Interreg

Co-funded by the European Union

Romania - Hungary

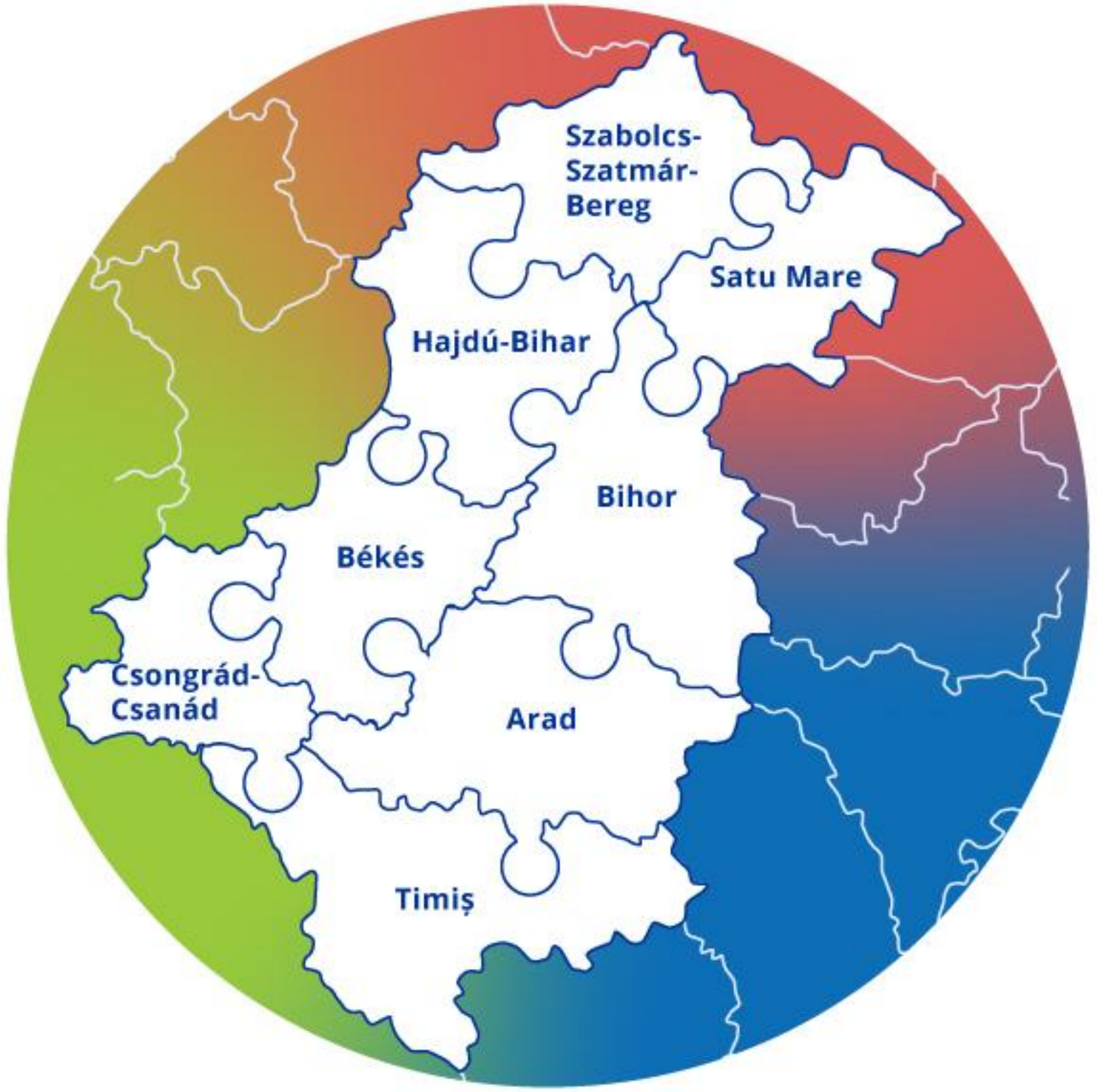
GOVERNMENT OF ROMANIA

HUNGARIAN GOVERNMENT

Project title: *Exchange of Experience for research and usage of Artificial Intelligence' Advanced Techniques in finance, accounting and business administration in Bihor-Hajdu Bihar Region*

Project Acronym: E2U-AI, **JEMS Code:** ROHU00120

Course 1: Artificial intelligence methods applied in finance, accounting and business administration



interreg-rohu.eu

ORADEA, June 2025

Grey Data Analysis
Grey Numbers

Lecturer: Camelia Delcea, Professor Habil.
Bucharest University of Economic Studies
Economic Informatics and Cybernetics Department
E-mail: camelia.delcea@csie.ase.ro
Phone: 0040 769 652 813

Road Map

1. Introduction to Grey Systems Research
2. Grey Numbers and their Mathematics
3. Practical Application

Next

1. Introduction to Grey Systems Research

1. Introduction to Grey Systems Research

Development History

Grey Systems Theory Appearance

In 1982, Professor **Julong Deng's** paper titled "**The Control Problems of Grey Systems**" was the first paper on grey systems to be published in the Systems and Control Letters journal (Deng 1982a).

In that same year, Professor Deng also published "**Grey Control System**" in Chinese and the paper was published by the Journal of Huazhong University of Science and Technology (Deng 1982b).

The publication of these two seminal articles indicated that a new and cross-sectional discipline named grey system theory came into the world.

Grey Systems Theory Appearance

Statistical methods

Uni-variate analysis - Beaver (1967), Tamari (1966), Moses and Liao (1987);

Multi discriminate analysis – Altman (1968), Deakin (1972), Edmister (1972), Blum (1974);

Linear probability model – Myer and Pifer (1970);

Stochastic model: logit – Ohlson (1980); probit – Zmijewski (1984).

Grey Systems Theory Appearance

Artificial intelligence methods:

Decision tree – Marais, Patel and Wolfson (1984), Frydman, Altman and Kao (1985);

Fuzzy set theory – Zimmerman (1991);

Grey systems theory – Lin and Liu (2005);

Case based reasoning – Jo, Han and Lee (1997), Park and Han (2002);

Genetic algorithms – Varreto(1998);

Support vector machine – Min and Lee (2005);

Data envelopment analysis – Cielen and Vanhoff (2004);

Rough sets theory - Dimitras et. al. (1999), McKee (2003);

Neural networks:

- BPNN = back propagation trained neural networks – Tam and Kiang (1992), Wilson and Sharda (1994), Bell (1997), Atyle (2001);
- ANN = artificial neural networks – Nasir et. al.(2000), Lei et. al. (2009);
- PNN = probabilistic neural networks – Yang, Pltt and Platt (1999);
- SOM = self organizing map – Serrano-Cinca (1996), Kaski et al.(2001), Lee et al.(2005);
- CASCOR = cascade correlation neural networks – Lacher et al. (1995)

Grey Systems Theory Appearance

In a speech generically titled "**Why the world is grey**?", given in 2011 at a conference dedicated to grey systems theory Professor Andrew AM, CEO of WOSC (World Organization of Systems and Cybernetics) points to some aspects of grey systems theory.

The speaker starts his argument by considering the observation made by the WOSC President, Vallee, that "**we live in a grey world**". Thus, Andrew [5] shows that many of the mathematical methods and models used in analyzes are not as robust as they are supposed to be. An example in this direction is given **by the assumption** that the methods are based on the fact that **data are normally distributed or that they follow a Gaussian distribution**. In fact, the speaker argues that these distributions are quite rare in practical applications [5].

Grey Systems Theory Appearance

In the 40 years of development, the grey systems theory has attracted a series of **prominent scholars** such as, but not limited to: Professor Liu SF, Professor Yang Y, Professor Xie NM, Professor Lin Y, Professor Javed SA, Professor Hipel K, Professor Salmeron JL, Professor Mi C, Professor Yuan C, Professor Javanmardi E, Professor Mierzwiak R, Professor Khuman A [8, 11–21], who have enriched the knowledge related to grey systems theory and have contributed to its continuous development.

Grey Systems Theory Appearance

A series of **associations and societies** have promoted over the years the grey systems theory:

- International Association of Grey Systems and Decision Sciences (IAGSUA),
- Grey Systems Society of China (GSSC),
- Grey Systems Society of Pakistan (GSSP),
- Polish Scientific Society of Grey Systems (PSGS),
- Chinese Grey System Association (CGSA),
- Grey Systems Committee (IEEE Systems, Man, and Cybernetics Society),
- Centre for Computational Intelligence (De Montfort University), etc. [23–27].

The associations have been actively involved in organizing conferences, roundtables, meetings, events, journals, all of them with focus on the development of the grey systems theory.

Grey Systems Theory Appearance

The fields of applicability for the grey systems theory are vast and they include, but they are not limited to: **agriculture** [28], **forestry** [29], **electricity** [30], **manufacturing** [31], **online social networks** [32], **healthcare** [33, 34], **aviation** [35], **traffic safety** [36, 37], **tourism** [38], **telecommunications** [39], **exploring human cognitive capacity** [40], **human preference**[41], **criminality assessment** [42], **pandemics** [43], **natural resources management** [44], **waste management** [45], **portfolio analysis** [46], **enterprises competences** [47], etc.

Grey Systems Theory Appearance

Also, the rapid development of grey systems theory and the interest manifested in the use of this new theory into practical applications by researchers from all around the world has highlight the need from the scholars involved in its development to **provide explanations** regarding different elements associated with the theory. As a result, in the scientific literature, one can find various papers containing in the title the "*explanation of terms of*" syntagm followed by the subject of the explanatory paper:

"Explanation of terms of grey models for decision-making" [48] ,

"Explanation of terms of grey numbers and its operations" [49],

"Explanation of terms of grey forecasting models" [50],

"Explanation of terms of grey incidence analysis models" [51],

"Explanation of terms of grey clustering evaluation models" [12],

"Explanation of terms of concepts and fundamental principles of grey systems" [52],

"Explanation of terms of sequence operators and grey data mining" [53].

Grey Systems Theory Evolution

Table 2. Main information about data

Indicator	Value
Timespan	1987:2021
Sources (Journals, Books, etc.)	286
Documents	869
Average years from publication	5.62
Average citations per documents	23.87
Average citations per year per doc	3.53
References	24,409

Table 3. Document contents

Indicator	Value
Keywords Plus (ID)	1238
Author's Keywords (DE)	2366

Table 4. Authors

Indicator	Value
Authors	1791
Author Appearances	2729
Authors of single-authored documents	68
Authors of multi-authored documents	1723

Table 5. Authors collaboration

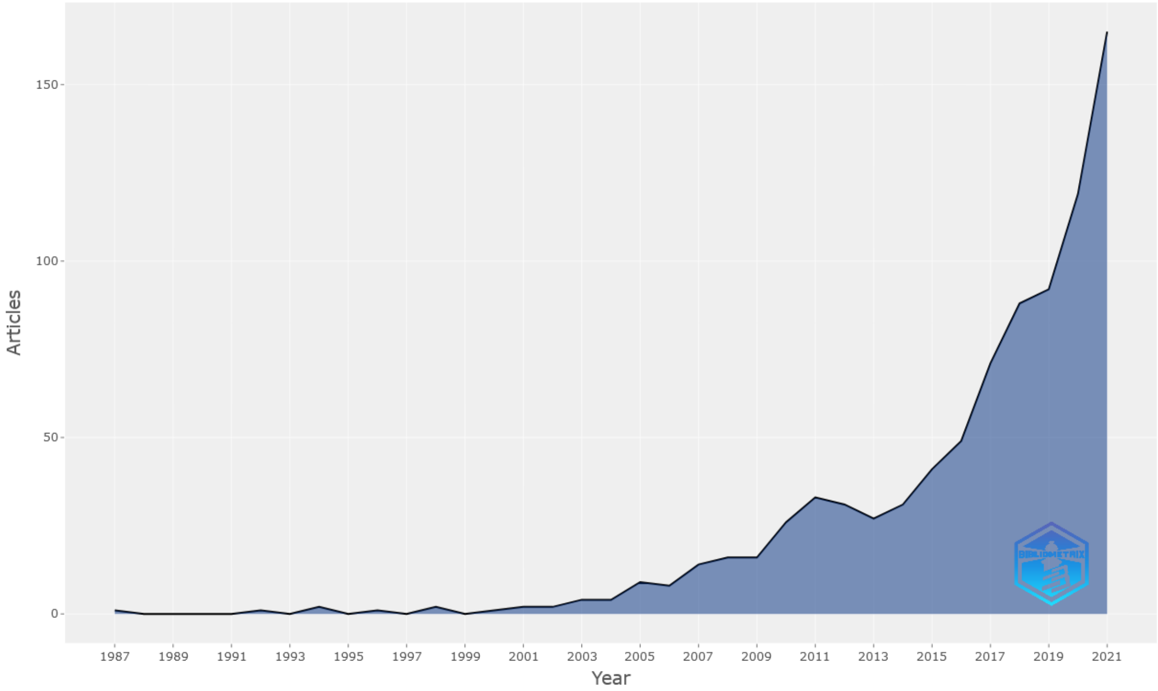
Indicator	Value
Single-authored documents	97
Documents per Author	0.485
Authors per Document	2.06
Co-Authors per Documents	3.14
Collaboration Index	2.23

