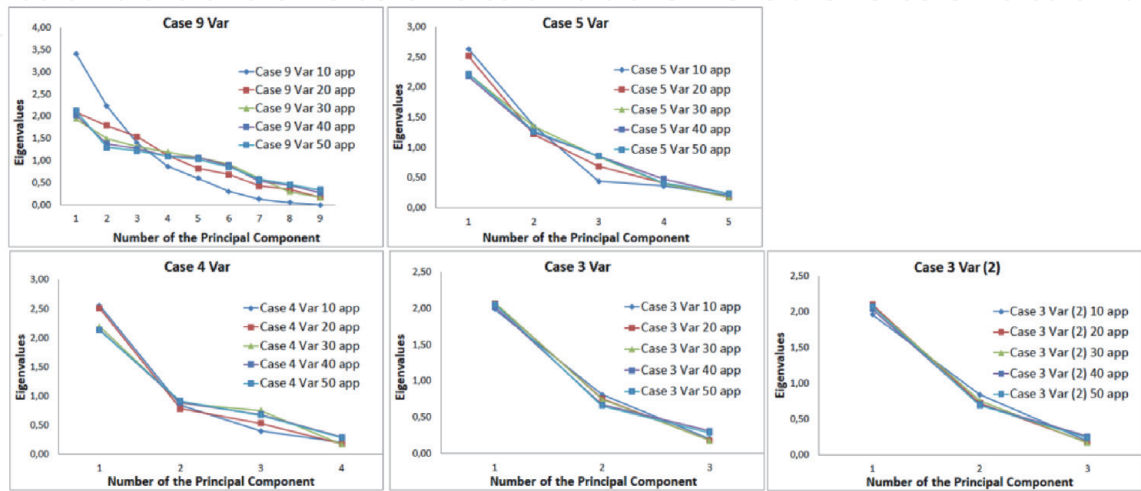


Figure 9.
Eigenvalues of cases 1 to 5.

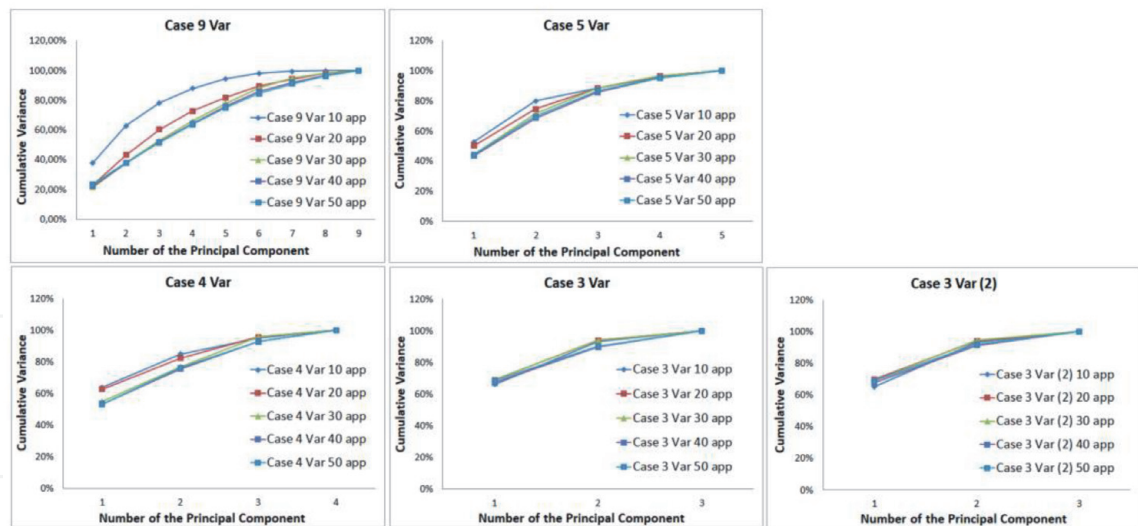


Figure 10.
Cumulative variances of cases 1 to 5.

eigenvalues of the components only tend to be stable starting in the sample with 30 applications. The same phenomenon can be observed by analyzing the cumulative variances of the samples. These results evidence that, for cases with nine variables that are intended to be executed, a sample of at least 30 patent applications, preferably 50 applications, is recommended to obtain better performance of the PCA method;

