

CHALLENGES FOR THE 21ST-CENTURY ANALYST

María Julia Martínez Martínez*

Abstract

New technologies have significantly increased the complexity of every task related to analysis. They facilitate the availability of a huge amount of data and offer a wide range of methodological possibilities. Faced with this situation, new analysts complement their training and experience with expertise in data science. They propose models, based on the scientific method, to study the phenomena they have to analyse. The present article seeks to show the conditions that a model must fulfil to be used as a predictive analytical tool. To illustrate it with an example, the CALP Model is described. This model has been developed to assess the terrorist threat phenomenon through the four CALP variables: Capacity, Activity, Lethality and Popularity.

Keywords: Analysis, information management, analytical methodology, data analysis, model, analytical tool, scenario analysis, predictive analytics, terrorist threat.

Resumen

Las nuevas tecnologías han incrementado notablemente la complejidad de todas las tareas ligadas al análisis. Permiten disponer de gran cantidad de datos y ofrecen un amplio abanico de posibilidades metodológicas. Ante esta situación, los nuevos analistas complementan su formación y experiencia con conocimientos especializados en ciencia de datos. Plantean modelos basados en el método científico para estudiar los fenómenos que tienen que analizar. En este artículo se muestran las condiciones que se deben cumplir para que un modelo se utilice como herramienta analítica con capacidad predictiva. Para ilustrarlo con un ejemplo, se describe el Modelo CALP, que se ha desarrollado para valorar el fenómeno de la amenaza terrorista a partir de las cuatro variables CALP: Capacidad, Actividad, Letalidad y Popularidad.

Palabras clave: Análisis, gestión de la información, metodología de análisis, análisis de datos, modelo, herramienta analítica, análisis de escenarios, análisis predictivo, amenaza terrorista.

* Analista de riesgos, mjulia.martinezmartinez@gmail.com

1. Introduction

New technologies have significantly increased the ability to store and distribute knowledge. The number of databases has multiplied considerably in recent decades, as well as their storage systems. The current global interconnection enables the almost immediate availability of a huge amount of data, which continues to grow at an exponential rate.

However, a vast amount of data has limited value for analysis if there is a lack of means to treat them, or to discriminate between the relevant and the unnecessary. Hence the importance of having tools to identify, obtain, and process large volumes of data.

Current technology also offers a wide range of methodological possibilities, so it is essential to make proper planning of the set of procedures and steps to follow during the analysis process.

This article seeks to provide a general overview of the way in which to focus the set of tasks associated with analysis at present. For that purpose, we review some aspects that are essential for a good analysis, as well as the conditions that must be met in order to obtain a model with predictive capacity.

First, the role of the current analyst is described, which is compared with his previous-century counterpart. In addition, some elements that contribute to increasing the quality of the analysis are indicated.

The following section identifies some of the properties that are expected from good information management, in general, and a useful database, in particular.

Subsequently, the importance of using the most appropriate analysis methodology is emphasized, as well as the necessary conditions to formulate a model. About the decisions to be made on methodological matters, special attention should be paid to everything related to data analysis.

Then, the most relevant aspects of the model specification are described. A model is understood as the simplified representation of reality under analysis. It is emphasized that its underlying content is as important as its representation.

The main steps to be taken for model estimation and diagnostics are reviewed consecutively. The requisites and conditions that are required for a model to be used in forecast are also indicated.

Finally, the CALP Model is proposed as a predictive analytical tool based on the scientific method. This is a model that studies the terrorist threat through the four CALP variables: Capacity, Activity, Lethality and Popularity.

2. The role of the analyst

The *role of the analyst* is evolving slowly, gradually. Their work is undergoing major changes and technology is accelerating the speed at which they occur. So, in order to adapt it, progress must be made.

The analysts of the 20th-century were characterized by having a great professional experience in the subject that they analyzed and a solid theoretical preparation, which could have been acquired both in an academic field, as well as through training courses held in their professional environment. Having an acceptable level of data analysis was rare and required receiving previously produced information.

Analysts of the last century were integrated into different categories according to their training and experience. They only undertook analyses of subjects related to their category. Depending on the nature of the subject, the analysis was conducted by an economic, financial, computer, psychological, sociological, legal profile, etc.

However, analysts in the 21st-century do not fit a standard profile in terms of training and experience. Their theoretical training is very diverse, which may be related to both the subject they analyze and to other similar ones. Present-day analysts' professional experiences are equally heterogeneous.

Furthermore, the complexity of the existing information management and processing systems have encouraged many of the current analysts to complement their previous training and experience with specialized knowledge and specific data science skills. Similarly, there are increasing numbers of analysts with theoretical knowledge and professional experience in multiple analytical methodologies.