1. Introduction to Grey Systems Research

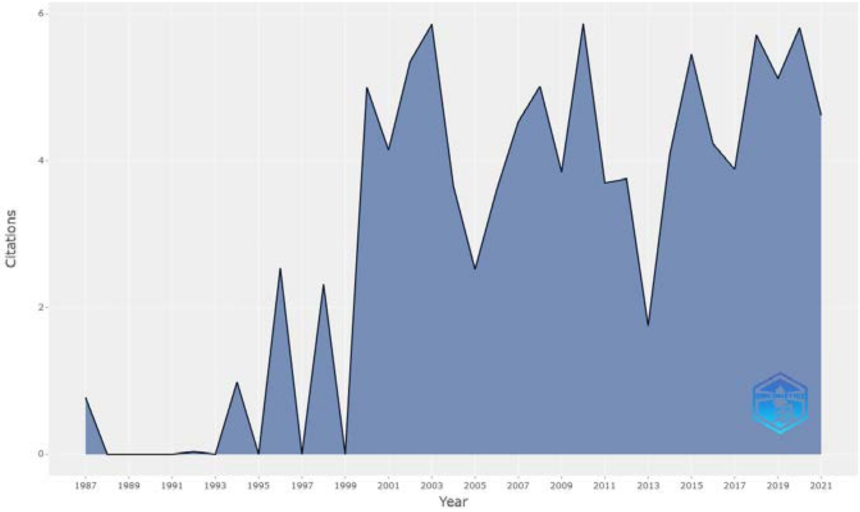
Development History

Grey Systems Theory Evolution

Annual Scientific Production



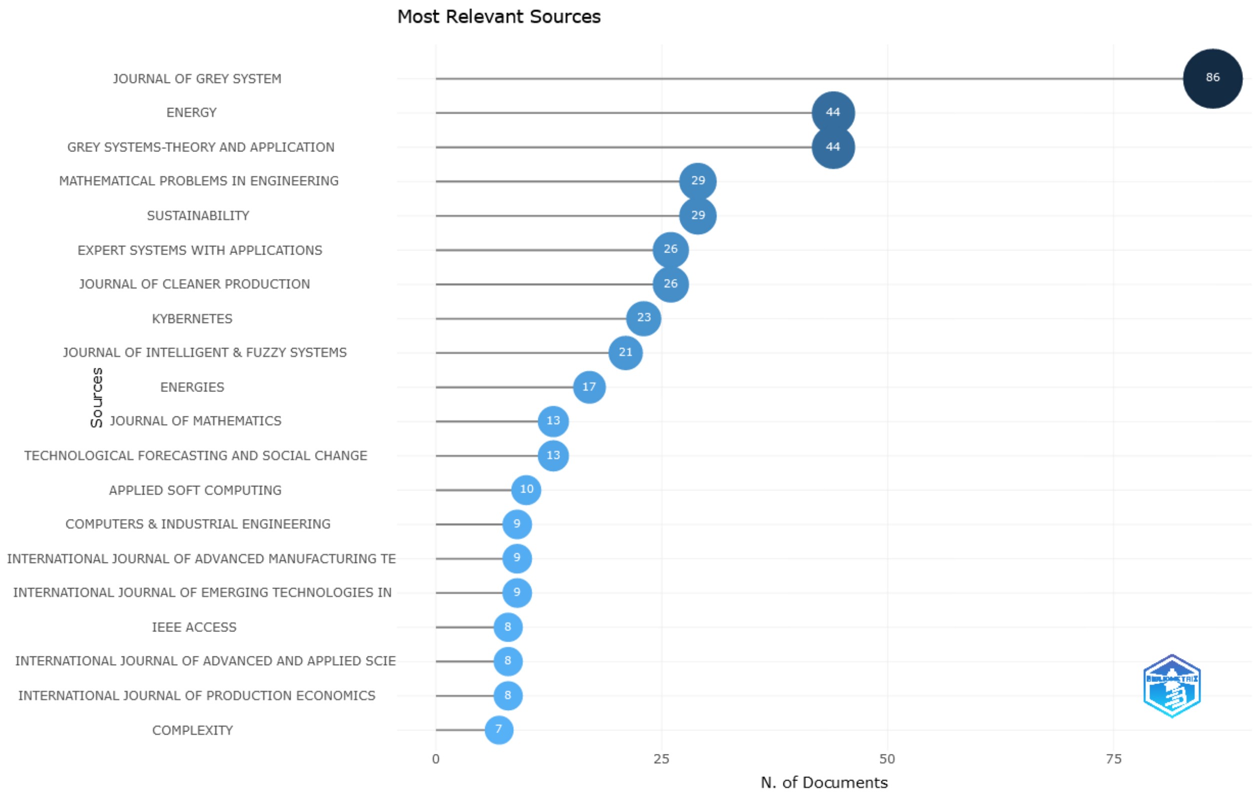
Average Article Citations per Year

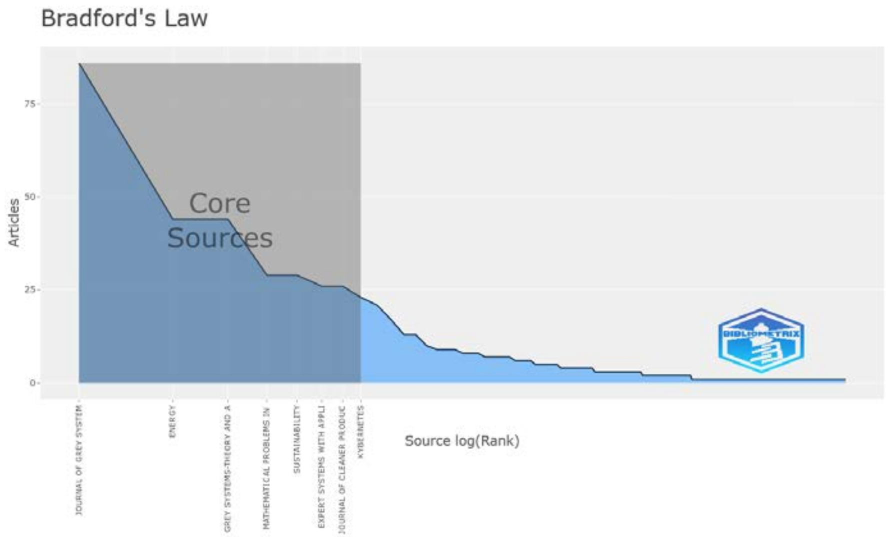


1. Introduction to Grey Systems Research

Development History

Grey Systems Theory Evolution

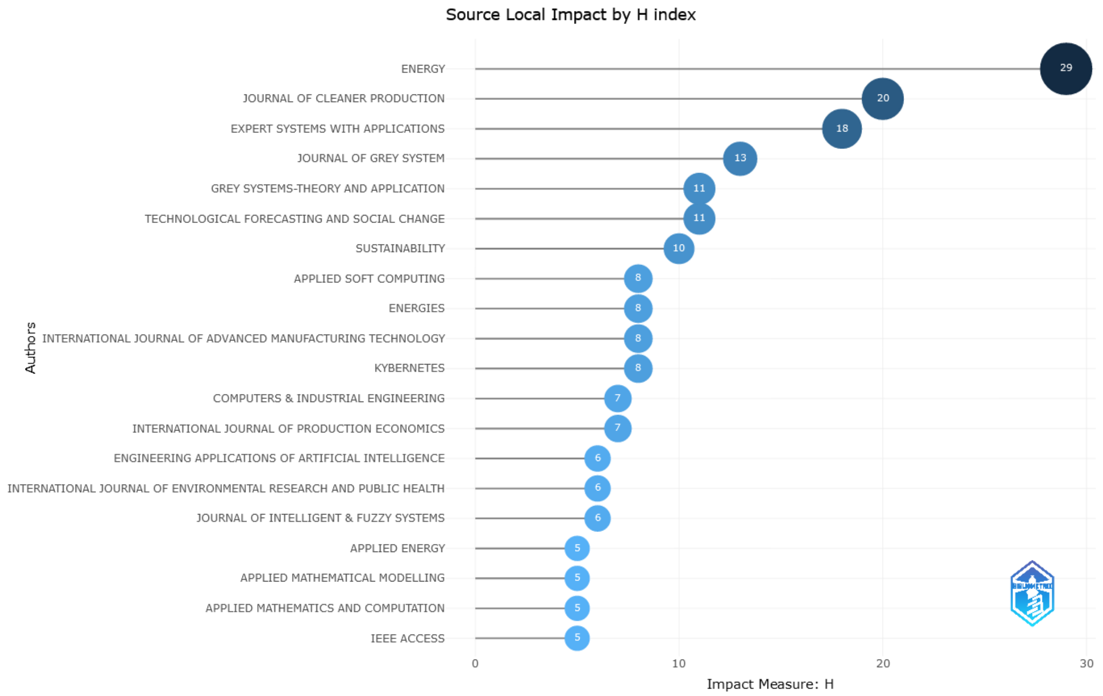


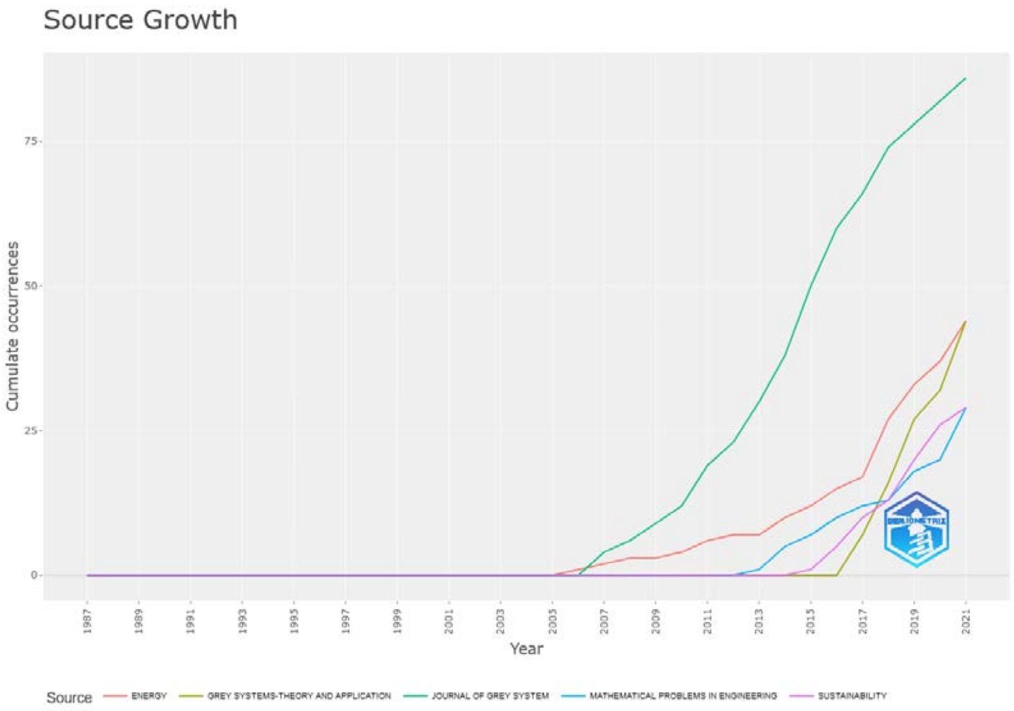


1. Introduction to Grey Systems Research

Development History

Grey Systems Theory Evolution

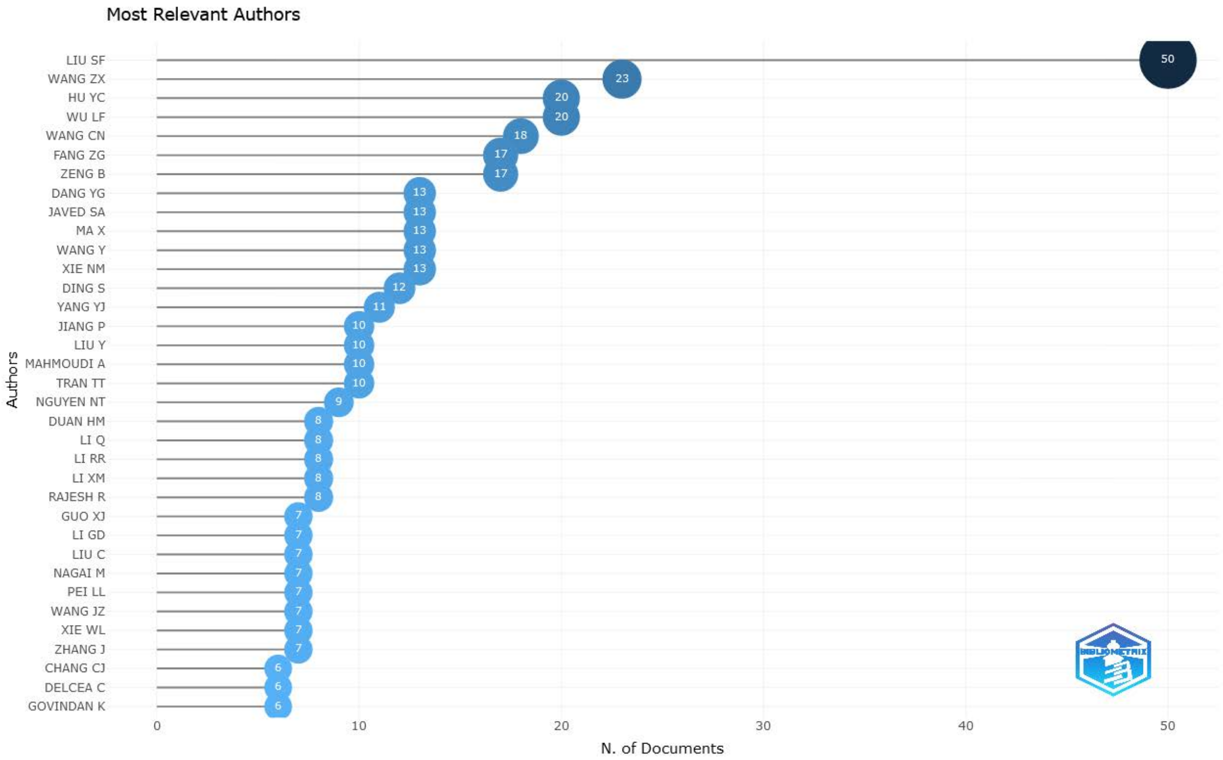


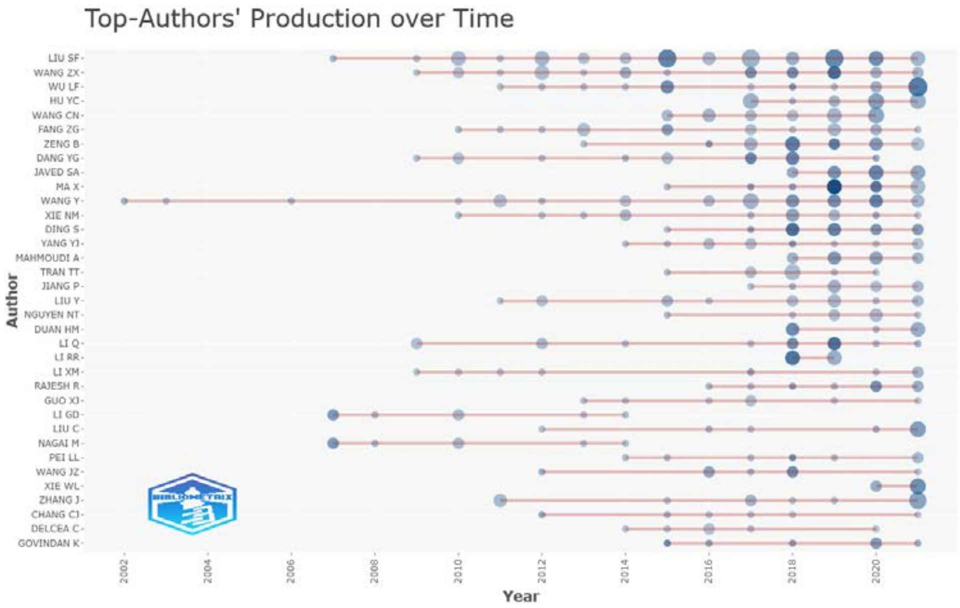


1. Introduction to Grey Systems Research

Development History

Grey Systems Theory Evolution

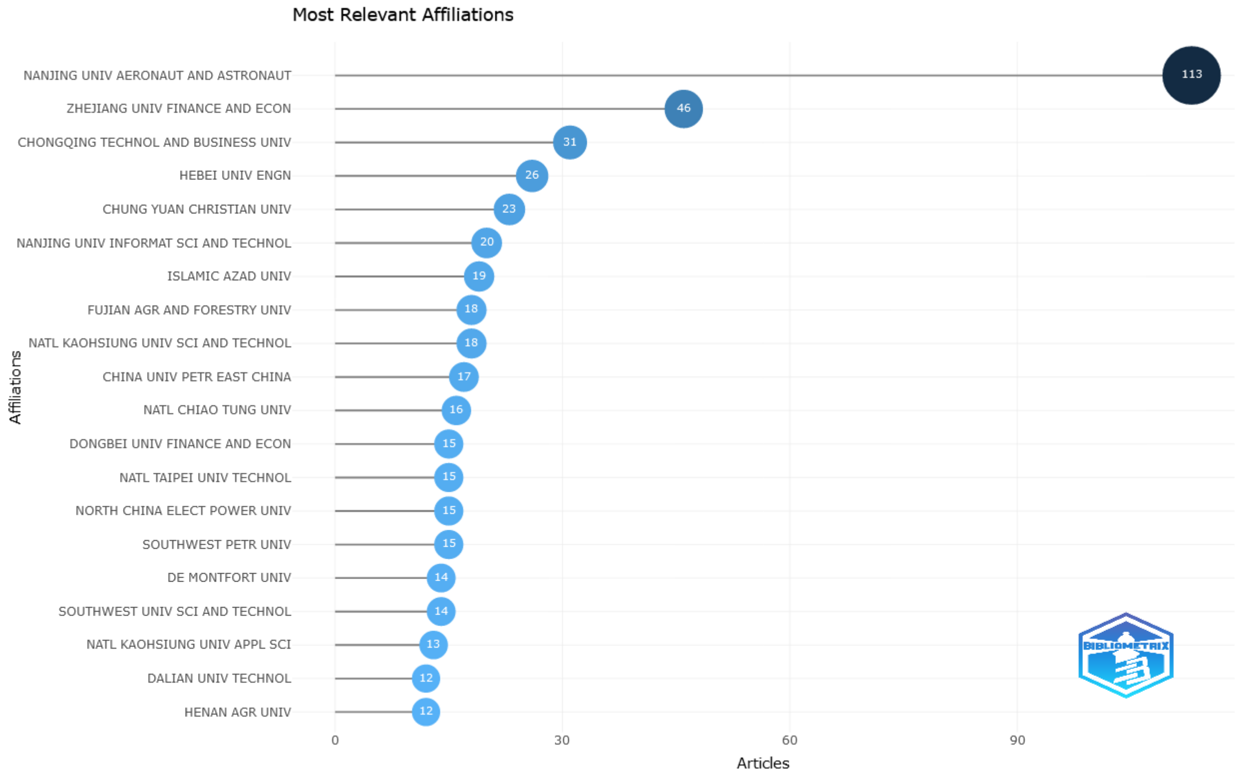


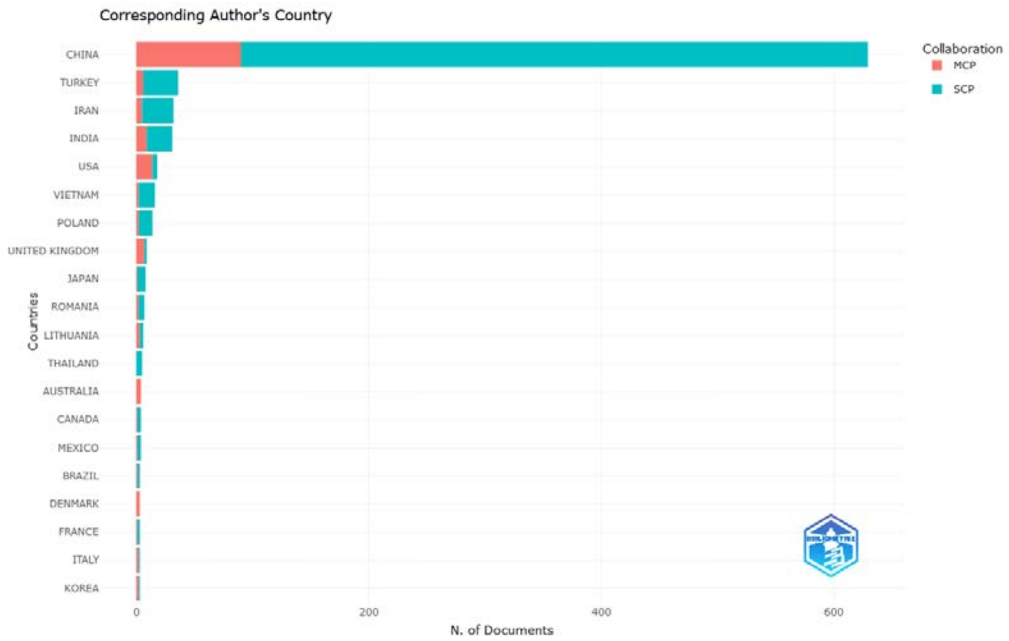


1. Introduction to Grey Systems Research

Development History

Grey Systems Theory Evolution



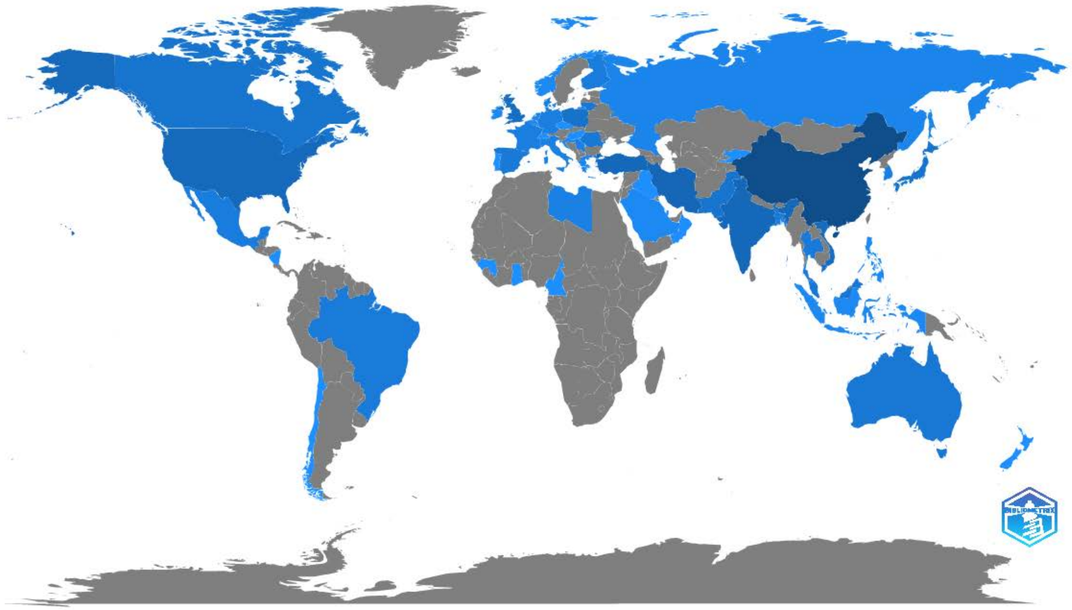


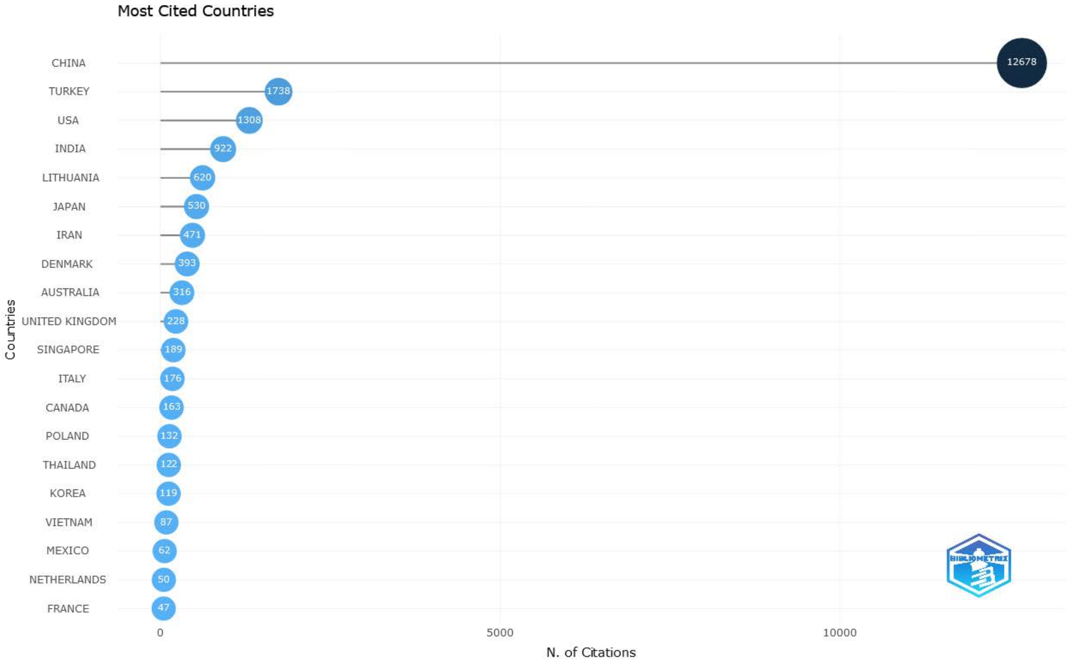
1. Introduction to Grey Systems Research