- Case 5 Var: only the simulation using a sample of 10 applications quite deviates from the others; the case using 20 applications shows a slight deviation; and cases using 30 to 50 applications are almost coincident, i.e., the eigenvalues of the components already tend to be stable starting in the sample with 20 applications. The same phenomenon can be observed by analyzing the cumulative variances of the samples. These results evidence that, for cases with five variables that are intended to be executed, a sample of at least 20 patent applications, preferably 30 applications, is recommended to obtain better performance of the PCA method;
- Case 4 Var: simulations using 10 and 20 applications slightly deviate from the others, and the cases using 30 to 50 applications are coincident, i.e., the eigenvalues of the components already tend to be quite stable starting in the sample with 10 applications, and very stable starting in 30 applications. The same phenomenon can be observed by analyzing the cumulative variances of the samples. These results evidence that, for cases with four variables that are intended to be executed, a sample of at least 10 patent applications, preferably 20 applications, is recommended to obtain better performance of the PCA method.
- Case 3 Var and Case 3 Var(2): all simulations, from samples of 10 to 50 applications, are virtually coincident, i.e., the eigenvalues of the components have excellent stability. The same phenomenon can be observed by analyzing the cumulative variances of the samples. These results evidence that, for cases with three variables that are intended to be executed, a sample of at least 10 patent applications, preferably 20 applications, is recommended to obtain better performance of the PCA method.

Figures 11 and 12 show the results of correlations of the IGC with time.

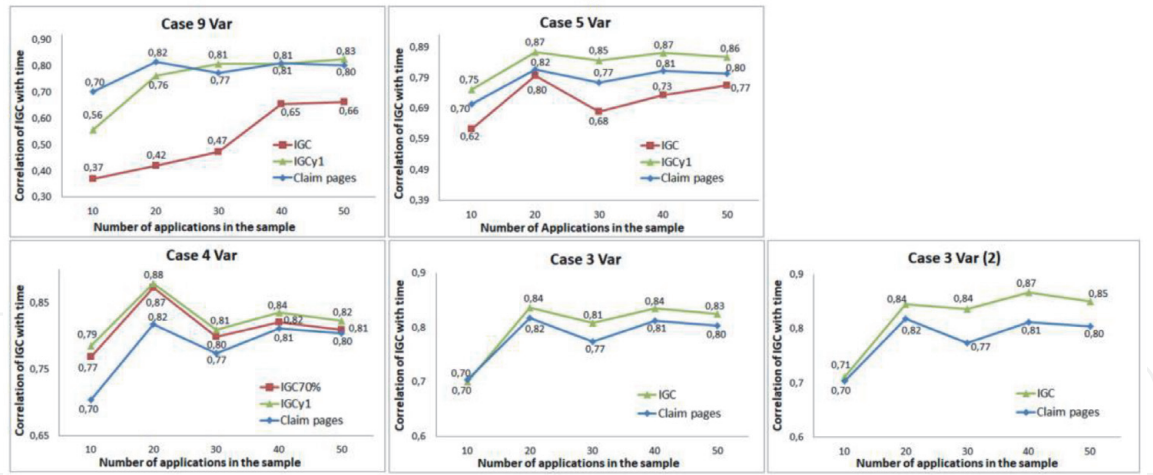


Figure 11.
Correlation of the IGC with the time for examination – cases 1 to 5.

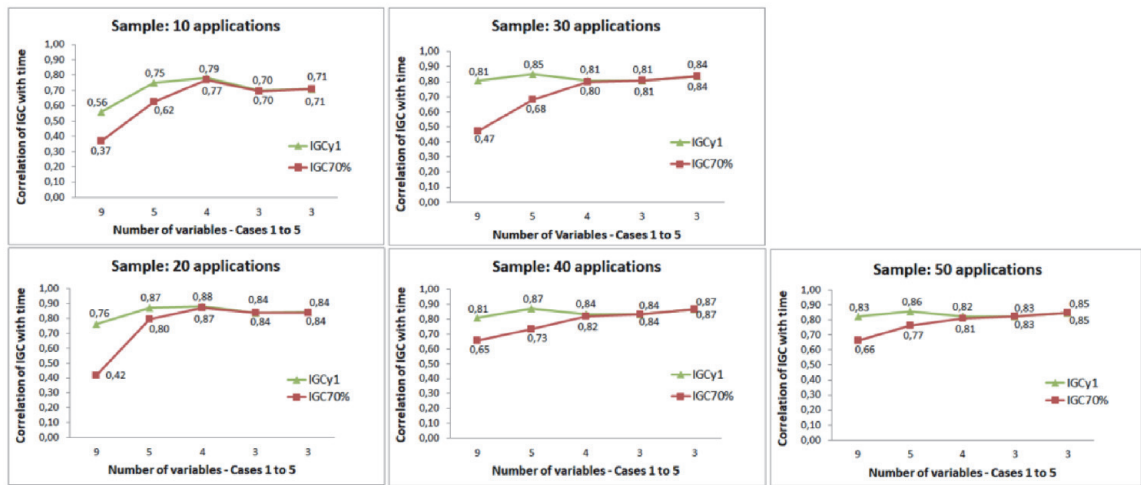


Figure 12.
Correlation of the IGC with the time for examination – cases 1 to 5 – By samples.