Two types of analysts currently coexist. On the one hand, there are those who have a solid academic preparation, great professional experience and few notions of data processing and methodology. On the other hand, there are other analysts whose theoretical knowledge may be more limited, but who excel in data processing and, moreover, widely dominate the different methodological alternatives.

The transition from the former analyst to the new one is unavoidable, unstoppable. It is an evolutionary adaptation in the purest Darwinian style. The final stage of the transition period will be reached at the end of the generational relay.

On the other hand, the *work of the analyst* consists in the study of a particular phenomenon. It usually concludes when they achieve the necessary knowledge to prepare a report describing its behaviour, its characteristics, its components, its causes, and, possibly, its future evolution.

An analysis report is understood as the finished product that results from the process of obtaining, treating, interpreting, and evaluating information related to the studied phenomenon.

The quality of the analysis report depends on the existing knowledge about this phenomenon, the amount of such knowledge that the analyst has been able to obtain, process, and understand. Finally, it also depends on the accuracy with which the analyst describes the phenomenon, for which it is required to find the balance between exhaustive explanation and necessary synthesis. In addition, a good report is the one that is produced at the right time.

The quality of the analysis report increases significantly with the availability of a model developed from a proper methodological basis. The analyst must know in detail how to use the model in order to assess the behaviour of the phenomenon under study. The ideal situation occurs when, previously, the analyst himself has been actively involved in the development of the model.

The quality of the report also improves considerably in those situations in which the analyst has a thorough knowledge of all the components of the phenomenon under study. This happens when he understands every stage of the information management process, in other words, when he knows perfectly how the data was obtained, processed, and structured.

3. Information management

Information management is the set of tasks associated with knowledge management that are needed during a research process. In the scope of an analysis, it comprises every step from the emergence of new pieces of information until the analyst has assimilated it and acquired it as a form of knowledge.

An analysis requires information that may be available in the form of pure raw data (either structured or unstructured), news or elaborate information.

In the field of knowledge management, each piece of information allows us to see only a part of reality. This was the case for the six blind men of the ancient Indian parable, who described the elephant from the part of the animal they had touched: a wall (side of the animal), a snake (trunk), a spear (tusk), a tree (feet), a fan (ear) or a rope (tail).

To understand the true nature of things, one cannot rely solely on the description of one part, even if it is very accurate. As long as the conditions of credibility, reliability and relevance are met, the use of various pieces of information from different sources contributes to more complete knowledge.

The work scheme of traditional analysis required a clear separation between tasks strictly related to obtaining and processing information, and the study itself.

The common tasks of obtaining and processing information are the *search* for data and information, the *verification* of its accuracy, authenticity and relevance, its *classification* in taxonomic units according to different categories, and its *integration* into a database, both in a structured and unstructured way.

In principle, it could be asserted that the value of a database depends on the quantity, quality, and detail of its content. The *quantity* is associated with the number of records it contains. It increases with the scope and extent of source searches, as well as the data loading processes. The *quality* increases as its records become more accurate, authentic, and relevant. The *detail* is the greater the more thorough the hierarchical, relational, and systematic sorting system of its content. In addition, a relational database makes it easier to extract data by reducing the number of queries.

It must be taken into consideration that databases that do not have the means to extract the information in a treatable format are not useful for analysis, as it happens with those

that are outdated due to the lack of agility in their data loading processes.

Therefore, the true value of a database depends primarily on the quantity, quality, and detail of updated content that can be extracted in a format that can be used in the analysis.

On the other hand, in the current interconnected society, the knowledge available in the form of records stored in databases is growing exponentially. More and more institutions are liberalizing their databases and granting public access to them.

Analysts have at their disposal an increasing number of records to study. Besides having the information provided by their own databases, in which access is restricted, they also have unlimited access to the huge amount of content provided by open sources.

Consequently, the value of an analysis depends both on the capacity of the analyst to extract information, and on his ability to study, understand, interpret, and evaluate it. The process of analysis will be more agile the more structured the data or the more treatable the information.

When data, information or news are stored and extracted from databases in an unstructured way, the analyst must transform them and provide them with a coherent and relational structure. Transformation and structuring is a process integrated into the treatment and analysis of data. Such procedures must have been previously specified, during the methodological definition stage.

4. Analysis methodology

The *analysis methodology* is the set of procedures and stages that are followed during a research process. It is especially applied to those studies focused on acquiring a precise knowledge of reality that, as a basic foundation, are approached with scientific rigor.

The selection of the most appropriate methodology is conditioned by the phenomenon to be studied. For this reason the phenomenon must be stated in advance, clearly and unambiguously. This implies delimiting the context in its three dimensions: the *space* or geographic location where it occurs, the *time* or interval in which it takes place, and the *population* it encompasses.