Development History

Grey Systems Theory Evolution

Country Scientific Production

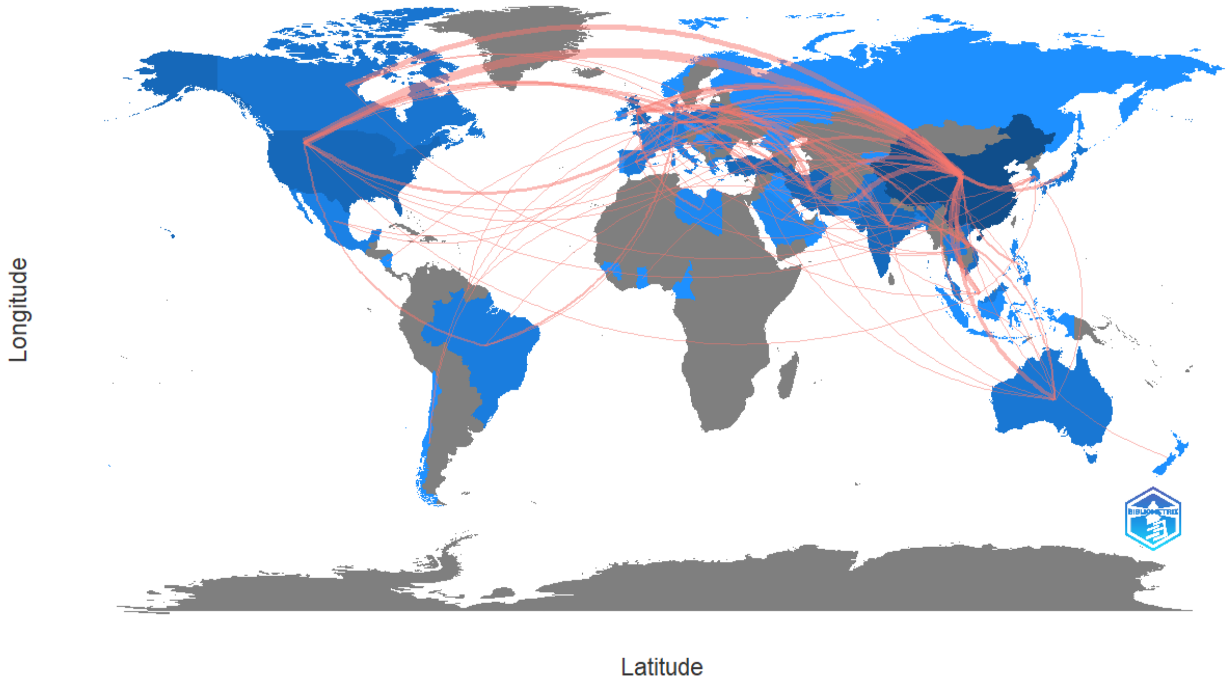


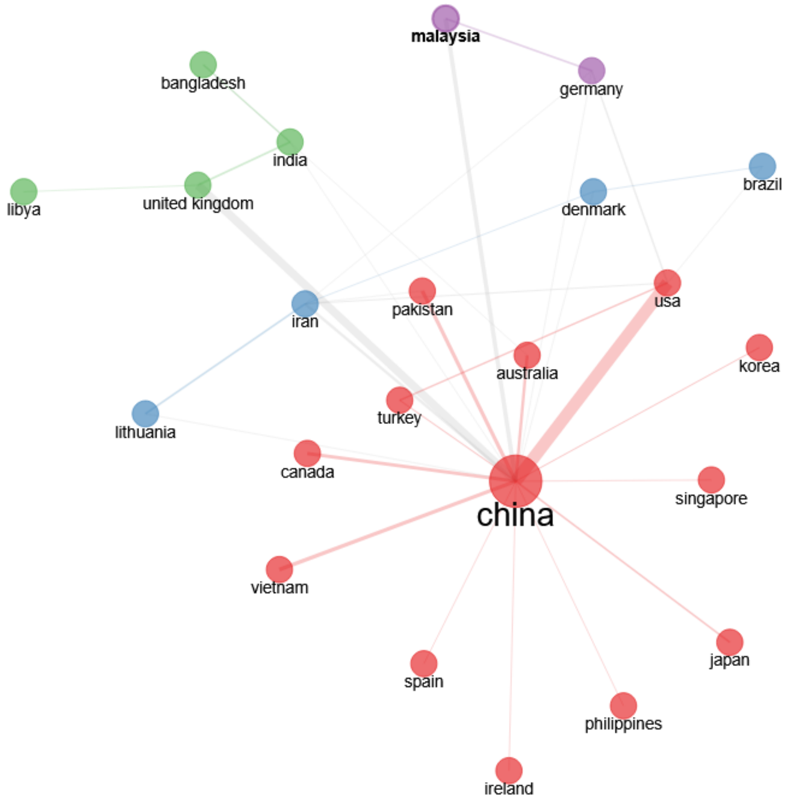


Development History

Grey Systems Theory Evolution

Country Collaboration Map





Grey Systems Theory Evolution

Table . No.	Paper (First Author, Year, Journal, Reference)	Number of authors	Country/ Countries	Digital object identifier (DOI)	Total citations (TC)	Total citations per year (TCY)	Normalized TC (NTC)
1.	Kayacan E, 2010, Expert Systems with Applications, [70]	3	Turkey, Canada	10.1016/j.eswa.2009.07.064	513	39.46	7.29
2.	Bai C, 2010, International Journal of Production Economics, [71]	2	China, USA	10.1016/j.ijpe.2009.11.023	506	38.92	7.19
3.	Pao HT, 2011, Energy, [72]	2	Taiwan	10.1016/j.energy.2011.01.032	317	26.41	7.80
4.	Hashemi SH, 2015, Journal of Production Economics, [73]	4	Iran, USA, Germany	10.1016/j.ijpe.2014.09.027	302	37.75	7.91
5.	Akay D, 2007, Energy, [74]	2	Turkey	10.1016/j.energy.2006.11.014	297	18.56	4.37
6.	Li GD, 2007, Mathematical and Computer Modelling, [75]	3	Japan	10.1016/j.mcm.2006.11.021	269	16.81	3.96
7.	Tseng ML, 2009, Expert Systems with Applications, [76]	1	Taiwan	10.1016/j.eswa.2008.09.011	262	18.71	5.24
8.	Pao HT, 2012, Energy, [77]	3	Taiwan, China	10.1016/j.energy.2012.01.037	259	23.54	6.90
9.	Kumar U, 2010, Energy, [78]	3	China, Denmark	10.1016/j.energy.2009.12.021	252	19.38	3.58
10.	Xia XQ, 2015, Journal of Cleaner Production, [79]	3	China, Denmark	10.1016/j.jclepro.2014.09.044	221	27.62	5.79

Grey Systems Theory Evolution

No.	Paper (First Author, Year, Journal, Reference)	Title	Grey systems theory main elements	Data	Purpose	Theoretical/ Practical approach	Hybrid approach / Theories considered
1.	Kayacan E, 2010, Expert Systistems with Applications, [70]	Grey system theory-based models in time series prediction	Grey prediction models	US dollar to Euro parity dataset	To compare the efficiency of the grey models and modified grey models using Fourier series	Both	-
2.	Bai C, 2010, International Journal of Production Economics, [71]	Integrating sustainability into supplier selection with grey system and rough set methodologies	Grey numbers	Synthetic data	To expand the supplier selection methodology by adding sustainability attributes	Both	Rough set theory (RST)
3.	Pao HT, 2011, Energy, [72]	Modeling and forecasting the CO2 emissions, energy consumption, and economic growth in Brazil	Grey prediction models	Pollutant emissions, energy consumption, output data	To model, forecast, and analyze the connection between the output, energy consumption and CO2 emissions	Both	-
4.	Hashemi SH, 2015, Journal of Production Economics, [73]	An integrated green supplier selection approach with analytic network process and improved Grey relational analysis	Grey relational analysis	Manufacturing company data	To take into account the environmental issues when the selection of the suppliers is performed	Both	Analytic network process (ANP)
5.	Akay D, 2007, Energy, [74]	Grey prediction with rolling mechanism for electricity demand forecasting of Turkey	Grey prediction models	Electricity consumption data	To forecast the electricity demand	Both	-

Grey Systems Theory Evolution

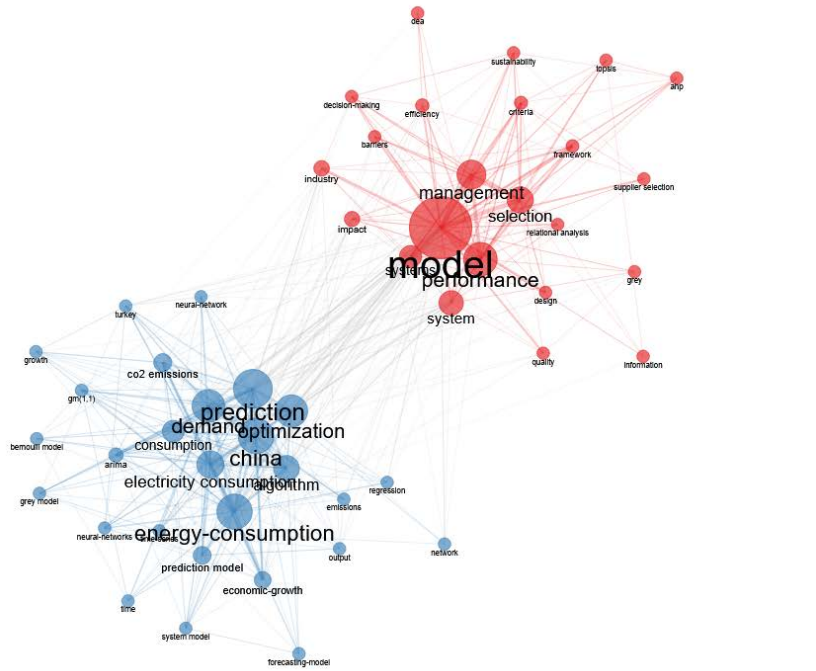
No.	Paper (First Author, Year, Journal, Reference)	Title	Grey systems theory main elements	Data	Purpose	Theoretical/ Practical approach	Hybrid approach / Theories considered
6.	Li GD, 2007, Mathematical and Computer Modelling, [75]	A grey-based decision-making approach to the supplier selection problem	Grey numbers	Synthetic data	To select the best supplied in a supplier selection problem	Both	-
7.	Tseng ML, 2009, Expert Systystems with Applications, [76]	A causal and effect decision making model of service quality expectation using grey-fuzzy DEMATEL approach	Grey numbers	Questionnaire	To rank the criteria in customer expectations	Both	Fuzzy
8.	Pao HT, 2012, Energy, [77]	Forecasting of CO2 emissions, energy consumption and economic growth in China using an improved grey model	Grey prediction models	Pollutant emissions, energy consumption, output data	To model, forecast, and analyze the connection between the output, energy consumption and CO2 emissions	Both	-
9.	Kumar U, 2010, Energy, [78]	Time series models (Grey-Markov, Grey Model with rolling mechanism and singular spectrum analysis) to forecast energy consumption in India	Grey prediction models	Conventional energy data: crude-petroleum consumption, coal, electricity consumption, natural gas consumption	To forecast the consumption of conventional energy	Both	-
10.	Xia XQ, 2015, Journal of Cleaner Production, [79]	Analyzing internal barriers for automotive parts remanufacturers in China using grey-DEMATEL approach	Grey numbers	Questionnaire	To analyze the internal barriers met by remanufacturers from the automotive parts industry and evaluate the causal barrieres encountered by the remanufacturers	Both	-

1. Introduction to Grey Systems Research

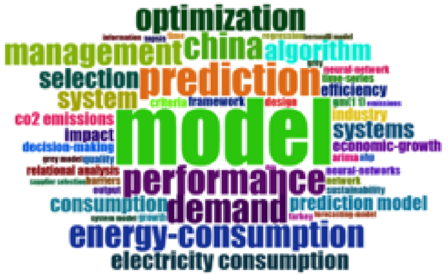
Development History

Grey Systems Theory Evolution

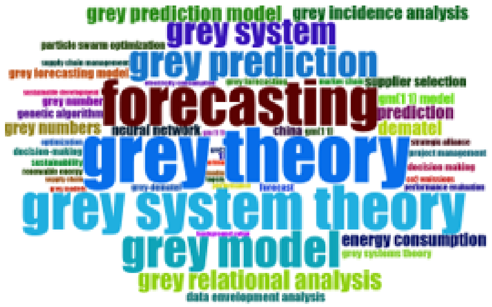
Words	Occurrences
model	148
prediction	74
performance	66
demand	63
china	60
energy-consumption	60
optimization	58
management	57
system	47
algorithm	46



Grey Systems Theory Evolution

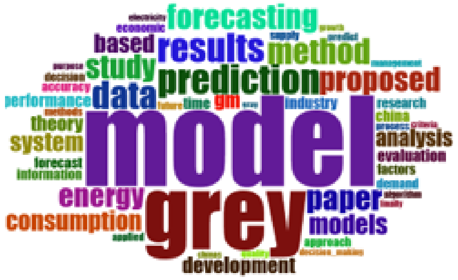


(A) Top 50 words based on keywords plus



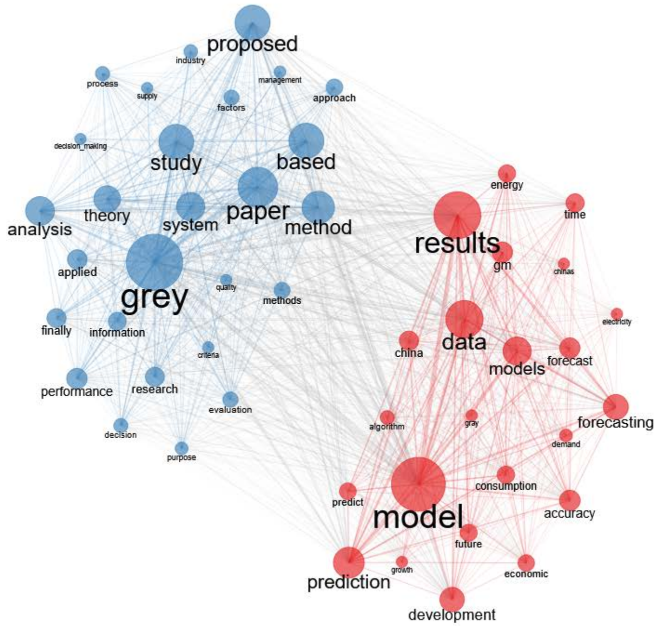
(B) Top 50 words based on authors' key-words

(C) Top 50 words based on title



(D) Top 50 words based on abstract

Edit



Grey Systems Theory Evolution

Table 10. Top 10 most frequent bigrams in abstracts and titles

Bigrams in abstracts	Occurrences	Bigrams in titles	Occurrences
grey model	352	grey model	109
energy consumption	242	grey prediction	56
grey prediction	206	energy consumption	48
prediction model	203	prediction model	43
grey system	173	grey system	32
proposed model	155	grey theory	28
system theory	146	system theory	25
gm model	137	neural network	24
supply chain	120	grey incidence	23
grey theory	116	supply chain	23

Table 11. Top 10 most frequent trigrams in abstracts and titles

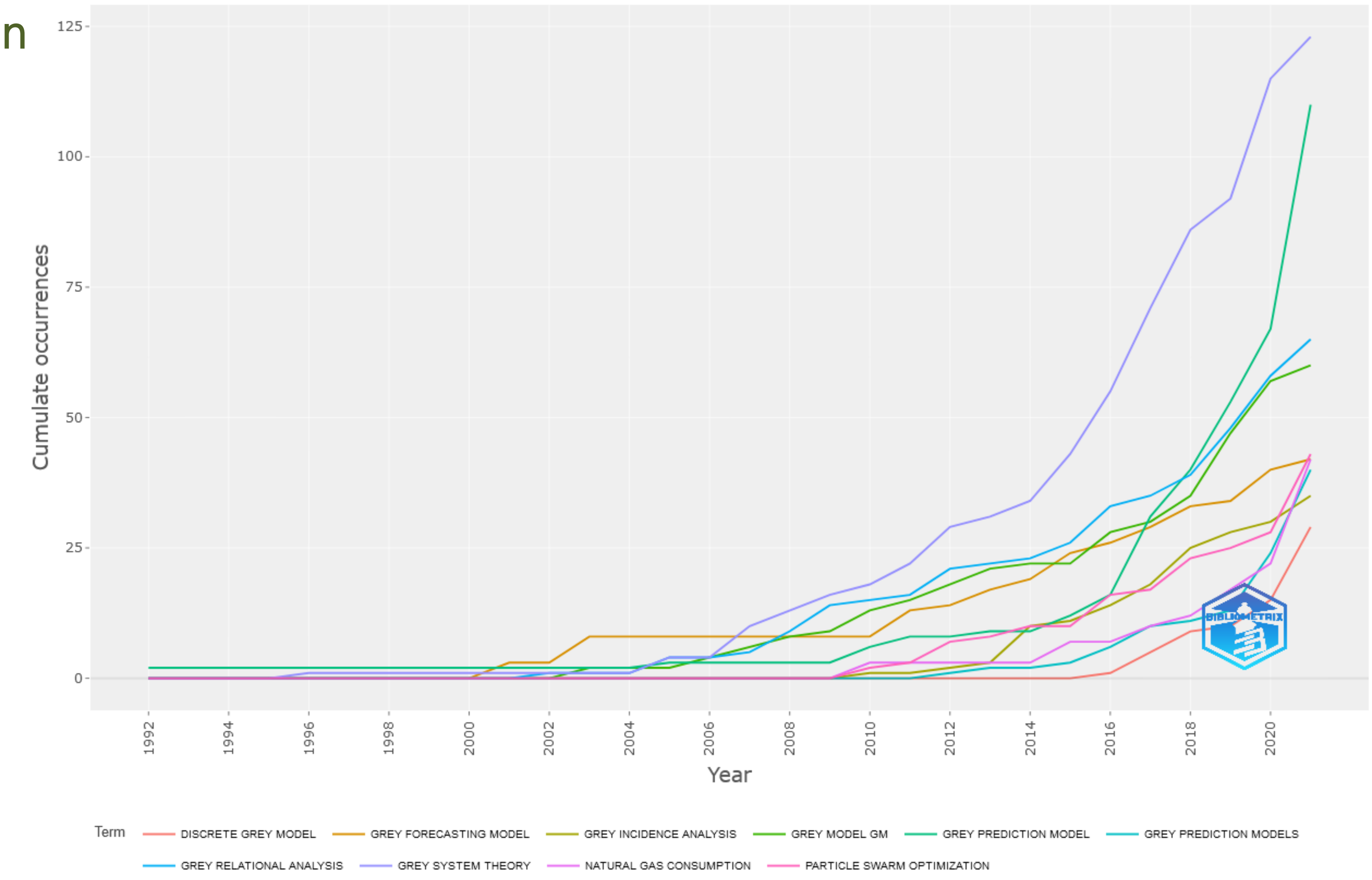
Trigrams in abstracts	Occurrences	Trigrams in titles	Occurrences
grey system theory	126	grey prediction model	28
grey prediction model	110	grey system theory	23
grey relational analysis	65	grey relational analysis	16
grey model gm	60	grey incidence analysis	13
particle swarm optimization	43	grey prediction models	11
grey forecasting model	42	natural gas consumption	11
natural gas consumption	42	grey bernoulli model	10
grey prediction models	41	grey forecasting model	10
grey incidence analysis	35	discrete grey model	8
discrete grey model	31	fractional grey model	8