The analysis of **Figures 11** and **12** indicates that:

- Case 9 Var: based on a sample with a total of 30 patent applications, there is a slight advantage of correlation with time for IGC_{Y1} , and both are very close to the value of 0.8;
- Case 5 Var: except for the sample with 20 applications, the larger the sample, the greater the correlation of the IGC with time. On the other hand, upon analysis of the IGC_{Y1} , that is, considering only the most significant principal component, basically related to the main direct volume variables, the correlation with time increases significantly, reaching even larger values (0.86). It is also worth to highlight that the correlation of the IGC_{Y1} with time increases a lot when the sample have 20 applications instead of 10; from then on, it seems to stabilize, oscillating around 0.85. This result proves to be consistent with the profile of the eigenvalues and cumulative variances analyzed, reinforcing the need for a sample with a least 20 applications. Finally, a very similar profile is noted in the three curves. Once again, the IGC_{Y1} has better results upon analysis of the correlation with time;
- Case 4 Var: the profile of the three curves is quite similar, and, based on the sample with 30 applications, the correlation with time of all ratios gets close to

0.80 and oscillates around that. It is also worth to highlight that both IGC curves show a good correlation with time even based on a sample with only 10 applications, a result that proves to be consistent with the profile of the eigenvalues and cumulative variances analyzed. Finally, it can be noticed that, although the correlation values of the three curves are close to each other, once again, the IGC_{Y1} (in this case also equal to the $IGC_{\lambda>1}$) has an advantage over the others;

- Case 3 Var: the profile of both curves is quite similar, and, based on the sample with 20 applications, the correlation with time of all ratios gets close to 0.80 and oscillates around that. It is also worth to highlight that both curves already show a reasonable correlation with time even based on a sample with only 10 applications, a result that proves to be consistent with the profile of the eigenvalues and cumulative variances analyzed. Finally, it can be noticed that, although the correlation values of both curves are close to each other, once again, the IGC (in this case $IGC = IGC_{70\%} = IGC_{\lambda>1} = IGC_{Y1}$) has an advantage.
- Case 3 Var (2): based on the sample with 20 applications, the correlation of the IGC with time is always above 0.84. It is also worth to highlight that both curves already show a reasonable correlation with time even based on a sample with only 10 applications, a result that proves to be consistent with the profile of the eigenvalues and cumulative variances. Finally, it can be noticed that this is the case in which the IGC (once again, $IGC = IGC_{70\%} = IGC_{\lambda>1} = IGC_{Y1}$) has the higher values of correlation with time, so it has an advantage regarding all variables individually. In addition to having a high correlation with time, this is the most indicated case for practical application, because, besides having only three variables and principal components with great stability, it does not require a division of the claims into independent and dependent, which greatly facilitates the data collection.

The analysis of **Figures 11** and **12** also indicates that the profile of the curves is quite similar, with an increasing trend for the correlation of the IGC with time in the beginning of all of them, i.e., when the number of variables decreases from nine to five, and, from five to three variables, the curves stabilize. These results reflect the previous analyses that showed that, when only the direct volume variables are selected (with their combinations varying), the trend was for obtaining higher and more stable correlations with time. Thus, although all nine variables of the study may contribute to the complexity of the patent application, in practice, the direct volume variables already represent well the examination effort/time.

It should also be noted that the correlations of the IGC_{Y1} with time remained high and almost constant for any of the cases analyzed with samples with twenty applications or more, showing a quite stable behavior regardless of the sample size. Consequently, for the problem in particular, the results converge showing that the IGC_{Y1} ratio seems more suitable to represent the examination effort/time. The results obtained suggest that Case 3 Var (2) is the one with the best cost–benefit relation for the performance of even more specific practical tests, whether because it captures the influences of the main variables of direct volume data, because it is simpler regarding obtaining and collecting data (as it does not require a division of the claims into independent and dependent), or because it has higher correlations of the IGC with time.

Figure 13 shows the classifications of the Sample applications by time and by the IGC_{Y1} . **Table 1** shows the applications in which there was divergence in the classification.