Some analysts direct their study to draw conclusions that are consistent with the knowledge they have previously acquired through their training and experience. The result is a biased report, whose findings are certain and resounding statements. Their predictions about what will happen in the future will be correct when the past is repeated, while they will be completely erroneous when unexpected events occur.

Those who perform this type of analysis can be called *followers of the revealed doctrine*. They are considered great experts in their fields of analysis and they were highly regarded in the past, mainly because of the large number of times they anticipated future events. However, this type of analysis is becoming obsolete in the current context, in which a constant stream of unexpected events takes place.

The current panorama is characterized by its complexity, which has led to a radical change in methods of analysis. There is an increasing number of *supporters of the scientific method*, whose approach is diametrically different from that of the followers of the revealed doctrine.

Analysts who follow the scientific method doubt their prior knowledge and do not accept any idea as absolute or definitive. Even though they use variables that reflect certain past facts that are highly likely to be repeated; however, they do not discard unexpected situations, which are incorporated into the analysis as variables in a specific manner.

In addition, the scientific method makes it possible to obtain a *model*, a simplified representation of the phenomenon under study. Studying phenomena using models is equivalent to combining the classic holistic approach with risk factor assessment techniques.

A model also allows completing the study with a *scenario analysis*, in which different alternative situations are simulated, including a wide range of them, from the most unrealistic to the most likely. The aim is to study how the phenomenon under analysis is affected by each one of the scenarios.

The result is a report that instead of offering a unique resounding statement, it has the necessary elements for describing the multiple ways in which the phenomenon under study may evolve, as well as the likelihood of occurrence of each of those situations. The advantage of producing a report that anticipates, for example, negative future events

early enough, is that it opens the door for action to be taken to change the course leading to those events.

The methodological aspects of the research are complemented by the *logical aspects*, which are the conceptual elements involved in the process and the relations that may exist between those elements. The main logical elements are: phenomenon, theory, hypothesis, variables, facts, data, verification and model.

From a methodological point of view, the analysis process using the scientific method begins with a previous theory that explains the behaviour of the phenomenon through a series of statements in which the features that characterize it are described and hypotheses are proposed.

Then, the variables that best represent the phenomenon and its properties are defined and the relations between those variables are explained.

Variables are measurable magnitudes that represent phenomena, facts, circumstances, situations, activities, entities, people, etc. that may be independent of each other or be linked by a relation of dependency, or a relationship of cause, or one of effect.

The next step is to check the validity of the previous theory and verify the hypotheses, based on the new information provided by the detailed and rigorous study of the variables and their relationships. In the case of the theory resulting invalid, it must be reframed so that all this new information is taken into account.

The research process concludes when a validated and verified theory is available, which is the necessary requirement to formulate the model that best explains the phenomenon that has been studied.

With regard to the decisions to be made on methodological matters, one of the most important is the selection of the most appropriate methods for the study of variables and their relationships.

Variables are measured in observation units from which data is extracted. The following section indicates some of the basic aspects that must be taken into account in the procedures for analysing this data.

5. Data analysis

The stage of *data analysis* provides information on the variables and their relationships. Data is the raw material for analysis. They are obtained quantitatively, qualitatively, or mixed form (both quantitative and qualitative).

In the academic field, the quantitative analysis is commonly associated with the natural sciences and the qualitative analysis with the social sciences. In scientific disciplines, quantitative data is often prioritized because it is more concrete and its measurement more accurate. However, qualitative data is full of nuances that the other kind does not always capture.

When restricting analyses to quantitative data, only one part of the reality is perceived and studied, whereas incorporating additional qualitative information results in a more complete knowledge thereof. Structuring, quantifying, and incorporating qualitative information into the analysis necessarily leads to a more detailed understanding of the variables.

Therefore, reports produced from mixed data analysis are more complete, as they contain more exhaustive and accurate descriptions.

However, the quantification process must be carried out with objectivity and scientific ground. The lack of rigour results in poor knowledge of the reality, which entails that the analysis report will be flawed with inaccuracies or erroneous conclusions, which undermines its quality.

On the other hand, the study of the variable can be focused according to its structural or conjuncture nature. It is structural when all the factors and circumstances remain unchanged over the course of time, whereas a conjunctural situation is a specific context at a particular moment in time. That is why the specified time interval is decisive and conditions that the data analysis is carried out in a static, dynamic, or mixed way.

Static analysis is a study whose approach is synchronous. It focuses on the behaviour of the variables at a specific time or in a given period of time, without considering their past evolution. It uses cross-sectional methods such as descriptive studies (which merely describe the variables and their characteristic features) or explanatory studies (which focus on their causes and effects). It is usually recommended for circumstantial situations analysis.