1. Introduction to Grey Systems Research

Development History

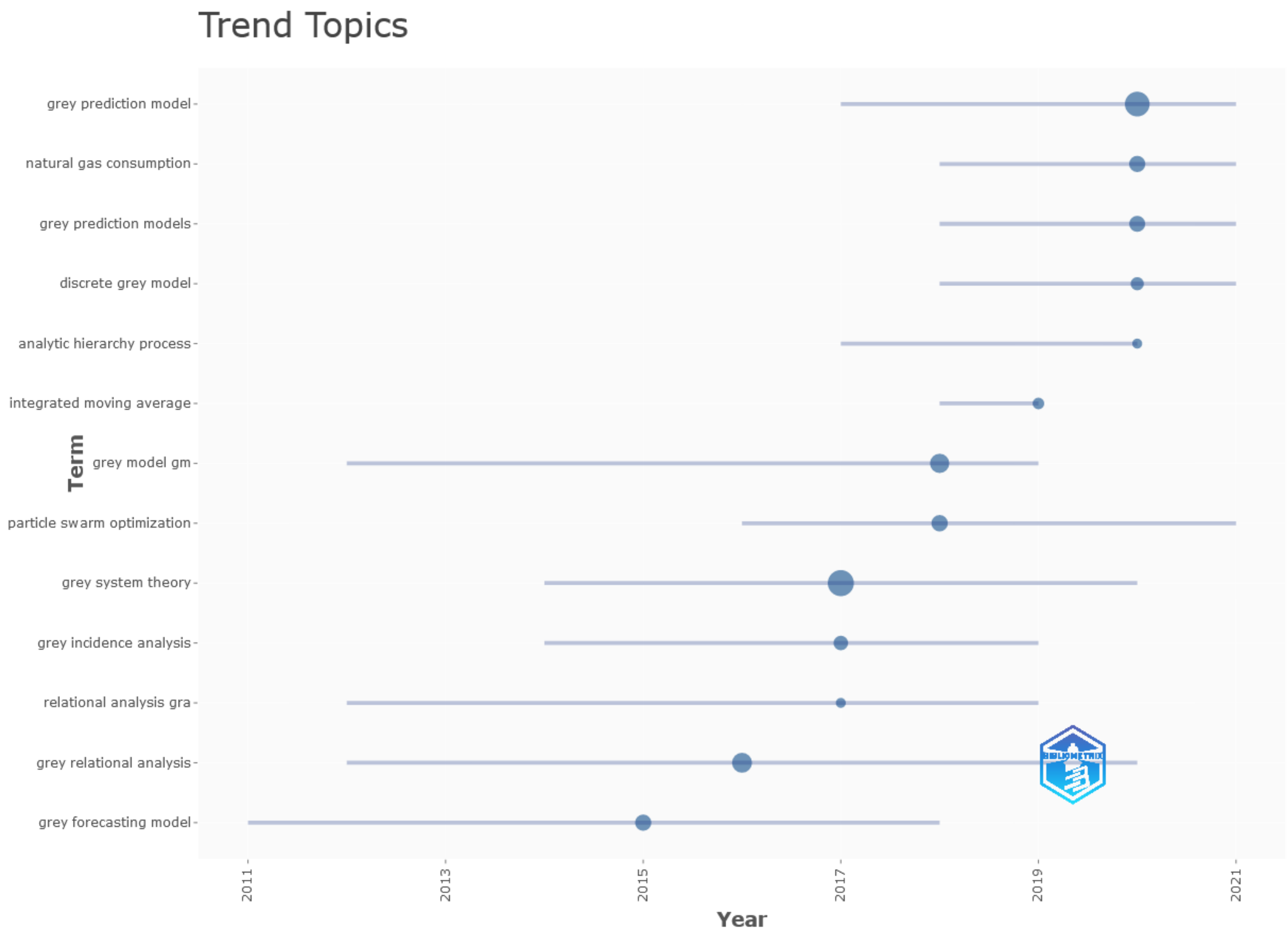
Grey Systems Theory Evolution

Word Growth



Development History

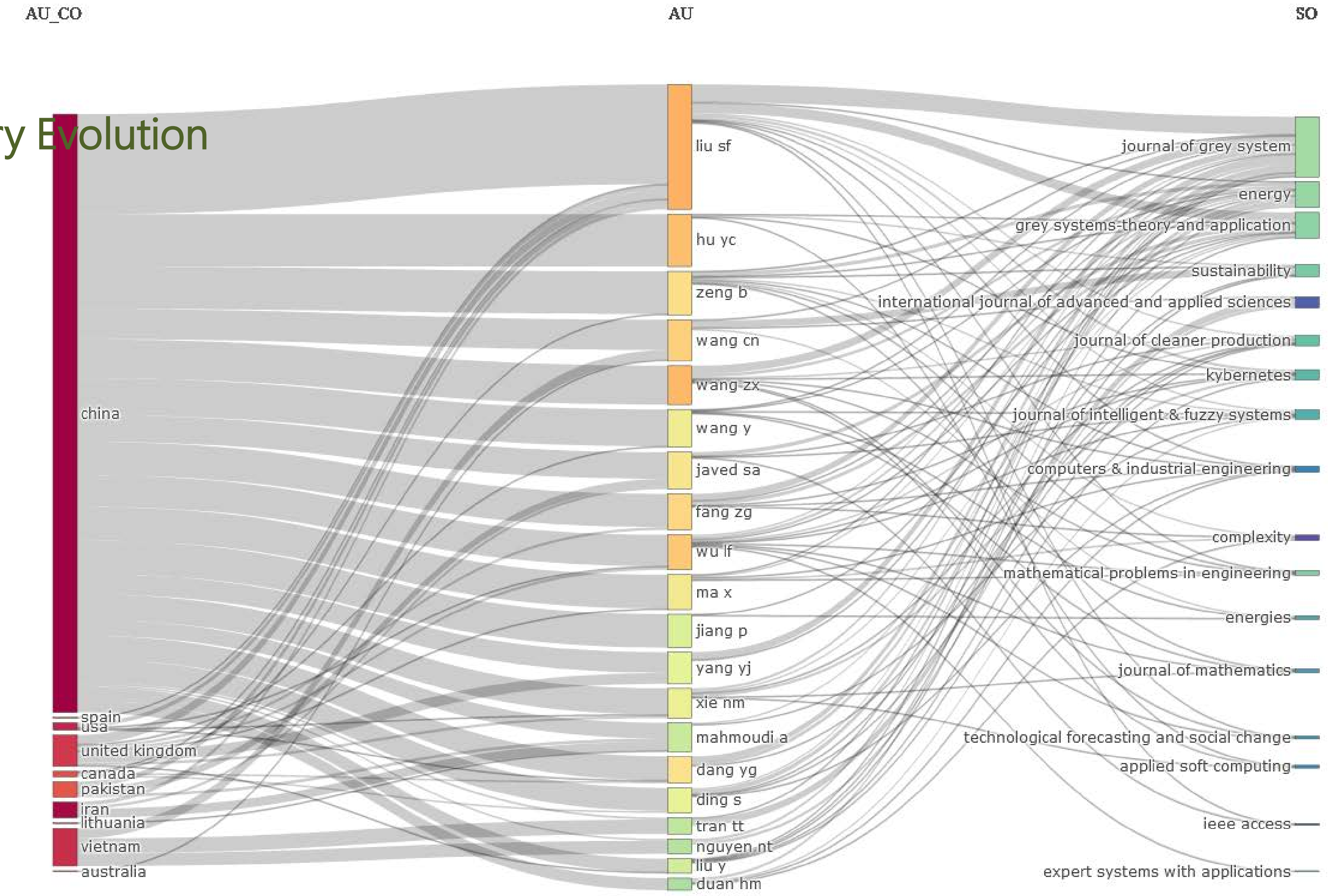
Grey Systems Theory Evolution



1. Introduction to Grey Systems Research

Development History

Grey Systems Theory Evolution



Elementary Concepts of Grey Systems

In the **theory of control**, scholars often make use of **colors** to describe the **degree of clearness of available information.**

For instance, **Ashby** refers to objects with unknown internal information as **black boxes**. This terminology has been widely accepted in the scientific community.

As another example, as a society moves toward democracy, citizens gradually demand **more information** regarding policies and the meanings of such policies. That is, citizens want to have an **increased degree of information transparency** (i.e. white information).

Elementary Concepts of Grey Systems

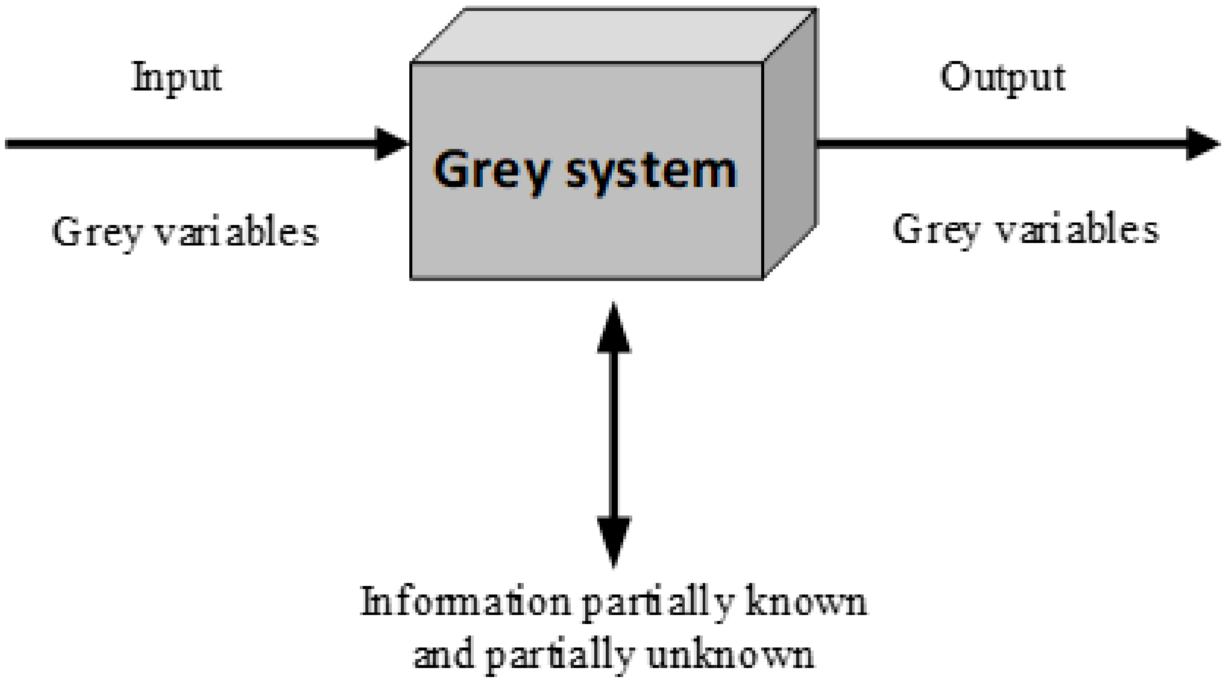
Thus, we use "**black**" to indicate unknown information, "**white**" to indicate completely known information and "**grey**" to convey partially known and partially unknown information.

Accordingly, **systems** with **completely known information** are regarded as **white**, while systems with **completely unknown information** are considered **black**, and systems with p**artially known information and partially unknown information** are seen as **grey**.

In this context, **incompleteness** in information is the fundamental meaning of "grey."

However, the meaning of "grey" can be expanded or stretched from **different angles** and in varied situations as presented in the following table:

Elementary Concepts of Grey Systems



Elementary Concepts of Grey Systems

Table 1: Extensions of the concept of "grey"

Situation/concept	Black	Grey	White
Information	Unknown	Incomplete	Completely known
Appearance	Dark	Blurred	Clear
Processes	New	Changing	Old
Properties	Chaotic	Multivariate	Order
Methods	Negation	Change for the better	Confirmation
Attitude	Letting go	Tolerant	Rigorous
Outcomes	No solution	Multi-solutions	Unique solution

Elementary Concepts of Grey Systems

The **research objects** of grey systems theory consist of **uncertain systems** that are known only partially with **small samples** and **poor information**.

The theory focuses on the generation and **excavation of partially known information** through **grey sequence operators of possibility functions** to enable an accurate description and understanding of the material world.

Characteristics of Uncertain Systems

The fundamental characteristic of uncertain systems is the **incompleteness** and **inadequacy** of their information.

Due to the **dynamics of system evolution**, the biological limitations of the human sensory system, as well as the constraints of relevant economic conditions and technological availabilities, **uncertain systems exist commonly**.

Incomplete Information

Incompleteness in information is one of the **fundamental characteristics** of uncertain systems.

The most **common situations** involving incomplete system information include cases where:

(1) Information about **system elements** (parameters) is incomplete;
(2) Information on the **structure** of the system is incomplete;
(3) Information about the **boundaries** of the system is incomplete; and
(4) Information on the system's **behaviors** is incomplete.

Incomplete Information

Incomplete information is a **common phenomenon** in our social, economic, and scientific research activities.

For instance, in **agricultural production**, even if we have exact information regarding plantation, seeds, fertilizers, and irrigation, uncertainties in areas such as labor quality, natural environment characteristics, weather conditions, and the commodity markets make it extremely difficult to **precisely predict the production output** and consequent economic value of agricultural fields.

For **biological prevention systems**, even if we know the relationship between insects and their natural enemies, it is still really difficult to **achieve the expected prevention effects** due to uncertainty regarding the relationships between insects and their baits, insects' natural enemies and their baits, and a specific kind of natural enemy with another kind of natural enemy.

Incomplete Information

As for the **adjustment and reform of pricing systems**, it is often difficult for policy makers to take actions because of the **lack of information regarding price elasticity of demand** and how price changes on a certain commodity would affect the prices of other commodities.

In **security markets**, even the brightest market analysts **cannot be assured of winning constantly** due to their inability to correctly predict economic policy and interest rate changes, management changes at various companies, the direction of political changes, investors' behavioral changes in international markets, and the effects of price changes in one block of commodities on another.

As for the **general economic system**, because there are no clear relationships between the "inside" and the "outside" of the system, and between the system itself and its environment, and because the boundaries between the inside and the outside of the system are difficult to define, it is also **difficult to analyze the effects of economic input on economic output**.

Incomplete Information

Incompleteness in available information **is absolute**, while **completeness** in information is **relative**.

Humans employ their **limited cognitive ability** to observe the infinite universe in order to try and obtain **complete information**.

However, it is impossible for us to do so. In fact, the concept of **large samples in statistics** represents **the degree of tolerance man has to incompleteness**.

In theory, when a sample contains at least 30 objects, it is considered “large.” However, in some situations, even when a sample contains thousands or several tens of thousands of objects, the true statistical laws of a given system still cannot be successfully uncovered.

Inaccuracies in Data

Another fundamental characteristic of **uncertain systems** is naturally occurring **inaccuracy** in available data.

In **grey systems theory**, the meanings of **uncertain and inaccurate** are roughly the same.

Both terms stand for **errors or deviations** from actual data values.

Based on the essence of how uncertainties are caused, inaccuracies can be categorized into three types:

- the conceptual,
- level, and
- prediction type inaccuracies.

Inaccuracies in Data – the Conceptual Type

Inaccuracies of the conceptual type emanate from the **expression of a certain event, object, concept, or wish.**

For instance, all such frequently used concepts as “large,” “small,” “many,” “few,” “high,” “low,” “fat,” “thin,” “good,” “bad,” “young,” and “beautiful” are **inaccurate** due to lack of clear definition.

It is very **difficult to use exact quantities** to express these concepts.

As a second example, suppose that a job seeker with an MBA degree wishes to get an annual salary offer of **no less than $150,000**, or that a manufacturing firm plans to control its rate of defective products to be **less than 0.1 %.** These are all cases of conceptual type inaccuracies.

Inaccuracies in Data – the Level Type

This kind of data inaccuracy is caused by **a change** at the level of research or observation.

This means that **the available data might be accurate** when seen at the level of the system of concern, that is, the **macroscopic level,** or at the level of the whole, that is, the cognitive conceptual level.

However, when data are **seen at a lower level**, that is, a microscopic level, or at a partial localized level of the system, they generally **become inaccurate**.

For example, the height of a person can be measured accurately to the unit of centimeters or millimeters. However, if the measurement has to be accurate to the level of one ten-thousandth micrometers, the former accurate reading will become extremely inaccurate.

Inaccuracies in Data – the Prediction or Estimation Type

Because it is **difficult** to have **complete understanding** of the laws of evolution, any prediction of the future tends to be inaccurate.

For instance:

- it is estimated that **two years from now, the GDP of a certain country will surpass** \$10 billion dollars;
- it is estimated that a **certain bank will attract savings** from individual residents of between \$70 thousand and \$90 thousand for the year 2017;
- it is predicted that in the coming years the **temperature** in Leicester, UK, during the month of June will not go beyond 30 °C, and so on.

Comparison among the Four Methods of Uncertainty Research

Uncertainty research	Grey system	Prob. statistics	Fuzzy math	Rough set
Research objects	Poor information	Stochastics	Cognitive	Boundary
Basic set	Grey number set	Cantor set	Fuzzy set	Approximate set
Describe method	Possibility func.	Density func.	Membership func.	Upper, lower appr.
Procedure	Sequence operator	Frequency	Cut set	Dividing
Data requirement	Any distribution	Known distribution	Known membership	Equivalent rel.
Emphasis	Intension	Intension	Extension	Intension
Objective	Law of reality	Historical law	Cognitive expression	Approx. approaching
Characteristics	Small data	Large sample	Depend on experience	Information form

Fundamental Principles of Grey Systems

Principle of Informational Differences "Difference" implies the existence of information. Each piece of information must carry some kind of "difference".

Principle of Non-Uniqueness The solution to any problem with incomplete and indeterminate information is not unique.

Principle of Minimal Information One characteristic of grey system theory is that it makes the most and best use of the "minimal amount of available information."

Principle of Recognition Base Information is the foundation on which people recognize and understand (nature).

Fundamental Principles of Grey Systems

Principle of New Information Priority The function of new pieces of information is greater than that of old pieces of information.

Principle of Absolute Greyness "Incompleteness" of information is absolute. Incompleteness and non-determinism of information have generality.

- Completeness of information is **relative** and temporary. It is the moment when the original non-determinism has just disappeared, and new non-determinism is about to emerge. Human recognition and understanding of the objective world have been improved over time through continued supplementation of information. With endless supply of information, man's recognition and understanding of the world also become endless. That is, **greyness of information is absolute** and will never disappear.

Next

2. Grey Numbers and their Mathematics

Grey Numbers Definition and Representation

A **grey system** is described with:

- grey numbers,
- grey sequences,
- grey equations, or
- matrices.

Here, grey numbers are the elementary "atoms" or "cells", and their **exact values are unknown**. In applications, a grey number stands for an **indeterminate number** that takes its possible value **within an interval or a general set of numbers**.

A grey number is generally represented using the symbol " $\otimes$ "
There are several types of grey numbers, as discussed below.

Grey Numbers Definition and Representation

(1) Grey numbers with only a lower bound: This kind of grey number $\otimes$ is represented as $\otimes \in [\underline{a}, \infty)$ or $\otimes(\underline{a})$, where $(\underline{a})$ stands for the definite, known lower bound of the grey number $\otimes$. The interval $[\underline{a}, \infty)$ is referred to as the field of $\otimes$.

For example, the weight of a celestial body which is far away from the Earth is a grey number containing only a lower bound, because the weight of the celestial body must be greater than zero. However, the exact value of the weight cannot be obtained through normal means. If we use the symbol $\otimes$ to represent the weight of the celestial body, we then have that $\otimes \in [0, \infty)$.

Grey Numbers Definition and Representation

(2) Grey numbers with only an upper bound: This kind of grey number $\otimes$ is written as $\otimes \in (-\infty, \bar{a}]$ or $\otimes(\bar{a})$, where $\bar{a}$ stands for the definite, known upper bound of $\otimes$.

A grey number containing only an upper bound is a grey number with a negative value, but its absolute value is infinitely great. For example, the opposite number of the weight of the celestial body mentioned above is a grey number with only an upper bound.

Grey Numbers Definition and Representation

(3) Interval grey numbers: This kind of grey number $\otimes$ has both a lower bound $\underline{a}$ and an upper bound $\bar{a}$, written $\otimes \in [\underline{a}, \bar{a}]$.

For example, for an investment opportunity, there always exists an upper limit representing the maximum amount of money that can be mobilized. For an electrical equipment, there must be a maximum critical value for the equipment to function normally. The critical value could be for a maximum voltage or for a maximum amount of current allowed to be applied to the equipment. At the same time, the values of investment, voltage, and current are all greater than zero. Therefore, the amount of dollars that can be used for a specific investment opportunity, and the voltage and the current requirements for the electrical equipment are all examples of interval grey numbers.

Grey Numbers Definition and Representation

(4) Continuous and discrete grey numbers: This kind of grey number takes only a finite number or a countable number of potential values and is known as discrete. If a grey number can potentially take any value within an interval, then it is known as continuous.

For example, if a person's age is between 30 and 35, his or her age could be one of the values 30, 31, 32, 33, 34, 35. Thus, age is a discrete grey number. As for a person's height and weight, they are continuous grey numbers.

Grey Numbers Definition and Representation

(5) Black and white numbers: Black numbers are represented as $\otimes \in (-\infty, +\infty)$; that is, when $\otimes$ has neither an upper nor a lower bound, then $\otimes$ is known as a black number. When $\otimes \in [\underline{a}, \bar{a}]$ and $\underline{a} = \bar{a}$, $\otimes$ is known as a white number.

For the sake of parsimony, in our discussion we treat black and white numbers as special grey numbers.

Grey Numbers Definition and Representation

(6) Essential and non-essential grey numbers: The former stands for a grey number that temporarily cannot be represented by a white number; the latter entails a grey number that can be represented by a white number obtained either through experience or through a certain method. The definite white number is referred to as the whitenization (value) of the grey number, denoted $\tilde{\otimes}$. Also, we use $\otimes(a)$ to represent grey number(s) with a as its whitenization.

A grey number is an uncertain number with its value in a specific range. The range can be regarded as a cover of the grey number. Therefore, an interval grey number $\otimes \in [\underline{a}, \bar{a}], \underline{a} < \bar{a}$ is very different from an interval number $[\underline{a}, \bar{a}], \underline{a} < \bar{a}$. An interval grey number $\otimes \in [\underline{a}, \bar{a}], \underline{a} < \bar{a}$ is only one value in interval $[\underline{a}, \bar{a}], \underline{a} < \bar{a}$. However, an interval number $[\underline{a}, \bar{a}], \underline{a} < \bar{a}$ is the whole interval $[\underline{a}, \bar{a}], \underline{a} < \bar{a}$.

Operations of Interval Grey Numbers

In what follows, let us look at the operations of interval grey numbers. Given grey numbers $\otimes_1 \in [a, b]$, $a<b$, and $\otimes_2 \in [c, d]$, $c<d$, let us use $*$ to represent an operation between $\otimes_1$ and $\otimes_2$. If $\otimes_3 = \otimes_1 * \otimes_2$, then $\otimes_3$ should also be an interval grey number satisfying $\otimes_3 \in [e, f]$, $e<f$, and for any $\tilde{\otimes}_1$ and $\tilde{\otimes}_2$, $\tilde{\otimes}_1 * \tilde{\otimes}_2 \in [e, f]$. The operation rules of interval grey numbers are discussed below

Operations of Interval Grey Numbers

Rule 1 (*Additive operation*). Assume that $\otimes_1 \in [a,b]$, $a<b$; $\otimes_2 \in [c,d], c<d$, then the following equation is called the sum of $\otimes_1$ and $\otimes_2$:

$$\otimes_1 + \otimes_2 \in [a+c, b+d]$$

Example 1 Assume that $\otimes_1 \in [3,4]$, $\otimes_2 \in [5,8]$, then $\otimes_1 + \otimes_2 \in [8,12]$.

Rule 2 (*Additive inverse*). Assume that $\otimes[a,b]$, $a<b$, then the additive inverse of $\otimes$ is given by:

$$-\otimes \in [-b,-a]$$

Example 2 Assume that $\otimes \in [3,4]$, then $-\otimes \in [-4,-3]$.

Operations of Interval Grey Numbers

Rule 3 (*Subtraction operation*). Assume that $\otimes_1 \in [a,b]$, $a < b$; $\otimes_2 \in [c,d]$, $c<d$, then the following is called the deviation $\otimes_1$ minus $\otimes_2$:

$$\otimes_1 - \otimes_2 = \otimes_1 + (-\otimes_2) \in [a-d, b-c]$$

Example 3 Assume that $\otimes_1 \in [3,4]$, $\otimes_2 \in [1,2]$, then:

$$\otimes_1 - \otimes_2 \in [3-2, 4-1] = [1,3], \otimes_2 - \otimes_1 \in [1-4, 2-3] = [-3,-1].$$

Rule 4 (*Multiplication operation*). Assume that $\otimes_1 \in [a,b]$, $a<b$; $\otimes_2 \in [c,d]$, $c<d$ then the following equation is called the product of $\otimes_1$ and $\otimes_2$:

$$\otimes_1 \cdot \otimes_2 \in [\min\{ac, ad, bc, bd\}, \max\{ac, ad, bc, bd\}]$$

Example 4 Assume that $\otimes_1 \in [3,4]$, $\otimes_2 \in [5,10]$, then:

$$\otimes_1 \cdot \otimes_2 \in [\min\{15, 30, 20, 40\}, \max\{15, 30, 20, 40\}] = [15, 40].$$

Operations of Interval Grey Numbers

Rule 5 (*Reciprocal*). Assume that $\otimes \in [a, b]$, $a < b$, $a \neq 0, b \neq 0$, $ab > 0$, then the following equation is called the reciprocal of $\otimes$:

$$\otimes^{-1} \in \left[\frac{1}{b}, \frac{1}{a}\right]$$

Example 5 Assume that $\otimes \in [2, 4]$, then $\otimes^{-1} \in [0.25, 0.5]$.