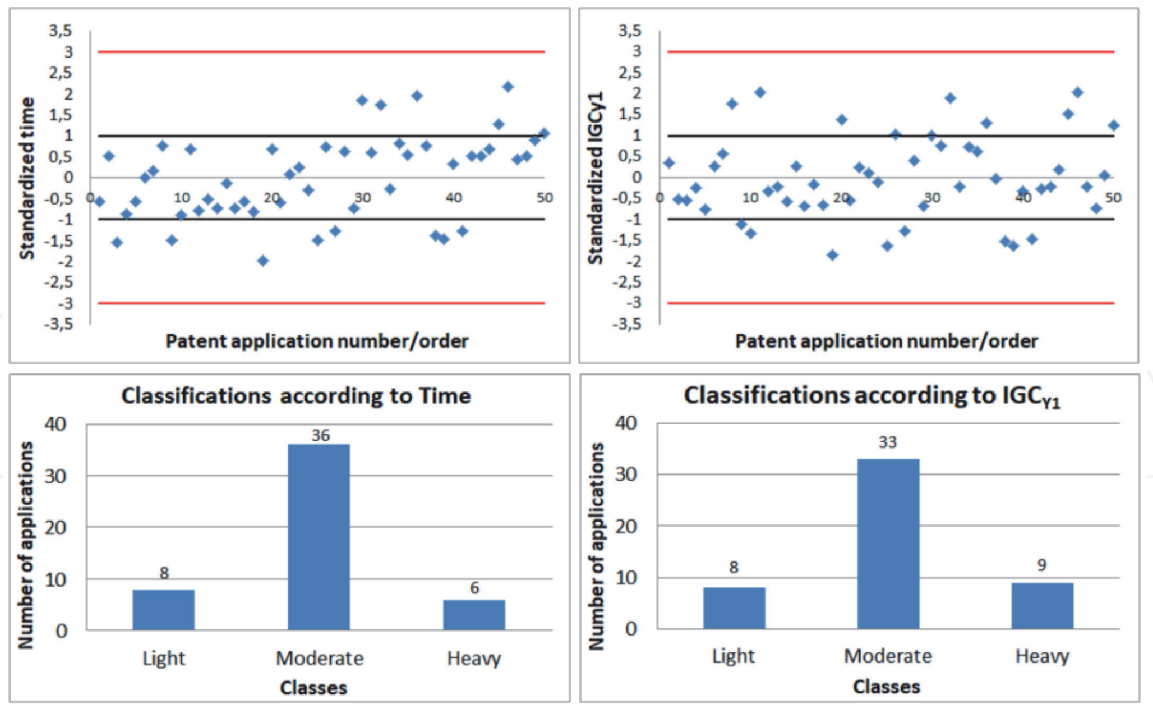


Figure 13.
Classification of applications of the standard sample (with time).

Comparison of Classifications according to Time and IGC _{Y1}					
Application	Claim Pages	Total Pages	Total Claims	Classification according to the Time	Classification according to IGC _{Y1}
3	2	34	5	Light	Moderate
8	5	50	15	Moderate	Heavy
10	2	13	3	Moderate	Light
11	6	37	19	Moderate	Heavy
20	5	34	16	Moderate	Heavy
26	4	36	15	Moderate	Heavy
30	4	56	8	Heavy	Moderate

Table 1.
Applications with divergent classifications.

By analyzing **Figure 13**, it can be verified that, when classified by time for examination, none of the applications from the sample was considered to be neither very light nor very heavy. Most applications were classified as moderate (36 or 72%), then light (8 or 16%), followed by heavy (6 or 12%). This result proves to be consistent with data obtained, as the standard sample is fairly homogeneous, shows applications of the same type of examination (using data from previous searches), and the time variable presented a moderate coefficient of variation (16.58%). Similarly to the classification by time, when classified by the IGC_{Y1}, none of the patent applications from the sample was considered to be neither very light nor very heavy. Most applications were classified as moderate (33 or 66%), then heavy (9 or 18%), followed by light (8 or 16%).

By analyzing **Table 1**, it can be verified that there was a total of seven patent applications with conflicting classifications by time and IGC_{Y1}. Therefore, this result shows that the classification of 43 of the 50 applications (86% of the total)

converged, in other words, it shows great similarity. It is important to note that the correlation of the IGC_{Y1} with time for the case under analysis was 0.85, i.e., the classification criterion proved to be efficient, managing to keep up with the capture tendency of this relation of the ratio with the time. More specifically to the differences found, there were four new classifications according to the IGC_{Y1} as heavy (applications 8, 11, 20, and 26). Such applications have higher than average number of claim pages, total number of pages, and total number of claims, showing a profile similar to the other five heavy applications with similar classifications and, so, its classification as heavy according to the IGC_{Y1} is warranted. On the other hand, it can be noticed that applications 20 and 26 showed IGC values close to one, i.e., to the classification limit between the moderate and heavy ranges. Consequently, there are two factors that may possibly explain this phenomenon: i) errors inherent in the mathematical model, which, although in small amounts, tend to occur depending on the variables, samples, and criteria adopted; and ii) measurement errors or deviations in time, which could move a classification close to the limit of the ranges.

Regarding application 10, classified as light according to the IGC_{Y1} and as moderate according to time, it is possible to note it is indeed a quite short application that, at first, would actually tend to be classified as light. Specifically in this case, the standardized time was very near to minus one (standardized time = -0.91), i.e., quite near to the limit between moderate and light classes. Unlike the conflicting heavy cases (in which deviations probably occurred for reasons inherent in the model), in this case the tendency is that time measurement deviations may have caused the discrepancy.