Dynamic analysis has a diachronic nature. It focuses on the joint evolution of the variables and their relations, with special attention to the successive stages in which small changes, alterations, or large transformations occur. It uses longitudinal analysis techniques such as time series. It is recommended for retrospective studies (which are limited to historical research), counterfactual studies (hypothetical investigation of historical events) and prospective studies (which focus on the exploration or the prediction of the future development of the variable).

Mixed analysis, also known as panel data, combines a cross-sectional and temporal approach to study changes in the variables and their relations. It uses time-series analysis techniques and nonlinear dynamics. As the static and dynamic nature of the variables is considered together, it is especially recommended for the dynamic structural analysis, as well as for complex prospective studies.

The data analysis process begins once: (1) it has been established which type of analysis is most appropriate according to the nature of the study; (2) all the qualitative and quantitative information necessary to understand the behaviour of the variables has been obtained; and (3) the information has been structured in the form of data. The complexity of the relationships will condition the use of univariate, bivariate, or multivariate analysis techniques.

There are two essential stages in data analysis: their classification into categories and the comparison of those categories.

Classification consists of grouping what is similar and separating what is different. In order to do this, it is essential to define classification criteria, which are the rules to discriminate values and to determine the conditions that must be met for data to fall into a particular category.

The more criteria (and therefore the more different categories) are defined, the more elaborately the variables and their relationships can be studied. A good classification system leads to more thorough knowledge.

The classification process must comply with three fundamental rules: (1) the classification criteria must be sole, (2) the process has to be complete and exhaustive, without omitting any single element, and (3) the categories must be exclusive so that no datum can be included in different categories simultaneously. In addition, the entire set of data has to be treated as one more element.

Another essential stage in data analysis is the *comparison between categories*. The comparison criteria must be homogeneous and the same for every category. The more comparisons are made between the different categories; the more interactions can be evaluated.

One of the basic tools to study the comparison between categories is correlation analysis. Correlation is a dimensionless measure of the linear relationship between two variables. Its value is a number between zero and one. A near-zero correlation occurs between independent variables or with variables whose relationships are nonlinear. Correlations significantly higher than zero are found in variables that have a linear relation and vary together, to a greater or lesser degree. A correlation close to one implies that the magnitude of the change of one variable is similar to that of the other.

The value of correlation is an important measure of the magnitude of changes in relationships. The sign is also a relevant indicator. A positive correlation means that the changes have the same direction, meaning that if the value of one variable increases, the value of the other increases as well, and that when the value of the variable decreases, the value of the other decreases likewise. A negative correlation implies that any increase in the value of one of the variables leads to a decrease in the value of the other, and vice versa.

It is important to note that the analysis of the correlations should not lead to any hasty conclusions about the dependence of the variables, since it is possible that the reasons for said correlation are spurious in nature. This occurs when they evolve over a period of time in a certain way not causally, but casually.

Covariation is not the same as the relationship between variables. Covariation reflects that two variables vary together, while the relationship entails some type of connection. For a dependency relation to exist, it is not enough that the variables vary together over a certain period of time, but there must be an underlying structural link.

6. Model specification

After completing the data analysis, the analyst has acquired enough knowledge to describe the behaviour of the studied phenomenon and has the necessary elements to build the model.

The *specification of the model* consists in developing a simplified representation of the analyzed phenomenon. In this task, the underlying knowledge that has been acquired during the study process, expressed in terms of variables and their relations, is as important as the way in which it is represented.

The acquired *underlying knowledge* is incorporated into the model in the form of conceptual and relational elements. Conceptual elements represent the phenomenon, its main components, and other related phenomena, circumstances, situations or events. Relational elements are used to represent the connections that exist between conceptual elements.

The conceptual elements of the model are the variables, while the relational elements represent the link that exists between those variables.

The relationships are links of interdependence by cause-effect. They have intensity and direction. The intensity measures the magnitude of the relationship between one variable with another one. The direction indicates the course of the connection path, and it is displayed with a positive or negative sign.

While the model must reflect reality as accurately as possible, there must be balance between precision and complexity. Employing a large number of variables leads to a more complex model, but this does not necessarily mean that it is better.

The incorporation of new variables must be carried out with caution since the introduction of redundant variables creates the problem of multicollinearity, which undermines the explanatory value of the model. Therefore, only those variables that optimize the explanatory capacity of the model should be included.

On the other hand, the *representation of the model* can take various forms, such as graphics or identity.

The *graph* is a means of representation characterized by the use of visual resources. It is a very effective tool for displaying content. It consists of a network made up of nodes and connections. Nodes represent concepts and connections show the way they are related.

Concept maps and mind maps are graphics commonly used in analysis. Charts are also common. They are a specific type of graphic representation in which magnitudes are displayed by using mainly numerical components.