Rule 6 (*Division*). Assume that $\otimes_1 \in [a, b]$, $a < b$; $\otimes_2 \in [c, d], c < d$, and $c \neq 0, d \neq 0, cd > 0$, then the following is called the quotient of $\otimes_1$ division by $\otimes_2$:

$$\otimes_1/\otimes_2 = \otimes_1 \times \otimes_2^{-1} \in \left[\min\left\{\frac{a}{c}, \frac{a}{d}, \frac{b}{c}, \frac{b}{d}\right\}, \max\left\{\frac{a}{c}, \frac{a}{d}, \frac{b}{c}, \frac{b}{d}\right\}\right]$$

Example 6 Assume that $\otimes_1 \in [3, 4], \otimes_2 \in [5, 10]$, then:

$$\otimes_1/\otimes_2 \in [\min\{\frac{3}{5}, \frac{3}{10}, \frac{4}{5}, \frac{4}{10}\}, \max\{\frac{3}{5}, \frac{3}{10}, \frac{4}{5}, \frac{4}{10}\}] = [0.3, 0.8].$$

Operations of Interval Grey Numbers

Rule 7 (*Scalar multiplication*). Let $\otimes \in [a, b]$, $a < b$, and k a positive real number, then the following is called the product of scalar k with grey number $\otimes$:

$$k \cdot \otimes \in [ka, kb]$$

Example 7 Assume that $\otimes \in [2, 4]$, and k=5, then $5 \times \otimes \in [10, 20]$.

Rule 8 (*Power*). Let $\otimes \in [a, b]$, $a < b$, k k a positive real number, then the following equation is called the kth power of the grey number $\otimes$:

$$\otimes^k \in [a^k, b^k]$$

Example 8 Assume that $\otimes \in [2, 4]$, and k = 5, then $\otimes^5 \in [32, 1024]$.

Comparing Grey Numbers

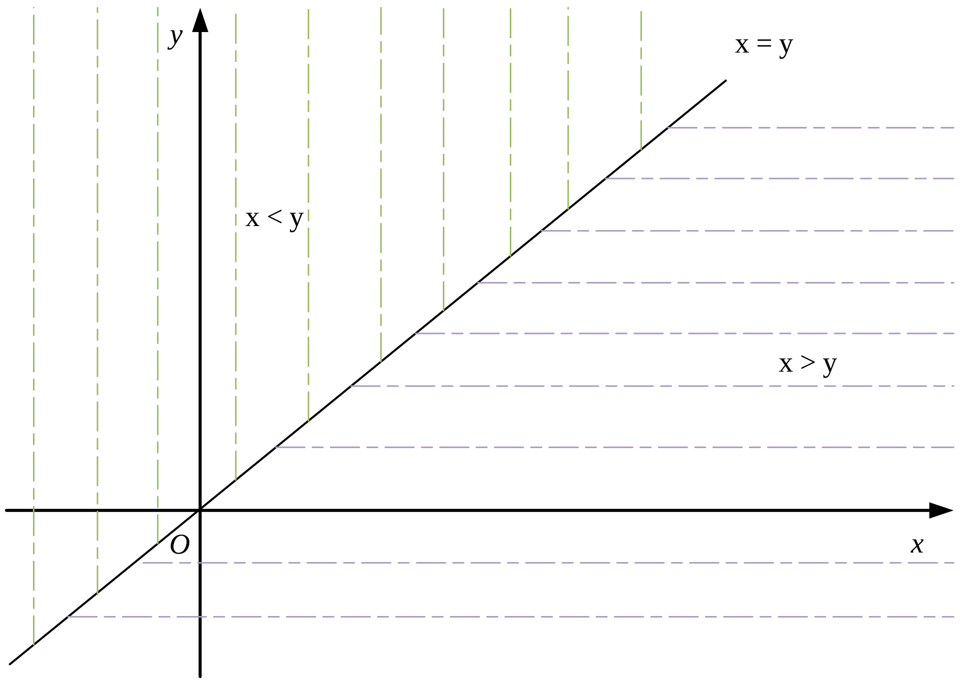
The rules for comparing both discrete and interval grey numbers have been discussed by [18], by taking into account the probability interval that can be attributed to each number.

Let us consider two independent grey numbers noted as $\otimes_{a} \in [\underline{a}, \overline{a}]$ **and** $\otimes_{b} \in [\underline{b}, \overline{b}]$**, with** $\underline{a} < \overline{a}$ **and** $\underline{b} < \overline{b}$. We will try to address the comparison issue first through a graphical approach and then by using the idea included by the probability density function.

For the moment, we shall refer to the comparison of the two independent grey numbers by **not taking into account their nature (discrete or continuous)** in order to better explain how the comparison can be seen on a general case. Later, we will move to discussing specifically each of the two situations (discrete and continuous grey numbers).

Comparing Grey Numbers

Graphically, in order to compare two grey numbers, one can consider a two-dimensional xOy system, in which on one of the axes, Ox, we will consider the probability value range of grey number $\otimes_a$, while on the other axes, Oy, the probability value range of grey number $\otimes_b$.



Comparing Grey Numbers

Considering the probability density functions of the two grey numbers, one can write:

$$\int_{-\infty}^{+\infty} f(x)dx = 1 \tag{2.7}$$

$$\int_{-\infty}^{+\infty} f(y)dy = 1 \tag{2.8}$$

Where: $f(x)$ is the probability density function of $\otimes_a$ and $f(y)$ is the probability density function of $\otimes_b$.

Comparing Grey Numbers

Consequently, the joint probability function of the two grey numbers, $\otimes_a$ and $\otimes_b$, noted as $f(x,y)$, is [18]:

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)dxdy = 1 \tag{2.9}$$

Therefore [18]:

$$p(\otimes_a > \otimes_b) = \int\int_{x>y} f(x,y)dxdy \tag{2.10}$$

Three results are possible depending on the value of $p(\otimes_a > \otimes_b)$ [18]:

- When $p(\otimes_a > \otimes_b) = 1$ then $\otimes_a > \otimes_b$
- When $p(\otimes_a > \otimes_b) = \mathrm{p}$, with $0 < \mathrm{p} < 1$ then $\otimes_a >_p \otimes_b$ – in this case, by $>_p$ it has been noted that the probable value of $\otimes_a$ is grater than the probable value of $\otimes_b$ with the probability p
- When $p(\otimes_a > \otimes_b) = 0$ then $\otimes_a < \otimes_b$

Comparing Discrete Grey Numbers

In order to discuss the discrete grey numbers comparison, we will consider in the following two discrete grey numbers $\otimes_a$ and $\otimes_b$ described as follows:

$$\otimes_a = \left\{ {}^{d_{a1}}/_{p_{a1}}, {}^{d_{a2}}/_{p_{a2}}, \dots, {}^{d_{an}}/_{p_{an}} \right\}$$

and

$$\otimes_b = \left\{ {}^{d_{b1}}/_{p_{b1}}, {}^{d_{b2}}/_{p_{b2}}, \dots, {}^{d_{bm}}/_{p_{bm}} \right\}$$

Where:

- p_{ai} is the probability of $\otimes_a$ at the point d_{ai}
- p_{bi} is the probability of $\otimes_b$ at the point d_{bi}

And:

$$\sum_{i=1}^{n} p_{ai} = 1$$

$$\sum_{j=1}^{m} p_{bj} = 1$$

Comparing Discrete Grey Numbers

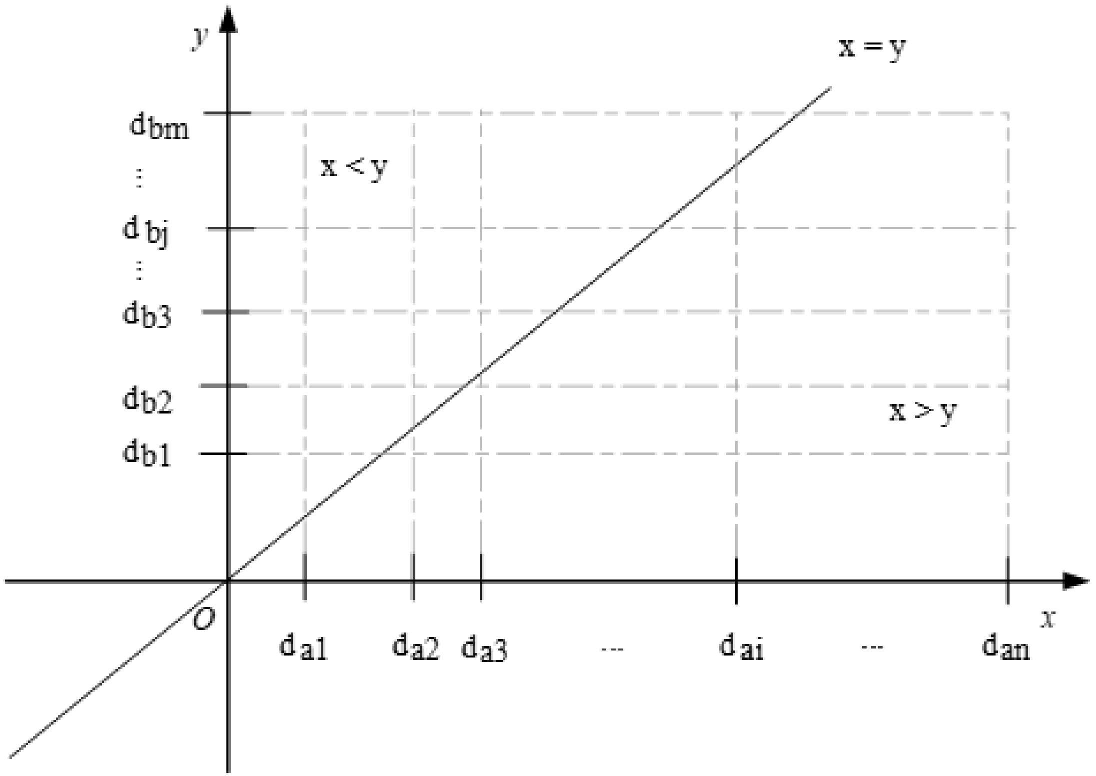


Fig . 3 Example on comparing two discrete grey numbers

Comparing Discrete Grey Numbers

If one considers the graphical approach as mentioned above and keeps the same notations, using the bidimensional space xOy, with Ox representing the probability value range of the discrete grey number $\otimes_a$ and Oy representing the probability value range of the discrete grey number $\otimes_b$, the same results can be obtained, namely [18]:

- The straight line $x = y$ represents the case in which the probable value of $\otimes_a$ is equal to the probable value of $\otimes_b$
- While when $x > y$ the probable value of the grey number $\otimes_a$ is greater than the probable value of the grey number $\otimes_b$, and
- When $x < y$ the probable value of the grey number $\otimes_a$ is lower than the probable value of the grey number $\otimes_b$.

Fig. 3 provides a graphical representation on the comparison of two discrete grey numbers.

Choosing two random values from the value-covered sets of the two discrete grey numbers, noted through d_{ai} – the random value in the value covered set of $\otimes_a$, and d_{bj} – the random value in the value covered set of $\otimes_b$, one can determine the following probability [18]:

$$p(d_{ai} > d_{bj}) = \begin{cases} 0, & d_{ai} < d_{bj} \\ 0.5 p_{ai} p_{bj}, & d_{ai} = d_{bj} \\ p_{ai}, & p_{bj} d_{ai} > d_{bj} \end{cases}$$

Comparing Discrete Grey Numbers

Which conducts to [18]:

$$p(\otimes_a > \otimes_b) = \sum_{i=1}^{n} \sum_{j=1}^{m} p(d_{ai} > d_{bj})$$

Which is equivalent to:

$$p(\otimes_a > \otimes_b) = \sum_{j=1}^{m} \sum_{i=1}^{n} p(d_{ai} > d_{bj})$$

Therefore, three cases are possible [18]:

- If $p(\otimes_a > \otimes_b) = 1$ then $\otimes_a > \otimes_b$
- When $p(\otimes_a > \otimes_b) = \mathrm{p}$, with $0 < \mathrm{p} < 1$ then $\otimes_a >_p \otimes_b$ – as mentioned above, in this case, by $>_p$ it has been noted that the probable value of $\otimes_a$ is grater than the probable value of $\otimes_b$ with the probability p
- If $p(\otimes_a > \otimes_b) = 0$ then $\otimes_a < \otimes_b$

Comparing Discrete Grey Numbers

Let us consider two discrete grey numbers having the following form:

$$\otimes_a = \left\{{}^{d_{a1}}/_{p_{a1}}, {}^{d_{a2}}/_{p_{a2}}, {}^{d_{a3}}/_{p_{a3}}, {}^{d_{a4}}/_{p_{a4}}, {}^{d_{a5}}/_{p_{a5}}\right\} = \left\{{}^{5}/_{0.1}, {}^{6}/_{0.2}, {}^{7}/_{0.4}, {}^{7.5}/_{0.2}, {}^{8}/_{0.1}\right\}$$

and

$$\otimes_b = \left\{{}^{d_{b1}}/_{p_{b1}}, {}^{d_{b2}}/_{p_{b2}}, {}^{d_{b3}}/_{p_{b3}}, {}^{d_{b4}}/_{p_{b4}}\right\} = \left\{{}^{6.5}/_{0.2}, {}^{7}/_{0.4}, {}^{7.4}/_{0.3}, {}^{8.2}/_{0.1}\right\}$$

Comparing Discrete Grey Numbers

First, it can be observed that the requirements related to the sum of probabilities being equal to 1 are fulfilled:

$$\sum_{i=1}^{5} p_{ai} = 0.1 + 0.2 + 0.4 + 0.2 + 0.1 = 1$$

$$\sum_{j=1}^{4} p_{bj} = 0.2 + 0.4 + 0.3 + 0.1 = 1$$

Comparing Discrete Grey Numbers

Next, the probabilities of each d_{ai} being greater than each d_{bj} are determined based on the equation (2.11):

$$p(d_{a1} > d_{b1}) = p(5 > 6.5) = 0$$
$$p(d_{a1} > d_{b2}) = p(5 > 7) = 0$$
$$p(d_{a1} > d_{b3}) = p(5 > 7.4) = 0$$
$$p(d_{a1} > d_{b4}) = p(5 > 8.2) = 0$$
$$p(d_{a2} > d_{b1}) = p(6 > 6.5) = 0$$
$$p(d_{a2} > d_{b2}) = p(6 > 7) = 0$$
$$p(d_{a2} > d_{b3}) = p(6 > 7.4) = 0$$
$$p(d_{a2} > d_{b4}) = p(6 > 8.2) = 0$$
$$p(d_{a3} > d_{b1}) = p(7 > 6.5) = p_{a3}p_{b1} = 0.4 * 0.2 = 0.08$$
$$p(d_{a3} > d_{b2}) = p(7 > 7) = 0.5p_{a3}p_{b2} = 0.5 * 0.4 * 0.4 = 0.08$$
$$p(d_{a3} > d_{b3}) = p(7 > 7.4) = 0$$
$$p(d_{a3} > d_{b4}) = p(7 > 8.2) = 0$$
$$p(d_{a4} > d_{b1}) = p(7.5 > 6.5) = p_{a4}p_{b1} = 0.2 * 0.2 = 0.04$$
$$p(d_{a4} > d_{b2}) = p(7.5 > 7) = p_{a4}p_{b2} = 0.2 * 0.4 = 0.08$$
$$p(d_{a4} > d_{b3}) = p(7.5 > 7.4) = p_{a4}p_{b3} = 0.2 * 0.3 = 0.06$$
$$p(d_{a4} > d_{b4}) = p(7.5 > 8.2) = 0$$
$$p(d_{a5} > d_{b1}) = p(8 > 6.5) = p_{a5}p_{b1} = 0.1 * 0.2 = 0.02$$
$$p(d_{a5} > d_{b2}) = p(8 > 7) = p_{a5}p_{b2} = 0.1 * 0.4 = 0.04$$
$$p(d_{a5} > d_{b3}) = p(8 > 7.4) = p_{a5}p_{b3} = 0.1 * 0.3 = 0.03$$
$$p(d_{a5} > d_{b4}) = p(8 > 8.2) = 0$$

Comparing Discrete Grey Numbers

Thus, according to (2.12), we can get the probability of $\otimes_a$ to be greater than $\otimes_b$:

$$p(\otimes_a > \otimes_b) = \sum_{i=1}^{5}\sum_{j=1}^{4} p\left(d_{ai} > d_{bj}\right) = 0.08 + 0.08 + 0.04 + 0.08 + 0.06 + 0.02 + 0.04 + 0.03 = 0.430$$

Therefore, one can affirm that $\otimes_a$ is greater than $\otimes_b$ with the probability of 0.430. This affirmation is equivalent to saying that $\otimes_b$ is greater than $\otimes_a$ with the probability of $1 - 0.430 = 0.570$.

Comparing Interval Grey Numbers

Let us consider two interval grey numbers and we will note them through $\otimes_a$ and $\otimes_b$, with:

$$\otimes_a \in [\underline{a}, \overline{a}]$$

and

$$\otimes_b \in [\underline{b}, \overline{b}],$$

with $\underline{a} < \overline{a}$ and $\underline{b} < \overline{b}$.

For each of the two considered interval grey numbers, one can define a probability density function noted as $f(\cdot)$ [18]:

$$\int_{\underline{a}}^{\overline{a}} f(x)dx = 1$$

$$\int_{\underline{b}}^{\overline{b}} f(y)dy = 1$$

where through $f(x)$ we have noted the probability density function for $\otimes_a$, while through $f(y)$ we have noted the probability density function for $\otimes_b$.

Comparing Interval Grey Numbers

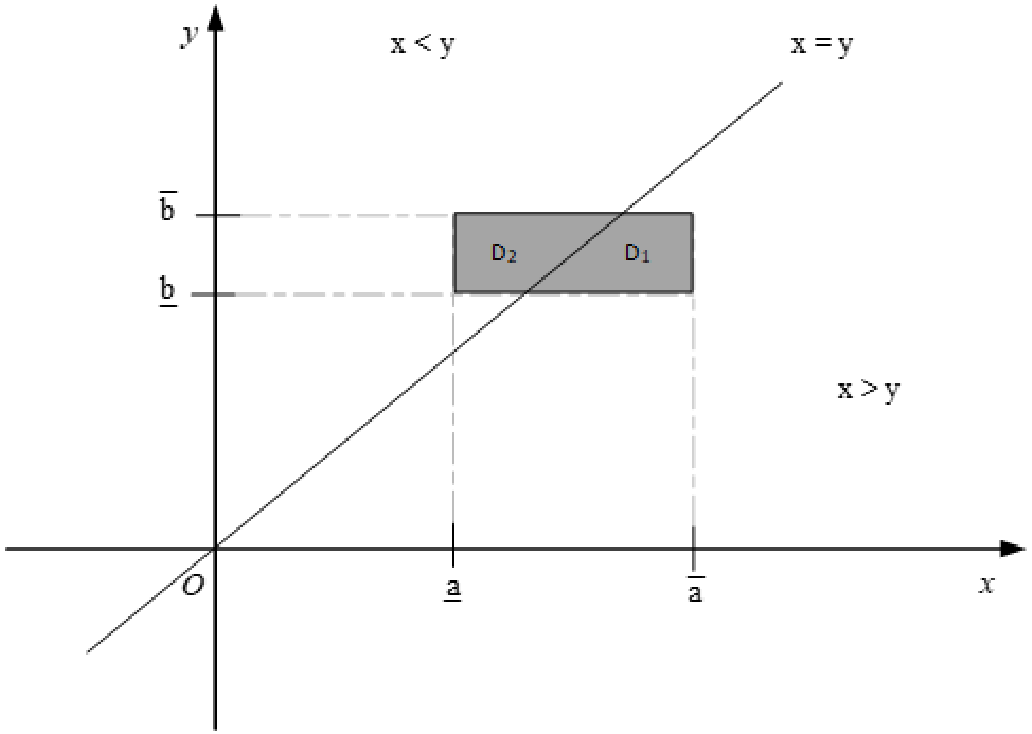


Fig . 4 Example on comparing two interval grey numbers when considering the area resulted by the points $(\underline{a}, \underline{b})$, $(\underline{a}, \bar{b})$, $(\bar{a}, \underline{b})$ and $(\bar{a}, \bar{b})$

Comparing Interval Grey Numbers

Using the idea of joint probability, Xie and Liu [18] define the probability that $\otimes_a$ is greater than $\otimes_b$, noted as $p(\otimes_a > \otimes_b)$, as being:

$$p(\otimes_a > \otimes_b) = \frac{\int \int_{D_1} f(x,y)dxdy}{\int \int_{D_1+D_2} f(x,y)dxdy} \tag{2.11}$$

where through D_1 it has been noted the area positioned in the right-side of the $x = y$ line, while through D_2 it has been noted the area positioned in the left-side of the $x = y$ line, as depicted in **Fig** . 4.