In the case of application 3, classified as moderate according to the IGC_{Y1} and as light according to time, we notice that it is an application with few claims and few claim pages, but with a quite high count of total pages. Depending on the specific examination procedure and the need to better understand the description and the figures, time may lead to a moderate or light classification. Hence, it is a type of application difficult to classify *a priori* and, in this case, the model was more conservative, classifying it as moderate.

Finally, application number thirty, classified as moderate according to the IGC_{Y1} and heavy according to time, showed a IGC_{Y1} virtually equal to one ($IGC_{Y1} = 0.995$), reaching the exact limit between the moderate and heavy classification ranges. It ends up being a case similar to the discrepancies of the heavy applications, due to issues inherent in the mathematical model.

In short, it can be verified that the model manages to represent quite satisfactorily the examination time/effort, and for cases in the threshold of the criteria adopted, few discrepancies occur, and, in these cases, the discrepancies occur only in the adjacent ranges. In other words, any discrepancies that occur are occasional, not rough, and reasonable given the limitations inherent in this kind of model and research.

5.3 Analysis of the redistribution logic: simulations with the final redistribution sample

Ten patent applications of ten examiners under analysis were selected to compose the final redistribution sample, amounting to one hundred (100) patent applications to be examined. The steps of the proposed methodology were strictly followed, but, due to the fact that the purpose of this case was to obtain a sample to apply the model validated with the standard sample with time (our reference), all variables of interest were obtained of first examinations already published, collected between May and July 2020, and all sample applications are also using data from previous searches by international offices (in the context of the “backlog

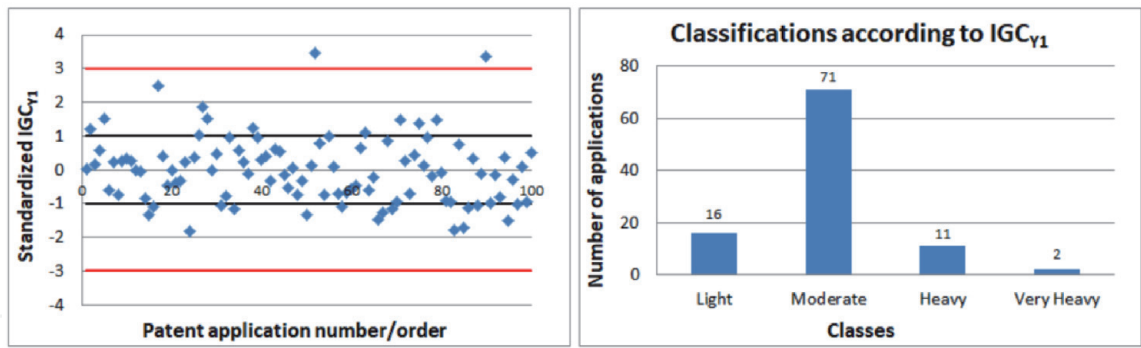


Figure 14.
classifications of the Redistribution sample applications.

Examiner	1	2	3	4	5	6	7	8	9	10
New ZAE	70,00%	50,00%	40,00%	60,00%	40,00%	50,00%	30,00%	10,00%	90,00%	40,00%
Original ZAE	80,00%	50,00%	40,00%	40,00%	10,00%	0,00%	20,00%	10,00%	30,00%	0,00%

Figure 15.
Percentage of applications within the ZAE before and after the new distribution.

combat plan”, i.e., without executing specific prior art search). The case chosen for the redistribution simulation was Case 5 – Case 3 Var (2), given that this case obtained the best results in the validation tests with the standard sample with time. **Figure 14** shows the classifications of the redistribution sample applications.

None of the sample patent applications were classified as very light. Most applications were classified as moderate (71%), followed by light (16%), heavy (11%), and, finally, very heavy, which were only two (2%). It should be noted that the data on the IGC_{Y1} showed normal statistical distribution, similarly to the time and the IGC_{Y1} of the standard sample.

Figure 15 indicates that with the redistribution there was a better balance in the concentration of applications within the ZAE of each examiner. We note that in the case of six out of the ten examiners, there was an increased number of applications distributed to them and pertaining to their own ZAE, with an emphasis on examiners 5, 6, 9 and 10, with significant increases. Only examiner 8 remained with a poor concentration (10%) of applications within his ZAE, which may be explained by the fact that this examiner is a more “versatile” examiner of the division, and does not have such a well defined ZAE. Thus, the results suggest that this new configuration contributes for the examiners to work within their specific fields of expertise and knowledge.

To complement the cycle of the methodology and make a last comparison between the distributions, the IBD ratios of the original sample distribution and of its redistribution were calculated. Eq. (12) resulted in an IBD equal to 0.83 for the original case and an IBD equal to 0.9 for the redistribution, i.e., there was an increased IBD with the new distribution, showing that the medians of the applications of the examiners after redistribution get closer to the general median of the division. This corroborates the fact that the new distribution results in a trend for better balance regarding the volume of data and time/effort of the applications distributed to the examiners.