Specifying the model graphically is to represent variables using nodes, and their relationships by means of graphic links. Variables take the form of images or geometric figures, and relationships connect variables using unidirectional or multidirectional lines or vectors.

The *identity* is a means of representation in the form of a formula that transcribes a statement into mathematical language. It consists of two mathematical expressions separated by the equal sign ($=$). It is characterized by the use of elements such as numbers, letters, signs, and operators, some of which can be grouped in algebraic expressions.

Algebraic expressions are groups of letters and numbers separated by operators. Each group of the algebraic expression, which are called terms, is composed of a literal part and a coefficient, as well as an exponent when the expression is more complex.

The literal part is a variable whose specific value is not always fixed or known. It is usually represented by a letter of the Latin alphabet. The coefficient is a multiplicative factor that indicates the degree in which the variable it accompanies varies. It is represented by a letter of the Latin or Greek alphabet. Operators are algebraic symbols ($+$, $-$, $*$, $/$, etc.) that indicate how the mathematical operations are performed on the elements between which they are placed.

When specifying the model in the form of identity, the variable representing the studied phenomenon is located on the left side, whereas the other variables that form the phenomenon or explain its behaviour are located on the right side in the form of algebraic expressions. These expressions are simple or complex, depending on the linearity or non-linearity of the relationship between the variables.

7. Estimation, diagnostic and use

The specification in the form of mathematical identity allows the estimation of the model, whose purpose is to obtain the quantitative value of the coefficients.

The first step is to summarise the data, using descriptive statistics tools. Each of the variables is measured, their frequency distribution, their centralization measures, such as mean, median, or mode; their dispersion measures, such as standard deviation or the coefficient of variation; and their shape, such as symmetry or kurtosis.

In addition, all related variables are studied together by multivariate frequency distribution, mean vectors, linear association measures (covariance and correlation), and the variances and covariance matrix.

In cases where the data are not complete but come from a sample population, the behaviour of that population must be generalized using statistical models of probability distribution. To do this, hypotheses about the distribution function are raised, and the probability of each of the possible values is calculated. The model is estimated using statistical inference techniques.

Statistical inference techniques use parametric and nonparametric methods, depending on their purpose. In addition, they are classified into classic and Bayesian methods according to the type of data they use.

When the estimation of the model stage is completed, its validity must be evaluated. This is the stage *diagnosis of the model*. It is then that the assessment of the model that has been previously formulated takes place, this is done in both the study of complete populations and in those cases of inferences from samples of data. In order to do this, the hypotheses raised are contrasted, especially those related to the parameters and the structure of the model.

Concerning the contrasts of statistical significance, a measure of the level of discrepancy is established, as well as the rule that makes it possible to judge whether such a discrepancy is acceptable or not. In any contrast, two types of errors can be made: the Type I error, which consists of rejecting the hypothesis when it is true, and the Type II error, which is made when accepting a hypothesis that is false.

Likewise, the properties of the estimators must also be evaluated. Bias, efficiency, robustness (consistency), and convergence are to be assessed.

Once the contrasts are statistically significant, the result is generalized and the model is considered verified.

In regard to the *use of the model*, its explanatory capacity determines whether it can be used to describe the phenomenon investigated in the analysis report. In addition, its predictive capacity conditions its use to obtain predictions. Those predictions can be incorporated into the report as possible future scenarios. Assuming conditions are propitious, the model can be used as a predictive tool.

Concerning prediction, it is important to highlight that variables fulfil a double function: they have an explanatory role and an instrumental role. The phenomenon is represented in the model by what is called the *target variable*, since obtaining its resulting value is the purpose of the prediction. The circumstances, situations, or events that affect the phenomenon are represented by *instrumental variables*, as they are the instruments that operate as a *transmission channel of changes*. This means that any alteration in one of the instrumental variables due to external factors will be transmitted through the channel and will produce a modification in the phenomenon.

Being provided with a *predictive analytical tool* simplifies the scenario-generating process through simulations. This facilitates the measurement of the impact on the target variable as a result of any possible changes that contemplate the various simulated scenarios.

Early detection of any circumstances leading to unfavourable situations can be useful for prevention systems, provided that those systems have the effective means to take action to change the course of events from the path to negative scenarios.

Regarding the conclusions drawn from a model, caution should always prevail. And making a clear distinction between the objective (data and proven facts) and the subjective (opinions, interpretations, estimations, and generalizations) is particularly relevant.

Ultimately, a model does not reflect the reality in its entirety. A model is a simplified representation of reality, defined on the basis of methodological criteria.

8. Case study: CALP Model

CALP Model is the name given to a simplified representation of the *terrorist threat* phenomenon, whose behaviour is mainly explained by the four CALP variables: *Capacity*, *Activity*, *Lethality*, and *Popularity*.