Comparing Interval Grey Numbers

Further, let us define [18]:

$$\int\int_{D_1+D_2} f(x,y)dxdy = \sigma \tag{2.12}$$

Considering the position of the points $(\underline{a}, \underline{b})$, $(\underline{a}, \overline{b})$, $(\overline{a}, \underline{b})$ and $(\overline{a}, \overline{b})$, Xie and Liu [18] noted that there are six possible situations in which one can be when comparing two interval grey numbers:

- Situation 1: $\underline{b} < \overline{b} < \underline{a} < \overline{a}$
- Situation 2: $\underline{a} < \overline{a} < \underline{b} < \overline{b}$
- Situation 3: $\underline{a} < \underline{b} < \overline{a} < \overline{b}$
- Situation 4: $\underline{b} < \underline{a} < \overline{b} < \overline{a}$
- Situation 5: $\underline{b} < \underline{a} < \overline{a} < \overline{b}$
- Situation 6: $\underline{a} < \underline{b} < \overline{b} < \overline{a}$

which are discussed in the following using a one-dimensional and a bi-dimensional graphical approach along with a mathematical expression which provides the calculus of the $p(\otimes_a > \otimes_b)$.

The simplest situations in which one needs to compare interval grey numbers and with the most intuitive results are the ones in which the two interval grey numbers are not overlapping, namely Situation 1 and Situation 2.

2. Grey Numbers and their Mathematics

Comparison

Comparing Interval Grey Numbers

- Situation 1: $\underline{b} < \overline{b} < \underline{a} < \overline{a}$

The succession $\underline{b} < \overline{b} < \underline{a} < \overline{a}$ can be shortly rewritten as $\overline{b} < \underline{a}$, keeping in mind that $\underline{b} < \overline{b}$ and that $\underline{a} < \overline{a}$.

In order to better visualize this situation a one-dimensional graphical approach can be used as presented in **Fig .** 5. As it can be observed from the **Fig .** 5, in this case $\otimes_a$ is greater than $\otimes_b$, thus the probability that $\otimes_a$ is greater than $\otimes_b$ is equal to 1.

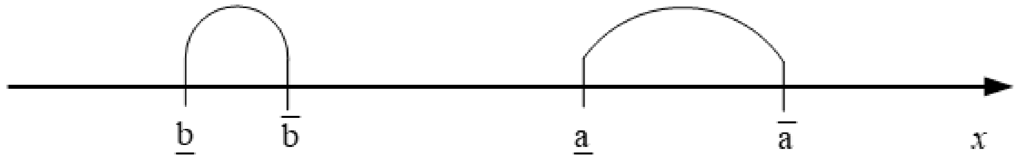


Fig . 5 One-dimensional representation of Situation 1

The probability that $\otimes_a$ is greater than $\otimes_b$, $p(\otimes_a > \otimes_b)$, can be determined mathematically by solving the following [18]:

$$p(\otimes_a > \otimes_b) = \iint_{D_1} {}^{f(x,y)dxdy}/_{\sigma} = 1 \qquad (2.12)$$

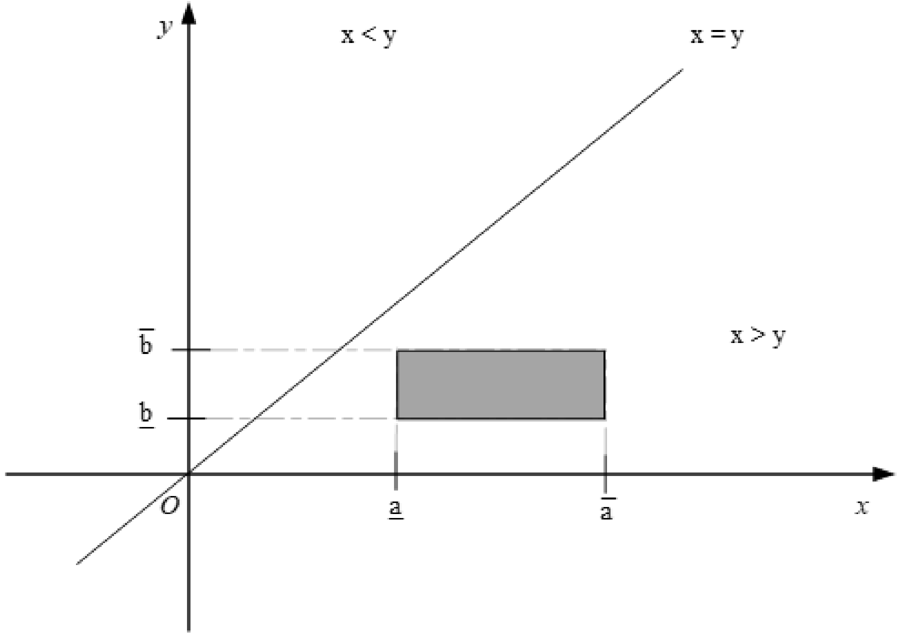


Fig . 6 Bi-dimensional representation of Situation 1

2. Grey Numbers and their Mathematics

Comparison

Comparing Interval Grey Numbers

- Situation 2: $\underline{a} < \overline{a} < \underline{b} < \overline{b}$

Even in this case, the succession $\underline{a} < \overline{a} < \underline{b} < \overline{b}$ can be shortly rewritten as $\underline{b} > \overline{a}$, knowing that $\underline{b} < \overline{b}$ and that $\underline{a} < \overline{a}$.
From the one-dimension graphical representation in Fig . 7 it can be observed that the probability that $\otimes_a$ is greater than $\otimes_b$ is equal to 0.

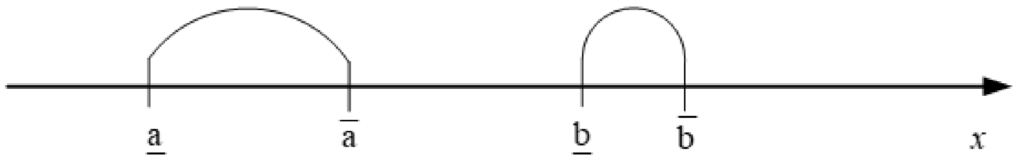


Fig . 7 One-dimensional representation of Situation 2

The same result is reached if one solves the following mathematical equation [18]:

$$p(\otimes_a > \otimes_b) = \iint_{D_1} f(x,y)dxdy \big/ \sigma = 0 \qquad (2.13)$$

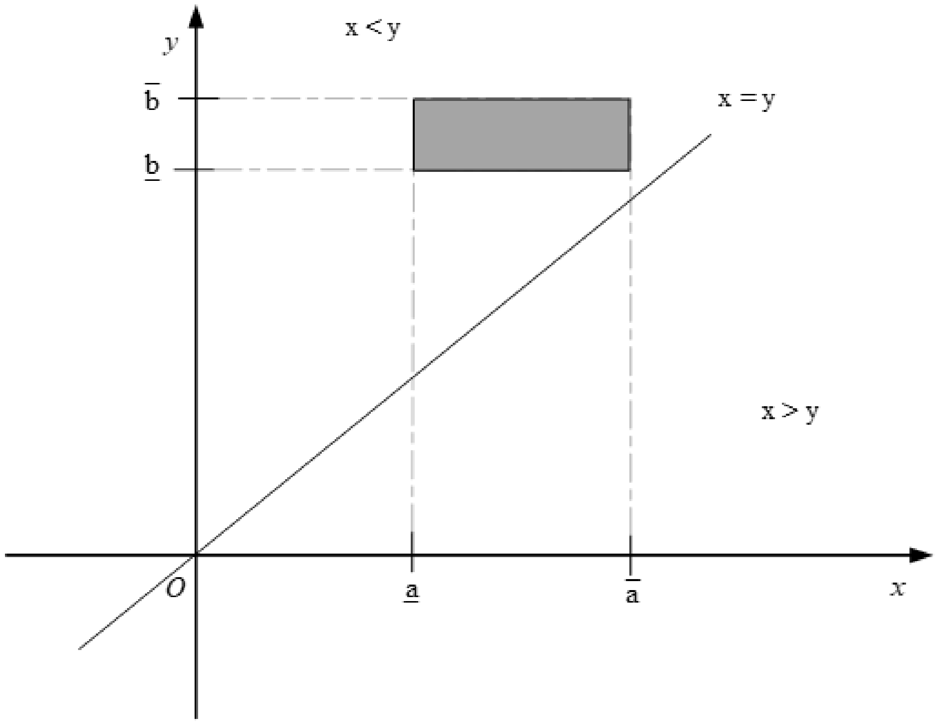


Fig . 8 Bi-dimensional representation of Situation 2

Comparing Interval Grey Numbers

- Situation 3: $\underline{a} < \underline{b} < \overline{a} < \overline{b}$

In this situation, the two interval grey numbers are partly overlapping as it results from the one-dimensional graphical representation in **Fig .** 9.

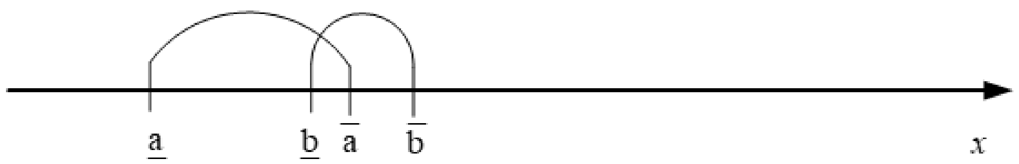


Fig . 9 One-dimensional representation of Situation 3

In order to determine the probability that $\otimes_a$ is greater than $\otimes_b$, one should solve [18]:

$$p(\otimes_a > \otimes_b) = \iint_{D_1} f(x,y)dxdy \Big/ \sigma = \int_{\underline{b}}^{\overline{a}} \int_{\underline{b}}^{y} f(x,y)dxdy \Big/ \sigma \qquad (2.14)$$

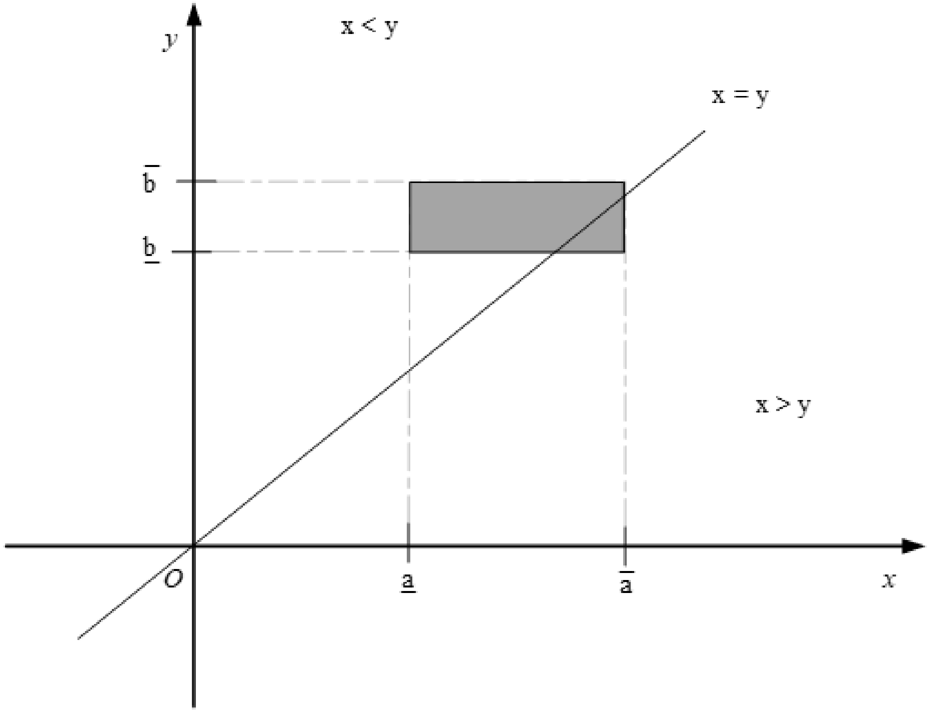


Fig . 10 Bi-dimensional representation of Situation 3

2. Grey Numbers and their Mathematics

Comparison

Comparing Interval Grey Numbers

- Situation 4: $\underline{b} < \underline{a} < \overline{b} < \overline{a}$

Even in the case of Situation 4, the two interval grey numbers are overlapping as depicted in Fig . 11. In this case, the probability that $\otimes_a$ is greater than $\otimes_b$ is determined as [18]:

$$p(\otimes_a > \otimes_b) = \iint_{D_1} f(x,y)dxdy/_{\sigma} = 1 - \iint_{D_2} f(x,y)dxdy/_{\sigma} = 1 - \int_{\underline{a}}^{\overline{b}} \int_{y}^{\overline{b}} f(x,y)dxdy/_{\sigma} \quad (2.15)$$

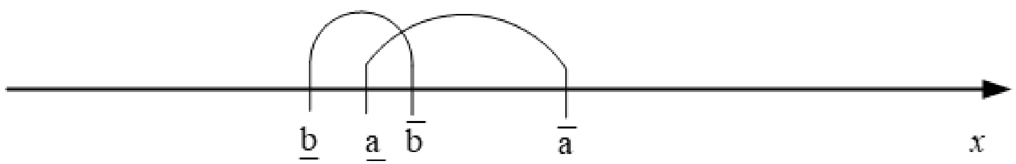


Fig . 11 One-dimensional representation of Situation 4

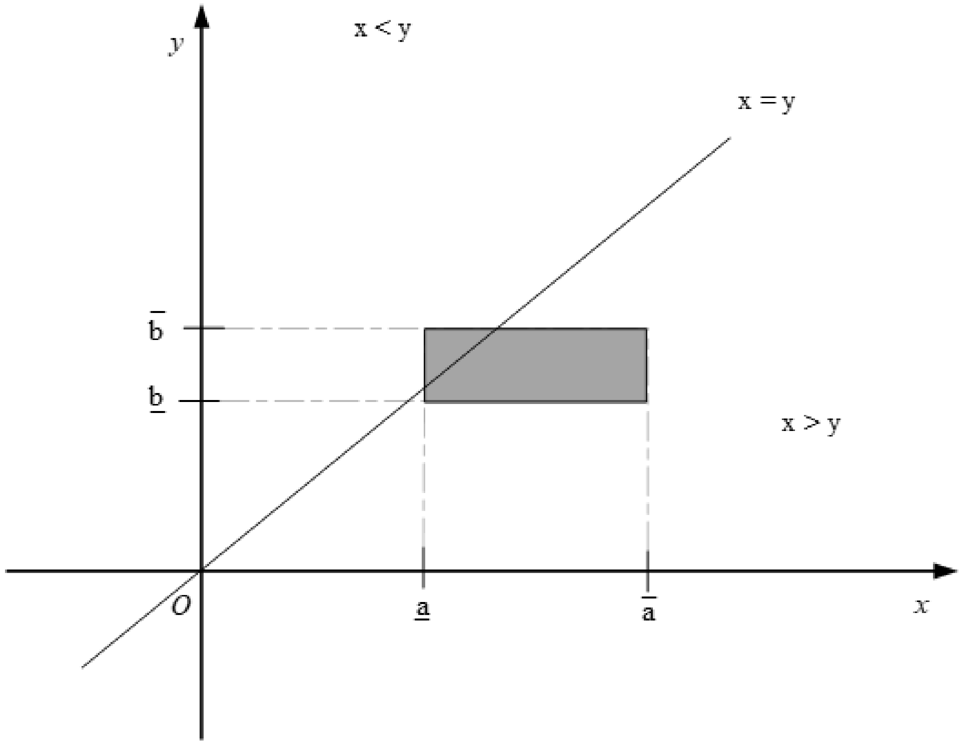


Fig . 12 Bi-dimensional representation of Situation 4

Comparing Interval Grey Numbers

- Situation 5: $\underline{b} < \underline{a} < \overline{a} < \overline{b}$

Given the succession of the values for $\underline{a}, \overline{a}, \underline{b}$ and $\overline{b}$, Situation 5 characterizes a case in which the interval grey number $\otimes_b$ include the interval grey number $\otimes_a$, as depicted in **Fig .** 13.

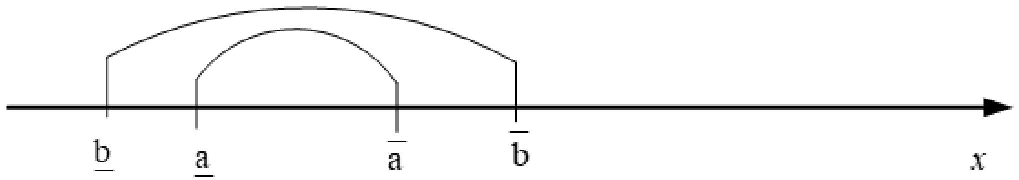


Fig . 13 One-dimensional representation of Situation 5

For determining the probability that $\otimes_a$ is greater than $\otimes_b$ the following equations should be solved as determined by Xie and Liu [18]:

$$p(\otimes_a > \otimes_b) = \iint_{D_1} f(x,y)dxdy \Big/ \sigma = \left[\int_{\underline{a}}^{\overline{a}} \int_{\underline{a}}^{y} f(x,y)dxdy + \int_{\underline{b}}^{\underline{a}} \int_{\underline{a}}^{\overline{a}} f(x,y)dxdy\right] \Big/ \sigma \qquad (2.16)$$

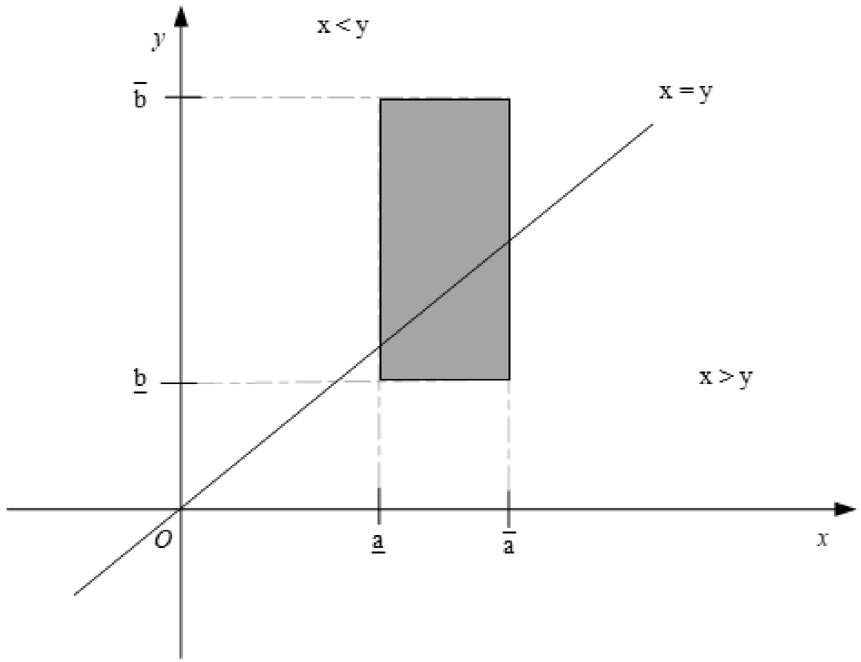


Fig . 14 Bi-dimensional representation of Situation 5

Comparing Interval Grey Numbers

- Situation 6: $\underline{a} < \underline{b} < \overline{b} < \overline{a}$

Even Situation 6 is an example of overlapping situation, but in this case the interval grey number $\otimes_a$ include the interval grey number $\otimes_b$ - Fig . 15.