6. Final considerations

In this study, ten possible variables were identified, relevant to the evaluation and distribution of the patent applications to the examiners. Among these variables,

the ones directly related to the voluminosity of a patent document, i.e., the volume of data that the examiner has to deal with when examining patents, were identified, namely: the number of pages of description, the number of claim pages, the number of pages of figures, the number of independent claims and the number of dependent claims.

With the application of the PCA in a first data sample, referred to as Initial Test Sample, it was verified that the components were consistent with the initial hypotheses. Based on this initial sample, containing a large number of applications examined over two years, the examiners' Specific Areas of Expertise (ZAE) were determined, that is, the IPC subclasses (technological areas) they examine the most according to their knowledge and work experience. These ZAE are highly relevant, as these subclasses are one of the criteria used to distribute patent applications to the examiners, and their comparison before and after any redistribution is important.

The patent applications were also classified in up to five classes: very light, light, moderate, heavy, and very heavy, and the classification had as a reference the IGC values, considering ranges equivalent to the average ratio plus one, three or more standard deviations. Then the applications were redistributed with emphasis on the examiners' ZAE and on the classifications. The results show that the medians of the examiners' applications approached the general medians of the division, suggesting that the new distribution is more balanced in volume of data than the original one. Moreover, with the new distribution, the examiners had the majority of their applications allocated within their respective ZAE, i.e., they would examine more applications in their specific areas of knowledge and preference, also suggesting that the new distribution contributes to better efficiency, quality, and motivation.

Additionally, the results obtained suggest that, although the five variables directly related to volume of data tend to be the ones that mostly impact the examination process, all ten variables selected, to some extent, influence the analysis of complexity of patent applications.

On the other hand, as complexity is something relative, to investigate if this complexity indeed captures the examination time/effort, a sensitivity analysis of the model developed was performed in order to verify the correlations of variables and IGC with time. In order to do so, it was then necessary to obtain a new sample of patent applications, referred to as Standard Sample, now with the additional collection of the examination time variable. In this context, simulations considering different variables and standard sample sizes were performed, with application of the PCA method and the model developed, including calculation of the IGC with different criteria and their correlations with time. The results obtained suggest that, for our specific problem, the IGC with greater efficiency and stability was IGC_{Y1} , i.e., using only the first principal component, the one which is most representative as to total data variance.

It is also worth noting that the case including only three variables (number of claim pages, total number of pages, and total number of claims) is the one recommended to perform even more specific practical tests, whether because it captures the influences of the main variables of direct volume of data, given the simplicity for data acquisition and collection (as it does not require separation of independent claims from the dependent ones), or because it has consistently higher correlations of the IGC_{Y1} with time, always close to 0.85.

Based on this new sample with the collection of the time for examination, the patent applications were once again classified into the five classes defined (very light, light, moderate, heavy and very heavy). Such classifications were carried out twice, the first time using the time for examination variable as a reference, i.e., the standard reference classification, and the second time using the IGC_{Y1} ratio, i.e., the

classification suggested by the model. Upon comparison between these classifications according to time and the classifications of the model, the results showed a strong similarity, as the model correctly classified 43 out of the 50 patent applications analyzed, a total of 86%.

After testing the mathematical model and the criteria for classification with the correlations with time, the following step was to perform a first practical complete redistribution test. In order to do so, it was necessary to collect a third and final sample, referred to as Final Redistribution Sample, with 100 patent applications, being 10 applications of 10 different examiners, all using data from previous searches by international offices, so that the profile of this new sample was similar to the profile of the standard sample, as our reference had already been tested. Based on this new sample, we determined the main central tendency statistics of the samples by examiner and calculated the Distribution Balancing Ratios (IBD) both for the original distribution and for the sample redistributed according to the IGC_{Y1}.

The results obtained with the new redistribution showed that there was a better balance in the examination concentration within the ZAE of each examiner, and in the samples of six out of the ten examiners analyzed, there was an increased number of applications distributed to them and pertaining to their own ZAE. Thus, there is evidence that this new configuration contributes for the examiners to work within their specific fields of expertise and knowledge and, consequently, to their efficiency and motivation. It should also be noted that the new redistribution produced a positive effect on the medians of the examiners' samples, which was mathematically quantified by calculating the IBD, which, in the original distribution, had a value of 0.83 and, after redistribution, increased to 0.90.

In short, our results suggest that the mathematical model is able to represent quite satisfactorily the examination time/effort for patent applications. Also, the logic proposed managed to achieve the goal of better balancing the examiners' workload distribution.