While the model that is proposed in this article is purely theoretical, its potential practical implementation could be an analysis tool for the systematic study of the phenomenon in an organized and structured manner.

It should be taken into consideration that threats are phenomena that are to be studied within a context of great uncertainty. This means that it is unclear whether or not events related to such threats will occur, neither is there any certainty whether said events could cause personal injuries or material damage.

The model being proposed is a multivariate structure with five variables (see Figure 1). The *target variable* is the *Threat*, whose behaviour is explained through four *instrumental variables*, which are *Capacity*, *Activity*, *Lethality*, and *Popularity*.

In the sphere of security, a threat is understood as that which has the capacity and decision to attack life or public and private interests. In the particular case of the terrorist threat, attacks directly endanger human life, security, the particular possessions of individuals, as well as public property.

For the practical use of this model, the Threat variable should be defined clearly and unambiguously according to its context, space and time.

The *context* of the Threat may be limited to a particular terrorist group or a set of more or less related groups. With regard to *space*, the study may focus on the global nature of the terrorist phenomenon or on the precise territory where the group, the groups and their operational cells have established. And concerning the *time*, the study may be narrowed to a precise point in time, or extended to study its evolution over a longer or shorter period.

On the other hand, the use of this model in practice requires a concrete and precise definition of each of the instrumental variables. It also requires the specification of the observation units, from which data are to be extracted in order to measure the variables.

Capacity has various connotations: operational capacity, destructive capacity, capacity to inflict damage and response capacity, among others. It can be quantified from their military strength, which includes the number of their troops, the weapons in their arsenals and their available financial resources.

Activity includes the series of direct and indirect actions that the terrorist group perpetrates with the purpose of instilling terror as a means of achieving its objectives. The acts of martyrdom may also be included, which are those that occur in terrorism for the purpose of creating martyrs or to

brutally and severely punish the rest of the people in the world, who do not share their radical vision.

Lethality is quantified based on the number of victims. It is usually limited to those directly affected by violent acts, both the attacked and the attackers themselves, including those who perpetrate acts of martyrdom. Indirectly affected victims may also be included, but obtaining this data can be extremely difficult.

Popularity is measured by the media impact. In order to instil terror, acts have to be widely disseminated by the media. Without exposure, they are not terrorist acts but mere violence without propagandistic power, which destroys their reason for existence.

With regard to the existing relations between variables being incorporated into the model, a positive causal connection is expected to occur. This means that any increase in Capacity, Activity, Lethality or Popularity poses a greater Threat. An increase in Capacity is an increase in resources, which leads to an increase in Activity. More Activity means more Lethality, which generates more Popularity. More Popularity, then, means more Capacity, either for the incorporation of new fighters in the terrorist force or for the reception of more economic contributions from supporters.

However, incorporating acts of martyrdom into the model can affect the relationships between variables, since in this case causal connections can be positive or negative. While acts of martyrdom generate a lot of popularity, there is no doubt that these actions decrease capacity, due to the death toll of combatants.

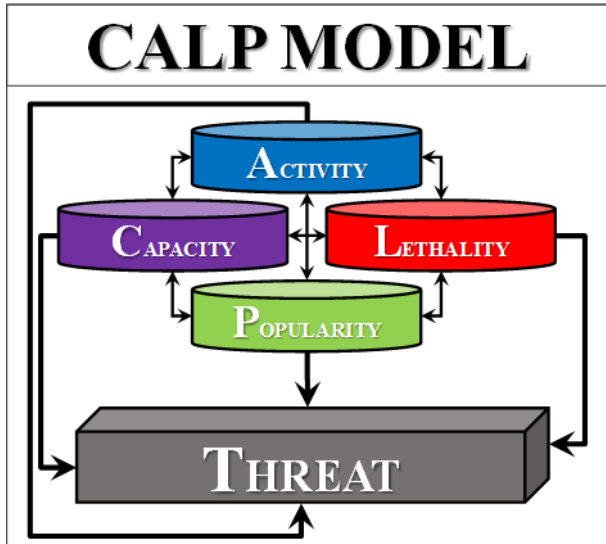
To display the CALP Model in the form of mathematical identity, a multivariate linear function is formulated. The target variable Threat (T) is composed of the sum of four algebraic expressions that represent the impact of the four instrumental variables: Capacity (C), Activity (A), Lethality (L), and Popularity (P). The coefficients α , β , γ , δ measure how intense is the effect of those four instrumental variables on the Threat.

$$\text{CALP} \rightarrow f(T) = f \begin{pmatrix} C \\ A \\ L \\ P \end{pmatrix} \rightarrow T = \alpha C + \beta A + \gamma L + \delta P$$

To present the CALP Model as a graph (see Figure 1), cubic and cylindrical shaped nodes are used to illustrate the variables of the model. The cube depicts the target variable, Threat, and the cylinders portray the instrumental variables,

Capacity, Activity, Lethality, and Popularity. Unidirectional and bidirectional vectors show the orientation of causal relationships.