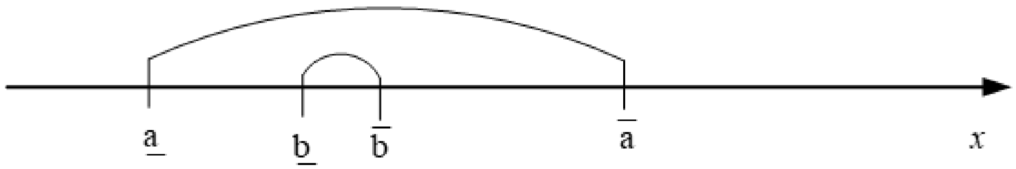


Fig . 15 One-dimensional representation of Situation 6

The probability that $\otimes_a$ is greater than $\otimes_b$ is determined as suggested by Xie and Liu [18]:

$$p(\otimes_a > \otimes_b) = \iint_{D_1} f(x,y)dxdy / \sigma = \left[\int_{\underline{a}}^{\overline{a}} \int_{\underline{a}}^{y} f(x,y)dxdy + \int_{\underline{b}}^{\underline{a}} \int_{\underline{a}}^{\overline{a}} f(x,y)dxdy\right] / \sigma \quad (2.17)$$

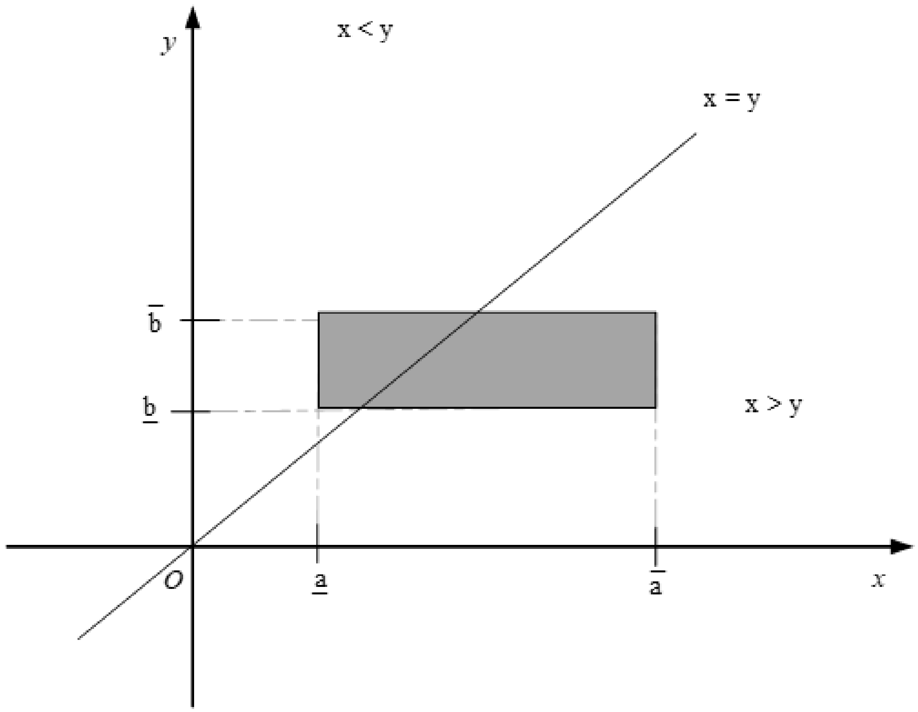


Fig . 16 Bi-dimensional representation of Situation 6

Comparing Interval Grey Numbers

2. Grey Numbers and their Mathematics

Comparison

Comparing Interval Grey Numbers

In the particular case in which the probabilities of any two values in both the covered-sets of the intervals grey numbers to be compared are equal, the probability function, $f(x, y) = 1$, and the probability that $\otimes_a$ is greater than $\otimes_b$ can be determined as [17, 18]:

$$p(\otimes_a > \otimes_b) = \frac{A_1}{A_1 + A_2} \tag{2.18}$$

where A_1 is the area of the rectangle characterized by the points $(\underline{a}, \underline{b})$, $(\underline{a}, \overline{b})$, $(\overline{a}, \underline{b})$, $(\overline{a}, \overline{b})$ and located in the right-side of the $x = y$ line, and A_2 is the area of the rectangle characterized by the points $(\underline{a}, \underline{b})$, $(\underline{a}, \overline{b})$, $(\overline{a}, \underline{b})$, $(\overline{a}, \overline{b})$ and located in the left-side of the $x = y$ line, as presented in **Fig .** 17.

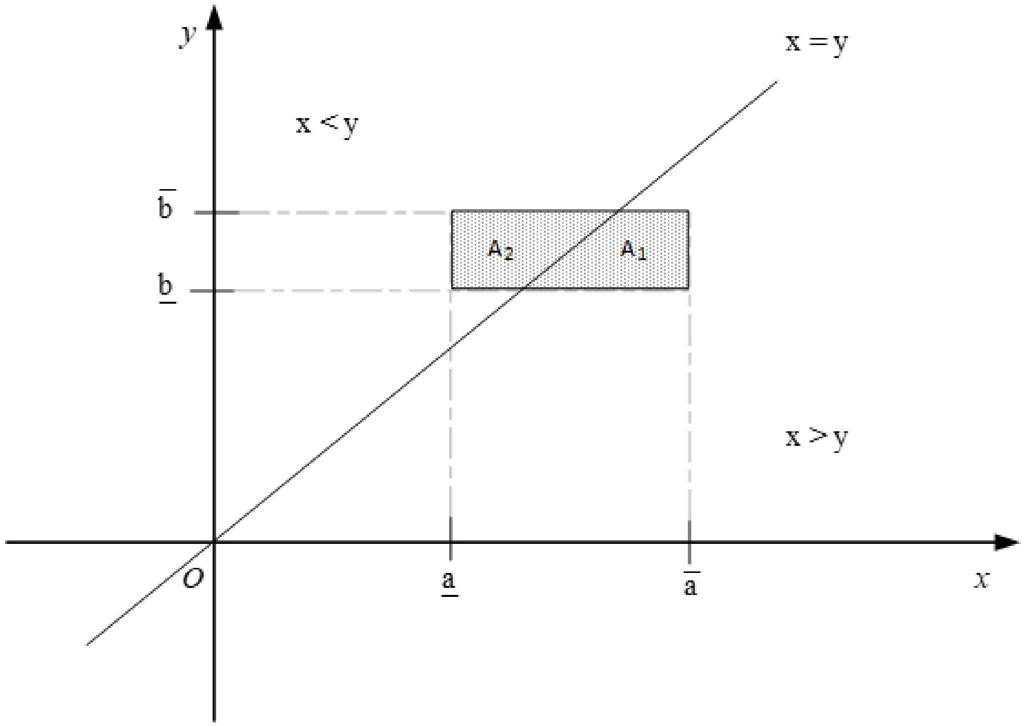


Fig . 17 Area-approach to comparing two interval grey numbers

Comparing Interval Grey Numbers

Furthermore, considering the particular case in which $\otimes_a$ is equal to $\otimes_b$ and in which the probabilities of any two values in both the covered-sets of the intervals grey numbers to be compared are equal, the situation depicted in **Fig .** 18 is encountered.

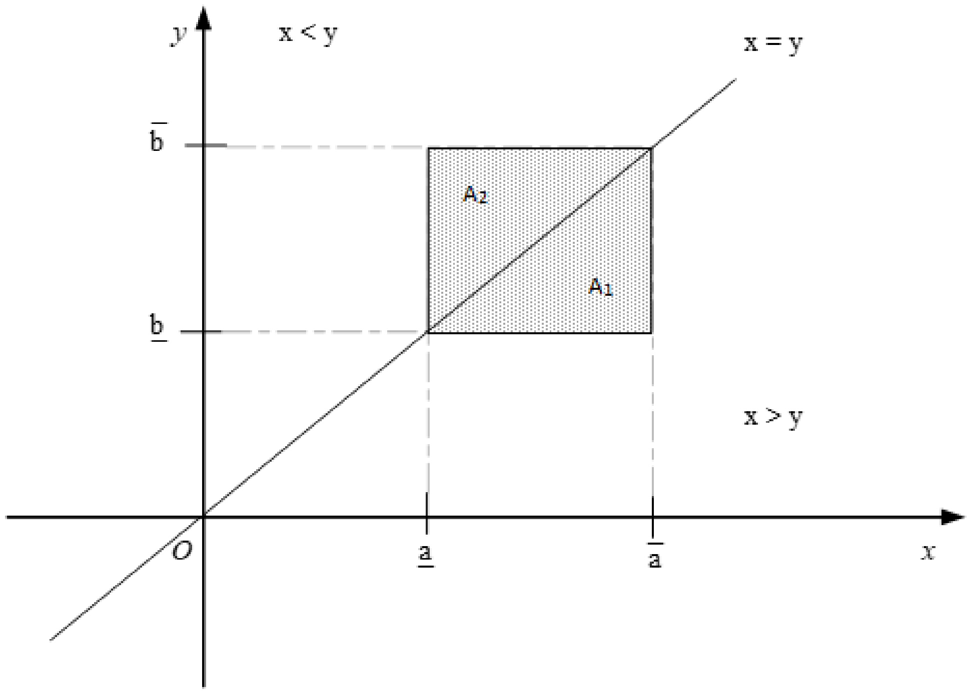


Fig . 18 Example of a particular case in which $\otimes_a$ is equal to $\otimes_b$ and in which the probabilities of any two values in both the covered-sets of the intervals grey numbers to be compared are equal

In this particular case, $A_1 = A_2$, and equation (2.18) becomes:

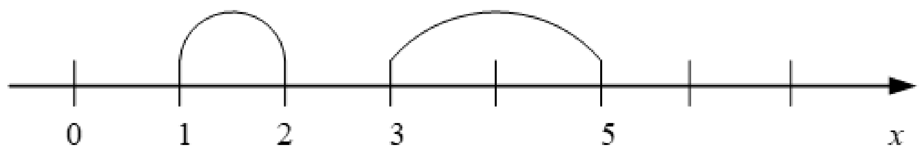
$$p(\otimes_a > \otimes_b) = \frac{A_1}{A_1+A_2} = \frac{A_1}{A_1+A_1} = \frac{A_1}{2 * A_1} = 0.5 \qquad (2.19)$$

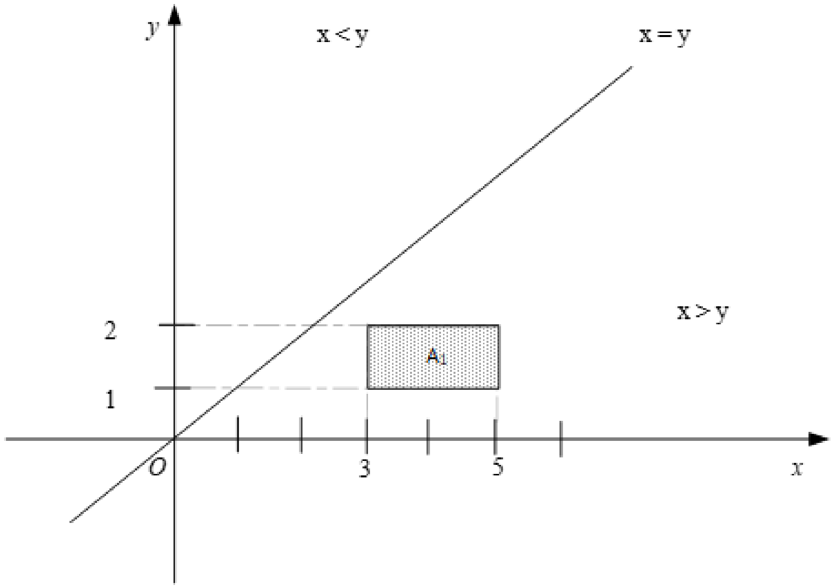
Comparing Interval Grey Numbers

- Numerical example 1: $\otimes_a = [3, 5]$ and $\otimes_b = [1, 2]$

In this case, the interval grey numbers are as in Fig. 19, so the result matches the Situation 1. The probability that $\otimes_a$ is greater than $\otimes_b$ is:

$$p(\otimes_a > \otimes_b) = 1$$





Otherwise, one can use equation (2.19) for determining the probability that $\otimes_a$ is greater than $\otimes_b$. Using the bi-dimensional representation of the two numbers (please see Fig. 20), it can be observed that $A_2 = 0$ and therefore:

$$p(\otimes_a > \otimes_b) = \frac{A_1}{A_1 + A_2} = \frac{A_1}{A_1 + 0} = \frac{A_1}{A_1} = 1$$

Comparing Interval Grey Numbers

- Numerical example 2: $\otimes_a = [2, 3]$ and $\otimes_b = [4, 6]$

In this case, the interval grey number $\otimes_b$ is higher than the interval grey number $\otimes_a$ (Fig . 21), and the probability that $\otimes_a$ is greater than $\otimes_b$ is 0.

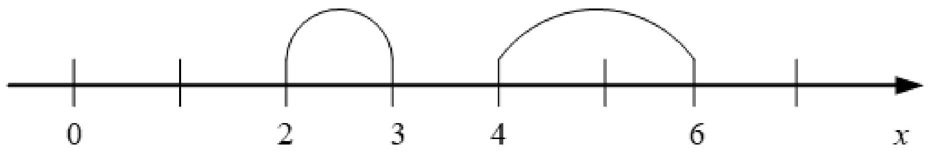


Fig . 21 Graphical representation in one-dimensional space for Numerical example 2

The same result is achieved by applying equation (2.19) in which $A_1 = 0$, as it can be observed from Fig . 22:

$$p(\otimes_a > \otimes_b) = 0$$

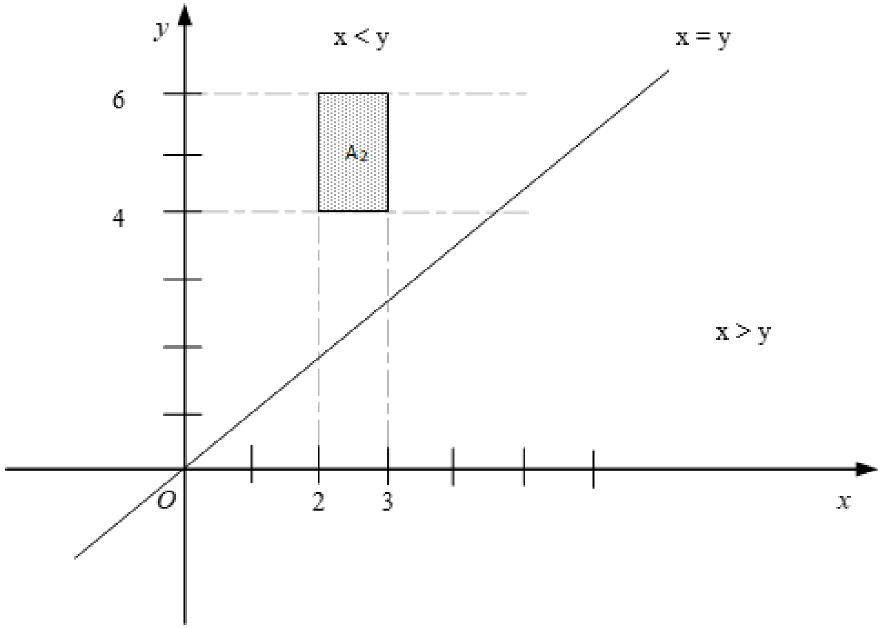


Fig . 22 Graphical representation in bi-dimensional space for Numerical example 2

Comparing Interval Grey Numbers

- Numerical example 3: $\otimes_a = [1, 4]$ and $\otimes_b = [3, 6]$

Numerical example 3 fits the Situation 3 presented above from a theoretical point of view. On a one-dimensional axe, the position of the two interval grey numbers $\otimes_a$ and $\otimes_b$ is as in **Fig .** 23. The bi-dimensional representation with the highlight of the two areas A_1 and A_2 is depicted in **Fig .** 24.

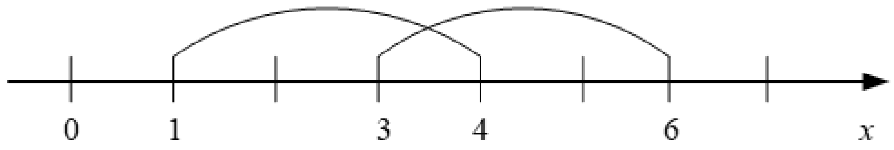


Fig . 23 Graphical representation in one-dimensional space for Numerical example 3

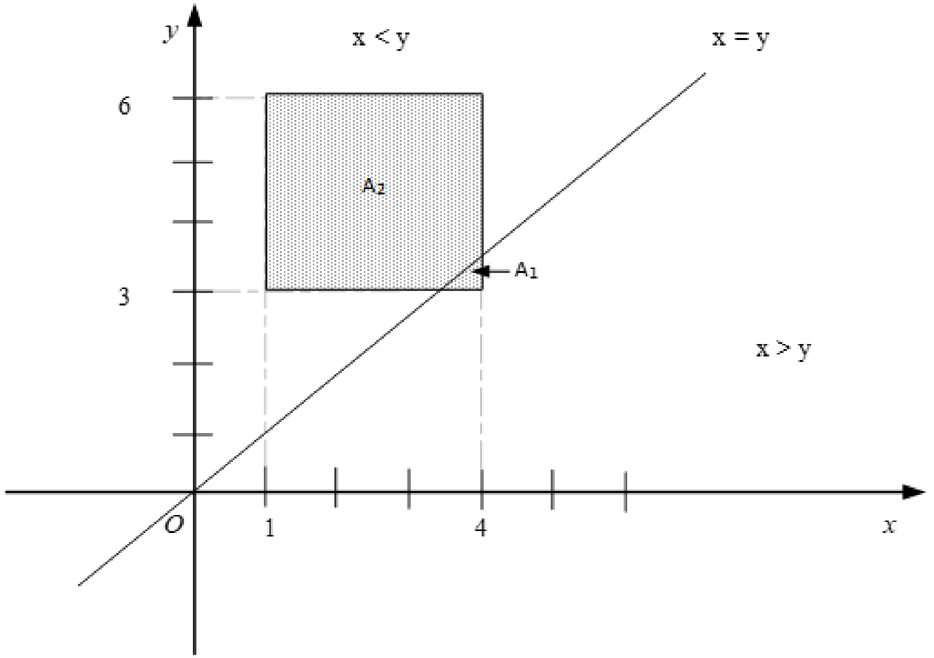


Fig . 24 Graphical representation in bi-dimensional space for Numerical example 3

To determine the probability that $\otimes_a$ is greater than $\otimes_b$, the following formula is applied:

$$p(\otimes_a > \otimes_b) = \frac{(\overline{a} - \underline{b})^2}{2(\overline{a} - \underline{a})(\overline{b} - \underline{b})} = \frac{(4-3)^2}{2 * (4-1)(6-3)} = \frac{1}{18} = 0.056$$

As a result, it can be stated that the probability that $\otimes_a$ is greater than $\otimes_b$ is 0.083, or that the probability that $\otimes_b$ is greater than $\otimes_a$ is $1 - 0.056 = 0.944$.

Comparing Interval Grey Numbers

- Numerical example 4: $\otimes_a = [1, 6]$ and $\otimes_b = [0, 4]$

In this case, it can be observed that we are in Situation 4 as $\underline{b} < \underline{a} < \overline{b} < \overline{a}$ – please see **Fig .** 25 and **Fig .** 26.

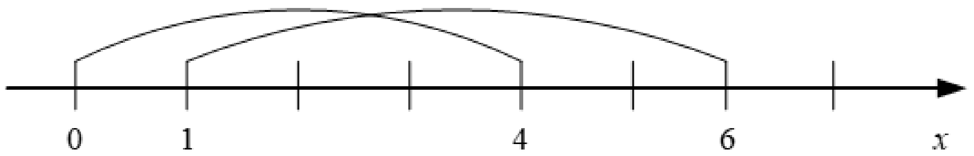


Fig . 25 Graphical representation in one-dimensional space for Numerical example 4

For determining the probability that $\otimes_a$ is greater than $\otimes_b$, we shall use:

$$p(\otimes_a > \otimes_b) = 1 - {(\overline{b} - \underline{a})^2}\Big/{[2(\overline{a} - \underline{a})(\overline{b} - \underline{b})]} =$$

$$1 - {(4-1)^2}\Big/{[2 * (6-1) * (4-0)]} = 1 - \frac{9}{40} = 0.775$$

Therefore, the probability that $\otimes_a$ is greater than $\otimes_b$ is 0.775, or that the probability that $\otimes_b$ is greater than $\otimes_a$ is $1 - 0.775 = 0.225$. Considering the bi-dimensional representation from **Fig .** 28 it was expected a high value for the value of the probability as the area represented by A_1 was considerable larger than the one represented by A_2.

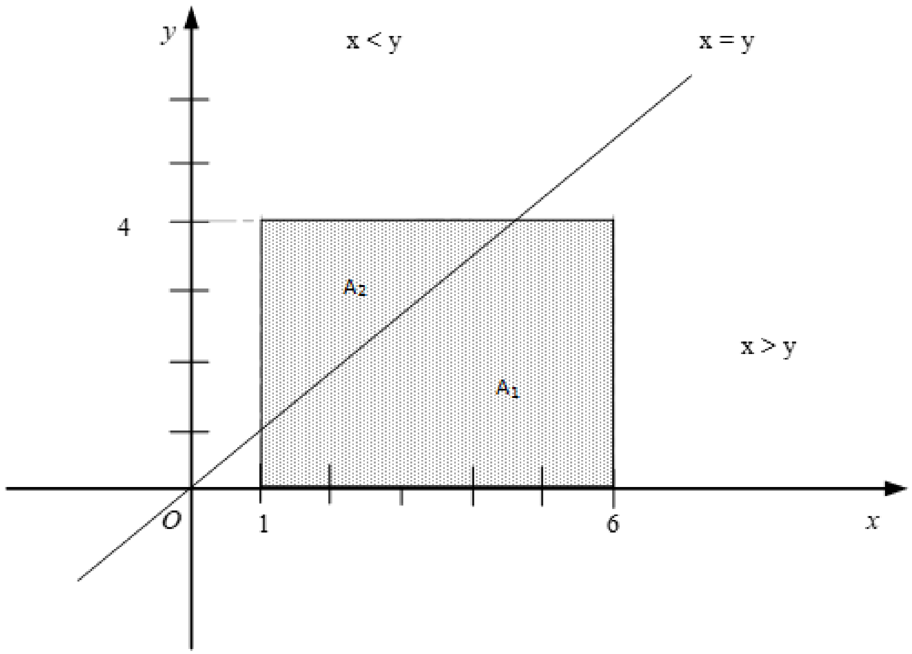


Fig . 26 Graphical representation in bi-dimensional space for Numerical example 4

Comparing Interval Grey Numbers

- Numerical example 5: $\otimes_a = [2, 4]$ and $\otimes_b = [1, 6]$

In this situation, $\underline{b} < \underline{a} < \overline{a} < \overline{b}$,as depicted in **Fig .** 27 and **Fig .** 28.