Author details


Cesar Vianna Moreira Júnior^{1*}, Daniel Marques Golodne²
and Ricardo Carvalho Rodrigues¹

1 Brazilian National Institute of Industrial Property (Intellectual Property Academy), Rio de Janeiro, Brazil

2 Brazilian National Institute of Industrial Property (Patent Office—DIRPA), Rio de Janeiro, Brazil

*Address all correspondence to: cvmjunior@gmail.com

IntechOpen

© 2021 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] INPI. Plano de Combate ao Backlog. Available at: <https://www.gov.br/inpi/pt-br/servicos/patentes/plano-de-combate-ao-backlog>. [Retrieved on: 2019-08-13].
- [2] INPI. Novo INPI: Protocolo de Madri e Plano de Combate ao Backlog de Patentes. Available at: <http://www.inpi.gov.br/noticias/protocolo-de-madri-e-plano-de-combate-ao-backlog-de-patentes/> [Retrieved on: 2020-03-02].
- [3] INPI. INPI tem novo Plano Estratégico com missão e visão 2021. Available at: <http://www.inpi.gov.br/noticias/inpi-tem-novo-plano-estrategico-com-missao-e-visao-2021/> [Retrieved on: 2020-03-02].
- [4] INPI. INPI e EPO lançam parceria técnica e estratégica na área de patentes. Available at: <http://www.inpi.gov.br/noticias/inpi-e-eipo-lancam-parceria-tecnica-e-estrategica-na-area-de-patentes/> [Retrieved on: 2020-03-03].
- [5] De Saint-Georges, M. & de la Potterie, B. V. P. A quality index for patent systems; *Research Policy*, 42(3), p. 704-719. 2013.
- [6] Philipp, M. Patent filing and searching: Is deflation in quality the inevitable consequence of hyperinflation in quantity? *World Patent Information*, 28(2), p. 117-121. 2006.
- [7] Kim, Y. K., & Oh, J. B. Examination workloads, grant decision bias and examination quality of patent office. *Research Policy*, 46(5), p. 1005-1019, 2017.
- [8] Golodne, D. M., & Silva, A. S. Managing patent examiners' workload by using patent application volume data. *Manuscript submitted for publication*, June, 2020.
- [9] BRAZIL. Law No. 9,279 of May 14, 1996 [Brazilian Industrial Property Law]. Regulating rights and obligations related to industrial property. Available at: http://www.planalto.gov.br/ccivil_03/leis/L9279.htm [Retrieved on: 2017-02-01].
- [10] INPI. INPI Normative Instruction No. 30 of 2013.
- [11] Soares, P. L. L., Moreira Jr., C. V., Rodrigues, R. C. & Braga, E. J. O Papel do Termo Caracterizante nos Documentos de Patentes na Delimitação de Direitos em Propriedade Industrial. *Revista da ABPI – nº 161 – Jul/Ago 2019*, p. 26-40. 2019.
- [12] Oliveira, L. G., Suster, R., Pinto, A. C., Ribeiro, N. M. & Silva, R. B. Informação em Patentes: Ferramenta Indispensável para a Pesquisa e o Desenvolvimento Tecnológico. *Química Nova*, Vol. 28, S36-S40. 2005.
- [13] ORGANIZAÇÃO MUNDIAL DA PROPRIEDADE INTELECTUAL – OMPI. Norma ST.9 - Manual de Informação e Documentação em Matéria de Propriedade Industrial. 2013. Available at: <https://www.wipo.int/export/sites/www/standards/es/pdf/03-09-01.pdf> [Retrieved on: 2019-11-01].
- [14] WORLD INTELLECTUAL PROPERTY ORGANIZATION – WIPO. International Patent Classification (IPC). Available at: <https://www.wipo.int/classifications/ipc/ipcpub> [Retrieved on: 2019-11-01].
- [15] Delvizio, E. Avaliação das estratégias da indústria para-petrolífera pelos respectivos depósitos de patentes. In: *Rio Oil & Gas Expo and Conference 2010*; Rio de Janeiro; 2010.
- [16] Harloff, D. & Wagner, S. Modeling the Duration of Patent Examination at European Patent Office. *GESY; Discussion Paper 170*; 2006.