Figure 1: The CALP model



The practical use of the CALP Model requires the extraction, processing and analysis of the data that the variables measure, as well as the estimation of the coefficients by the most suitable analytical method, based on methodological criteria. To display the resulting model as a mathematical identity, the symbols representing the coefficients are replaced by their corresponding estimated numeric values.

Given the appropriate conditions, the model can serve as a powerful *analytical tool* with great explanatory and, potentially, predictive capacity.

The purpose of the prediction is to obtain the quantitative value of the Threat from the equally quantitative values of the four instrumental variables. Capacity, Activity, Lethality and Popularity operate as a channel through which changes are transmitted toward the Threat. This analytical tool could also be used to generate alternative scenarios through simulations that quantify the variation of the Threat.

Furthermore, this tool could be useful for prevention systems, given that it allows the early detection of those unfavourable situations that aggravate the terrorist threat, as well as for quantifying the deterioration of those situations. If those prevention systems had the means to take actions on external factors that affect its capacities, its activities, its lethality, or its popularity, the course of events leading to the worst threat scenarios could be changed.

In the end, caution must prevail. No model can reflect reality in its entirety. The CALP Model is a simplified representation of the terrorist threat. Hence, its validity is conditioned by the methodological criteria adopted.

9. Conclusions

The challenges the 21st-century analyst currently faces are becoming more numerous and complex. The storage and distribution systems provide a huge amount of data. Global interconnection enables the almost immediate availability of all sorts of information and offers multiple analytical possibilities. Therefore, technology has significantly increased the complexity of all analysis-related tasks.

The *role of the analyst* is evolving slowly, gradually. Many analysts are complementing their training and experience with specialized knowledge and specific skills in data science. More and more analysts are excelling in the field of information management and processing systems. Likewise, they are mastering the various methodological alternatives that can be applied in each situation.

The *work of the analyst* is undergoing major changes and technology is accelerating the speed at which they occur. This work consists of the study of a particular phenomenon and in the preparation of a report describing its behaviour. The quality of the report improves noticeably when the analyst has a deep knowledge of the information management process. Using a model also improves significantly the quality of the report and, particularly, in those cases in which the analysts themselves are actively involved in the selection and development of its methodological criteria.

Information management is one of the most essential aspects of the analysis. To understand the true nature of a phenomenon, many different pieces of information from many sources must be used. The value of the analysis depends on the ability of the analysts to study, understand, interpret and assess the phenomenon under study, which is entirely dependent on their capacity to extract worthwhile information.

The *analysis methodology* is another crucial aspect of the analysis. It is the set of procedures and stages that are followed during a research process. The methodology based on the scientific method is the most suitable option in the complex current context. A combination of both the holistic approach and the factor assessment techniques enables the

analyst to develop a model to explain the behaviour of the phenomenon under observation.

During the *data analysis* stage, the variables and their relationships are measured. The structural or conjunctural nature of the study of the variables, as well as to the specified time interval, conditions the analysis to be static, dynamic, or mixed. Category classification and the comparison of these categories are critical stages within the data analysis process.

A simplified representation of the behaviour of the phenomenon is developed in the *model specification* stage. This representation can take different forms, such as a graph or an identity. The *graphic form* represents the variables and their relationships by means of visual resources such as nodes and connections. The *identity form* resorts to a formula to transcribe the description of the behaviour of the phenomenon into mathematical language.

In the *estimation of the model* stage, the quantitative value of the coefficients that accompany the variables is obtained. Its validity is assessed in the *diagnostic of the model* stage, using hypothesis contrasts on the parameters and on the structure of the model.

The *use of the model* depends on its explanatory and predictive capacity. The adopted methodological criteria determine whether the model can be used as a predictive analytical tool.

Having an adequate *analytical tool* for prediction simplifies scenario analysis. It helps to understand how the mechanism named *change transmission channel* works, as it measures the scale of the variations in the phenomenon when external factors alter the value of the other variables.

In general, analytical tools facilitate the production of analysis reports. In particular, predictive analytical tools are especially useful for prevention systems when they can take actions to change the course that leads to the most adverse scenarios.

The *CALP Model* is an example of a model based on the scientific method. Its potential practical implementation with the appropriate methodological criteria could transform it into an analytical tool to assess the terrorist threat with great explanatory and predictive capacity. It could be used to generate different alternative scenarios, which could lead to study in depth how any potential alteration in an

external factor ends up causing an impact on the level of the terrorist threat.

An analytical tool such as the CALP Model allows the early detection of unfavourable situations. Thereupon, it could be particularly useful for those who have the means to take preventive measures to keep the level of the terrorist threat from worsening.

However, and although analytical tools ease the burden of many of the tasks associated with analysis, they should be used with great caution. They are based on models that do not reflect the reality in its entirety but only the part of the reality under analysis, based on methodological criteria.

The current state of analysis is complex and, in view of the high speed of technological developments, it will inevitably become more complex in the future. However, this situation is an extremely beneficial opportunity for the 21st-century analyst. The range of methodological possibilities will be broadening even more, which will entail huge progress in the analysis models. Undoubtedly, these will also lead to a high development of the analytical predictive tools.

10. References

- Abad, F.J. et al. (2011) *Medición en ciencias sociales y de la salud*. Madrid: Editorial Síntesis.
- Artola, C. & Galán, E. (2012). Las huellas del futuro están en la web: construcción de indicadores adelantados a partir de las búsquedas en internet. *Banco de España Documentos Ocasionales*, N° 1203.
- CES. (2017). La digitalización de la economía. *Consejo Económico y Social Colección Informes*, 03/2017.
- Choi, H. & Varian, H. (2012). Predicting the Present with Google Trends. *Economic Record*, 88(1), pp 2-9.
- Cussac, J.L. & Giménez, F.F. (2017). Una metodología para el análisis de las amenazas a la seguridad, la evaluación de las respuestas y su impacto sobre los derechos fundamentales. *Cuadernos de estrategia*, No 188, p. 15-64.
- De la Corte Ibáñez, L. & Blanco Navarro, J. M. (2014) *Seguridad nacional, amenazas y respuestas*. Madrid: LID Editorial Empresarial.
- Dhrymes, P. (2017) *Introductory econometrics*. Cham: Springer.

- Enomoto, C. E. & Douglas, K. (2019). Do internet searches for islamist propaganda precede or follow islamist terrorist attacks. *Economics & Sociology*, 12(1), 233-247.
- Fernández-Montesinos, F.A. (2017). Terrorismo y contra-terrorismo. *Cuadernos de estrategia*, No 188, p. 65-120.
- Fred, Y. Y. (2017). *Scientific metrics: Towards analytical and quantitative sciences*. Singapur: Springer.
- Fuleky, P. (2020). *Macroeconomic Forecasting in the Era of Big Data: Theory and Practice*. Cham: Springer.
- Ghirelli, C. et al. (2019). Algunas implicaciones de las nuevas fuentes de datos para el análisis económico y la estadística oficial. *Banco de España Boletín económico*, n. 2.
- Gwani, A. A. & Aliyu, A. (2018). E-Statistics as an instrument for reliable source of data. *Journal of Physics*, Conference Series Vol. 1132, No. 1, p. 012070.
- Hammer, C. et al. (2017). Big Data: Potential, Challenges and Statistical Implications. *International Monetary Fund Staff Discussion Notes*, 17/06.
- Heuer, R. J. & Pherson, R. H. (2015) *Técnicas analíticas estructuradas para el análisis de inteligencia*. Madrid: Plaza y Valdés.
- Martínez, M. J. (2016). Riesgo de financiación del terrorismo: Vulnerabilidades de los canales de pago. *Revista de Pensamiento Estratégico y Seguridad*, 1(1), 66-78.
- Martínez, M.A. et al. (2010) *Gestión de la información y del conocimiento* Ediciones CEF.
- McLaren, N. & Shanbhogue, R. (2011). Using internet search data as economic indicators *Bank of England Quarterly Bulletin*, Q2.
- Neusser, K. (2016) *Time series econometrics*. Bern: Springer.
- Novales, A. (1989) *Econometría*. Madrid: McGraw Hill.
- Papoutsoglou, M. et al. (2019). Extracting Knowledge from on-line Sources for Software Engineering Labor Market: A Mapping Study. *Institute of Electrical and Electronics Engineers IEEE Access*, 7, 157595-157613.
- Peña, D. (1989). *Estadística Modelos y métodos*. Madrid: Alianza Editorial.
- Pfanzagl, J. (2014). *Mathematical Statistics*. Berlin: Springer.
- Ramsay, J. & Hooker, G. (2014). *Dynamic Data Analysis*. New York: Springer.
- Rubio, Z. et al (2018). Modelo de gestión del conocimiento apoyado en las tecnologías de información y comunicación *Criterio Libre*, vol. 16, no 28, p. 211-226.
- Thaler, R.H. (2015). *The making of behavioral economics misbehaving*. Nueva York: W. W. Norton & Company, Inc.
- Wright, S. (1921). Correlation and causation. *Journal of Agricultural Research*, Num. 20, p. 557-585.