- [17] Gambardella, A., Harhoff, D. & Verspagen, B. The Value of European Patents. *European Management Review*, 5(2), p. 69-84. 2008.
- [18] Kuhn, J. M. & Thompson, N.C. How to Measure and Draw Causal Inferences with Patent Scope. Working Paper. 2017.
- [19] Marco, A. C., Sarnoff, J. D. & de Grazia, C. Patent Claims and Patent Scope. Working paper. 2016.
- [20] Cockburn, I. M., Kortum, S. & Stern, S. Are All Patent Examiners Equal? The Impact of Examiner Characteristics. NBER Working Paper 8980. 2002.
- [21] Lemley, M. A. & Sampat, B. Examiner Characteristics and Patent Office Outcomes. *Review of Economics and Statistics* 94(3), p. 817-827. 2012.
- [22] Cesare, R. & Timothy, S. Patent Examiner Specialization. NBER Working Paper. 2018.
- [23] Archontopoulos, E., Guellec, D., Stevnsborg, N., de la Potterie, B. V. P. & Zeebroeck, N. V. When small is beautiful: Measuring the evolution and consequences of the voluminosity of patent applications at the EPO. *Information Economics and Policy*, 19 (2), p. 103-132. 2007.
- [24] Van Zeebroeck, N., Stevnsborg, N., de la Potterie, B. V. P., Guellec, D., & Archontopoulos, E. Patent inflation in Europe. *World Patent Information*, 30 (1), p. 43-52. 2008.
- [25] Van Zeebroeck, N., de la Potterie, B. V. P. & Guellec, D. Claiming more: the increased voluminosity of patent applications and its determinants. *Research Policy*, 38(6), p. 1006-1020. 2009.
- [26] Johnson, R. A., Wichern, D. W. *Applied Multivariate Statistical Analysis*, New Jersey: Prentice-Hall, Inc., 5 ed., 797p. 2002.
- [27] Rencher, A. C. *Methods of Multivariate Analysis*. John Wiley and Sons, 2 ed., 740p. 2002.
- [28] Paiva, A. P. *Metodologia de Superfície de Resposta e Análise de Componentes Principais em Otimização de Processos de Manufatura com Múltiplas Respostas Correlacionadas*. Tese (Doutorado) – Universidade Federal de Itajubá, Instituto de Engenharia Mecânica, 257p. 2006.
- [29] Herig, L. C. T. *O Modelo das Componentes Principais: apresentação e interpretações*. Tese (mestrado) – Universidade Federal do Rio de Janeiro, Faculdade de Engenharia. 222p. 1979.
- [30] Vicini, Lorena. *Análise multivariada da teoria à prática*. Monografia (Especialização) – Universidade Federal de Santa Maria, Faculdade de Estatística. 215p. 2005.
- [31] Hongyu, K., Sandanielo, V. L. M., Junior, G. J. de O. *Análise de componentes principais: resumo teórico, aplicação e interpretação*. *Engineering and Science*, V.1, 5ed. 2015.
- [32] Meira, C. T.; Pereira, I. G.; Farah, M. M.; Pires, A. V.; Garcia, D. A.; Cruz, V. A. R. Seleção de características morfofuncionais de cavalos da raça Manga larga Marchador por meio da análise de componentes principais. *Arquivo Brasileiro de Medicina Veterinária e Zootecnia*, 65, p. 1843–1848. 2013.
- [33] Paiva, A.L.C.; Teixeira, R.B.; Yamaki, M.; Menezes, G.R.O.; Leite, C. S.; Torres, R.A. Análise de componentes principais em características de produção de aves de postura. *Revista Brasileira de Zootecnia*, 39, p. 285–288. 2010.
- [34] Savenago, R.P., Caetano, S.L., Ramos, S.B., Nascimento, G.B.,

Schmidt, G.S., Ledur, M.C. Munari, D.P. (doutorado) - Universidade Estadual Paulista, Faculdade de Ciências Agrônômicas de Botucatu, 159p. 2008.

Estimates of genetic parameters, and cluster and principal components analyses of breeding values related to egg production traits in a White Leghorn population. *Poultry Science*, 90, p.2174-2188. 2011.

[35] Silva, M. C., Silva, J. D. G., Borges, E. F. Análises de componentes principais para elaborar índices de desempenho no setor público. *Revista Brasileira de Biomedicina*, v.33, n.3, p.291-309. 2015.

[36] Begnini, S., Almeida, L. E. D. F. Grau de desenvolvimento dos municípios da mesorregião oeste catarinense: caracterização e classificação. *Revista Interações*, v.17, n.4, p.547-560. 2016.

[37] Hongyu, K. Comparação do GGE-biplot ponderado e AMMI-ponderado com outros modelos de interação genótipo \times ambiente. Tese (Doutorado em Estatística e Experimentação Agronômica) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba. 155p. 2015.

[38] Kaiser, H. F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, v. 23, n. 3. p. 187-200. 1958.

[39] Cattell, R. B. The Screen Test for the Number of Factors. *Multivariate Behavior Research*, V. 1, p. 245-276. 1966.

[40] Scremin, M. A. A. Método para Seleção do Número de Componentes Principais com base na Lógica difusa. Tese de Doutorado em Engenharia de Produção, Universidade Federal de Santa Catarina, 124p. 2003.

[41] Sandanielo, V. L. M. Emprego de técnicas estatísticas na construção de índices de desenvolvimento sustentável aplicados a assentamentos rurais. Tese