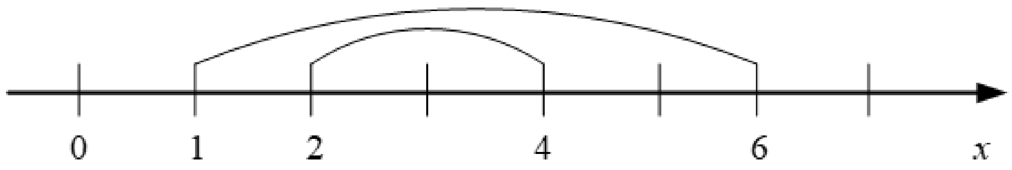


Fig . 27 Graphical representation in one-dimensional space for Numerical example 5

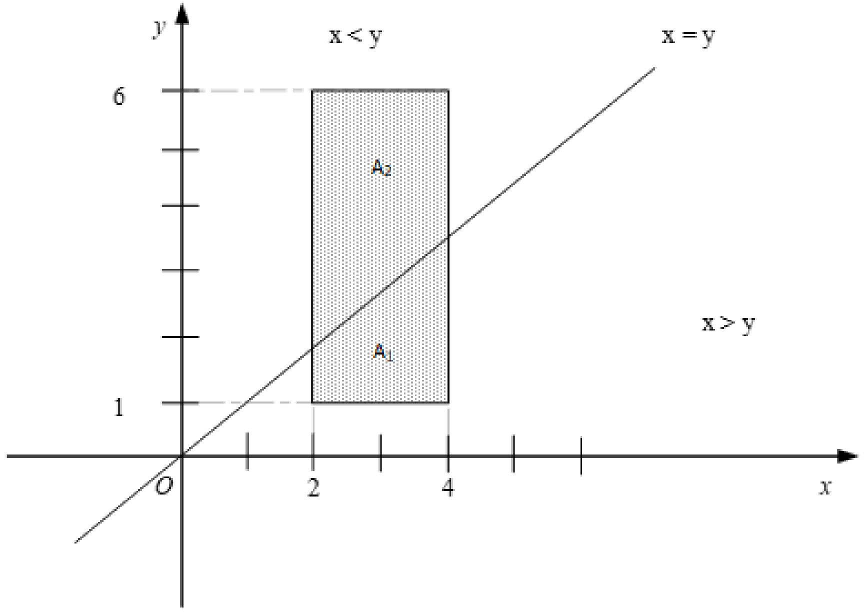


Fig . 28 Graphical representation in bi-dimensional space for Numerical example 5

The probability that $\otimes_a$ is greater than $\otimes_b$ is determined as:

$$p(\otimes_a > \otimes_b) = \left(\underline{a} + \overline{a} - 2\underline{b}\right) \Big/ \left[2(\overline{b} - \underline{b})\right] = (2 + 4 - 2 * 1) \Big/ [2 * (6 - 1)]$$

$$= 8/20 = 0.400$$

Comparing Interval Grey Numbers

- Numerical example 6: $\otimes_a = [0, 5]$ and $\otimes_b = [3, 4]$

It can be observed that $\underline{a} < \underline{b} < \overline{b} < \overline{a}$ in this situation. Fig . 29 and Fig . 30 depict the one-dimensional and bi-dimensional representation of the two interval grey numbers to be compared.

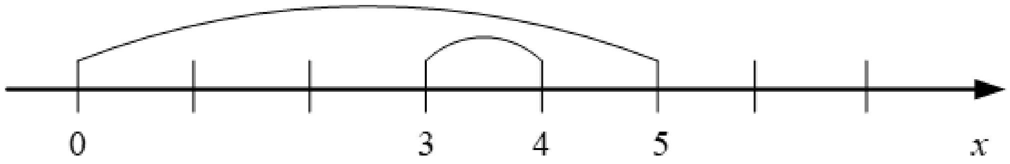


Fig . 29 Graphical representation in one-dimensional space for Numerical example 6

The probability that $\otimes_a$ is greater than $\otimes_b$ is given by:

$$p(\otimes_a > \otimes_b) = {(2\overline{a} - \overline{b} - \underline{b})}\Big/{[2(\overline{a} - \underline{a})]} = {(2 * 5 - 4 - 3)}\Big/{[2 * (5 - 0)]}$$
$$= {}^{3}/_{10} = 0.300$$

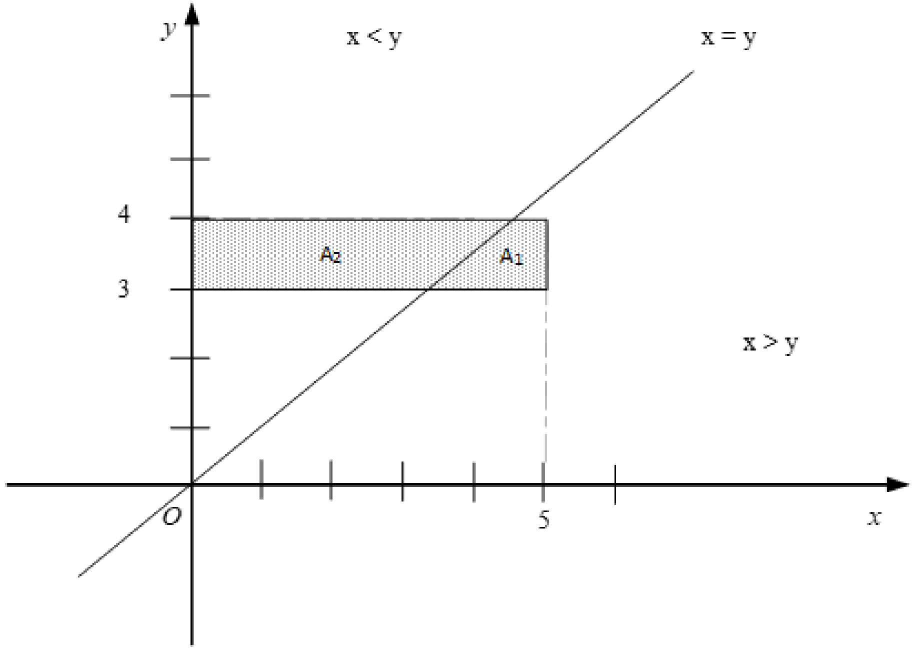


Fig . 30 Graphical representation in bi-dimensional space for Numerical example 6

Comparing Interval Grey Numbers

In order to ease the calculus implied by the interval grey numbers comparison on the particular situation in which the probabilities of any two values in both the covered-sets of the intervals grey numbers to be compared are equal, a Jupyter Notebook written in Python has been created and can be accessed at the link: https://github.com/liviucotfas/grey-systems-book/blob/main/grey-numbers-comparison-greater.ipynb

After clicking on "Open in Colab", one can insert the two grey numbers to be compared in the third input box (In [3]) by replacing the values [2, 4] and [3, 5] with the values of the grey numbers to be compared. Then, after pressing Runtime -> Run all the from the top-panel, the result of the comparison can be found in the output box Out [3]. Additionally, the program lists the situation in which the two grey numbers to be compared are.

Comparing Interval Grey Numbers

Open in Colab

Note: the `GreyNumber` class is used for representing grey numbers, while the `probability` function is used for computing the probability that a > b.

In [1]:
```
class GreyNumber(object):
    def __init__ (self, low, up):
        self.low = low
        self.up = up
```

In [2]:
```
def probability_greater(a : GreyNumber, b : GreyNumber) -> float:
    '''returns the probability that a > b'''
    if a.low > b.up:
        print("a.low > b.up")
        return 1
    elif b.low > a.up:
        print("b.low > a.up")
        return 0
    elif a.low < b.low < a.up < b.up:
        print("a.low < b.low < a.up < b.up")
        return (a.up-b.low)**2 / (2 * (a.up - a.low) * (b.up - b.low))
    elif b.low < a.low < b.up < a.up:
        print("b.low < a.low < b.up < a.up")
        return 1 - (b.up - a.low)**2 / (2 * (a.up - a.low) * (b.up - b.low))
    elif b.low < a.low < a.up < b.up:
        print("b.low < a.low < a.up < b.up")
        return (a.low + a.up - 2 * b.low)  / (2 * (b.up - b.low))
    elif a.low < b.low < b.up < a.up:
        print("a.low < b.low < a.up < b.up")
        return (2 * a.up - b.up - b.low) / (2 * (a.up - a.low))
    else:
        raise Exception("Sorry, this case is not supported")
```

Compare (≤) any two grey numbers (a1 and a2) by modifying the values below.

In [3]:
```
a = GreyNumber(2, 4) # change these values
b = GreyNumber(3, 5) # change these values
# a > b
probability_greater(a, b)
```

```
a.low < b.low < a.up < b.up
```

Out[3]:
```
0.125
```

Comparing Interval Grey Numbers

- Numerical example 1: $\otimes_a = [3, 5]$ and $\otimes_b = [1, 2]$ – see **Fig .** 32

```
a = GreyNumber(3, 5) # change these values
b = GreyNumber(1, 2) # change these values
# a > b
probability_greater(a, b)

a.low > b.up
1
```

Fig . 32 Solution for Numerical example 1

- Numerical example 2: $\otimes_a = [2, 3]$ and $\otimes_b = [4, 6]$ – see **Fig .** 33

```
a = GreyNumber(2, 3) # change these values
b = GreyNumber(4, 6) # change these values
# a > b
probability_greater(a, b)

b.low > a.up
0
```

Fig . 33 Solution for Numerical example 2

- Numerical example 3: $\otimes_a = [1, 4]$ and $\otimes_b = [3, 6]$ – see **Fig .** 34

```
a = GreyNumber(1, 4) # change these values
b = GreyNumber(3, 6) # change these values
# a > b
probability_greater(a, b)

a.low < b.low < a.up < b.up
0.05555555555555555
```

Fig . 34 Solution for Numerical example 3

- Numerical example 4: $\otimes_a = [1, 6]$ and $\otimes_b = [0, 4]$ – see **Fig .** 35

```
a = GreyNumber(1, 6) # change these values
b = GreyNumber(0, 4) # change these values
# a > b
probability_greater(a, b)

b.low < a.low < b.up < a.up
0.775
```

Fig . 35 Solution for Numerical example 4

Comparing Interval Grey Numbers

- Numerical example 5: $\otimes_a = [2, 4]$ and $\otimes_b = [1, 6]$ – see Fig. 36

```
a = GreyNumber(2, 4) # change these values
b = GreyNumber(1, 6) # change these values
# a > b
probability_greater(a, b)

b.low < a.low < a.up < b.up
0.4
```

Fig. 36 Solution for Numerical example 5

- Numerical example 6: $\otimes_a = [0, 5]$ and $\otimes_b = [3, 4]$ – see Fig. 37

```
a = GreyNumber(0, 5) # change these values
b = GreyNumber(3, 4) # change these values
# a > b
probability_greater(a, b)

a.low < b.low < a.up < b.up
0.3
```

Fig. 37 Solution for Numerical example 6

Comparing Interval Grey Numbers

Table 2. Class for representing grey numbers

```
class GreyNumber(object):
    def __init__(self, low, up):
        self.low = low
        self.up = up
```

Table 3. Function for comparing two grey numbers

```
def probability_greater(a : GreyNumber, b : GreyNumber) ->
float:
    '''returns the probability that a > b'''
    if a.low > b.up:
        print("a.low > b.up")
        return 1
    elif b.low > a.up:
        print("b.low > a.up")
        return 0
    elif a.low < b.low < a.up < b.up:
        print("a.low < b.low < a.up < b.up")
        return (a.up-b.low)**2 / (2 * (a.up - a.low) * (b.up -
b.low))
    elif b.low < a.low < b.up < a.up:
        print("b.low < a.low < b.up < a.up")
        return 1 - (b.up - a.low)**2 / (2 * (a.up - a.low) *
(b.up - b.low))
    elif b.low < a.low < a.up < b.up:
        print("b.low < a.low < a.up < b.up")
        return (a.low + a.up - 2 * b.low)  / (2 * (b.up -
b.low))
    elif a.low < b.low < b.up < a.up:
        print("a.low < b.low < a.up < b.up")
        return (2 * a.up - b.up - b.low) / (2 * (a.up -
a.low))
    else:
        raise Exception("Sorry, this case is not supported")
```

Comparing Interval Grey Numbers

Example 2. Let the two overlapping interval grey numbers $\otimes_1 = [1,5]$ and $\otimes_2 = [3,6]$. If $f(x) = (x-3)^2$ is the probability distribution of $\otimes_1$ and $f(y) = 1$ is the probability distribution of $\otimes_2$. Compare $\otimes_1$ and $\otimes_2$.

Comparing Interval Grey Numbers

Example 2. Let the two overlapping interval grey numbers $\otimes_1 = [1,5]$ and $\otimes_2 = [3,6]$. If $f(x) = (x-3)^2$ is the probability distribution of $\otimes_1$ and $f(y) = 1$ is the probability distribution of $\otimes_2$. Compare $\otimes_1$ and $\otimes_2$.

This example belongs to the second form

The joint probability density function is $f(x,y) = f(x) \cdot f(y) = (x-3)^2$

$$\sigma = \int\int_{D_1+D_2} f(x,y)dxdy = \int_3^6 \int_1^5 (x-3)^2 dxdy = 16,$$

$$P(\otimes_1 > \otimes_2) = \int_{a_2}^{b_1} \int_{a_2}^{y} f(x,y)dxdy \Big/ \sigma = \frac{1}{16} \int_3^5 \int_3^y (x-3)^2 dxdy = 0.0833.$$

Then $\otimes_1 \underset{0.0833}{>} \otimes_2$ or $\otimes_1 \underset{0.9167}{<} \otimes_2$. That is to say $\otimes_1$ is greater than $\otimes_2$ with the probability 0.0833 or $\otimes_1$ is less than $\otimes_2$ with the probability 0.9167.

Comparing Interval Grey Numbers

Example 3. Let the two inclusion interval grey numbers $\otimes_1 = [2,8]$ and $\otimes_2 = [3,6]$. If the probabilities of any two values in the value-covered set of $\otimes_1$ and $\otimes_2$. Compare $\otimes_1$ and $\otimes_2$.

Comparing Interval Grey Numbers

Example 3. Let the two inclusion interval grey numbers $\otimes_1 = [2,8]$ and $\otimes_2 = [3,6]$. If the probabilities of any two values in the value-covered set of $\otimes_1$ and $\otimes_2$. Compare $\otimes_1$ and $\otimes_2$.

This example belongs to the fifth peculiar form

$$P(\otimes_1 > \otimes_2) = (2b_1 - b_2 - a_2)(b_2 - a_2)/[2(b_1 - a_1)(b_2 - a_2)] = 21/36 = 0.5833.$$

Then $\otimes_1 \underset{0.5833}{>} \otimes_2$ or $\otimes_1 \underset{0.4167}{<} \otimes_2$. That is to say $\otimes_1$ is greater than $\otimes_2$with the probability 0.9833 or $\otimes_1$ is less than $\otimes_2$ with the probability 0.4167.

Next

3. Practical Application

While classical machine learning and deep learning approaches are increasingly used for sentiment analysis, lexicon-based ones have the advantage of providing reasonably good results without requiring the collection and annotation of large amounts of training data. Thus, they can be considered the preferred approach when the volume or the quality of the training data is not ad-equate for machine learning approaches [22].

Some of the most important factors that influence the performance of lexicon-based methods are the coverage of the lexicon and the accuracy of the affective values associated with the tokens. As highlighted in Table 1 sentiment lexicons can be constructed using both automatic and manual approach-es. While automatic approaches can have a very good coverage, containing large numbers of words and expressions, they are also commonly very noisy, which can negatively affect the performance of the sentiment classification algorithms [3]. On the other hand, manually crafting highly accurate lexicons with the help of independent human raters can be a tedious task, due to which, most manually created lexicons have limited coverage.

In this context, an approach based on grey numbers theory is proposed in Cotfas et al. [3], which investigates how existing lexicons can be combined in order to improve both coverage and accuracy. The approach heavily relies on the introduction of grey sentiment lexicons. Compared to traditional sentiment lexicons that assign a single numeric value to each token, a grey sentiment lexicon characterizes the affective value of the tokens using grey numbers [3]. The lexicons considered in the study are Maxdiff-Twitter, VADER and Sentiment140. Since word polarities can vary significantly across domains, the three lexicons have been chosen from the same domain, namely mi-croblogging. As shown in Table 4, the number of tokens and the ratio be-tween the number of negative and positive tokens varies greatly.

The Maxdiff lexicon [7] contains 1515 tokens and has been created manually through crowdsourcing using the Maxdiff approach. The polarity values of the tokens vary between [-1], for extremely negative, to [+1] for extremely positive. On the other hand, the VADER lexicon [4] contains approximately 7500 tokens and has been manually created with the help of 10 independent human raters, that have received training in order to accurately evaluate the tokens on a scale from [-4], corresponding to extremely negative, up to [+4], corresponding to extremely positive. The resulting lexicon includes only the tokens for which the mean sentiment rating was non-zero and the standard deviation of the ratings was below 2.5. Finally, the Sentiment140 lexicon [13] has been created in an automatic manner, starting from the emoticons contained in tweets. While the intensity of the words varies between minus infinity and plus infinity, save for a few exceptions, all the tokens have a polarity between [-5] and [5]. Out of the three lexicons, only Maxdiff includes a limited number of neutral terms.

https://github.com/cjhutto/vaderSentiment

Table 4. Comparison between the number of positive and negative tokens

	Negative	Neutral	Positive	Total
Maxdiff	726	14	775	1515
VADER	4173	-	3344	7517
Sentiment140 (unigrams)	24,154	-	38,311	62,465

The following steps have been taken by Cotfas et al. [3] in order to create the grey sentiment lexicons by combining several classical ones. First, the values in all the classic lexicons are normalized in the interval [-1,1]. Afterwards the grey sentiment value for each token is determined as the interval between the minimum and the maximum values with which the token appears in the considered classical lexicons.

Table 5. Grey sentiment values for several affective words

	Maxdiff-Twitter	VADER	Sentiment140	Grey Sentiment
horrible	-0.89	-0.625	-0.3902	[-0.8900, -0.3902]
bad	-0.5	-0.625	-0.2594	[-0.6250, -0.2594]
relaxed	0.688	0.55	0.2464	[0.2464, 0.6880]
happy	0.734	0.675	0.2392	[0.2392, 0.7340]
perfect	0.766	0.675	0.1958	[0.1958, 0.7660]
best	0.812	0.8	0.1574	[0.1574, 0.8000]

Four lexicons have been constructed in total by Cotfas et al. [3], three of them combining two by two the classical lexicons and a fourth one combining all three classical lexicons. In order to validate the benefits of grey sentiment lexicons, in comparison with the classical ones, in the following the base-line grey sentiment analysis algorithm proposed in [3] has been applied. The algorithm follows the approach proposed by Hutto and Gilbert [4], but without implementing the heuristics, such as the ones related to negations and booster words. While implementing also the heuristics would have resulted in a better classification performance, the base-line algorithm can nevertheless provide a relatively fair comparison of the analyzed lexicons.

During the first step of the algorithm, the tweet is divided into separate units, called tokens, which correspond to the words, mentions and hashtags in the tweet. All the resulting tokens are then converted to lowercase, in order to facilitate the identification of the tokens that are also present in the lexicon. Afterwards, the grey polarity scores for these tokens are added using the addition operation for grey numbers in order to compute the overall grey sentiment score of the tweet. As mentioned by Cotfas et al. [3], the grey scalar product could be used to increase the importance of tokens following a booster word, such as “extremely”.

In the following, the approach will be evaluated using the STS-Gold dataset [23] that contains 2034 tweets, out of which 632 are classified as positive tweets and 1402 are marked as negative tweets. The evaluation has been performed in terms of precision, recall and accuracy, computed from the True Positives (TP), True Negatives (TN), False Positives (FP) and the False Negatives (FN). The TP represents the number of real positive tweets classified as positive, while the TN contains the number of negative tweets correctly classified as negative. On the other hand, FP represents the number of real negative tweets classified incorrectly as positives, while FN is the number of real positive tweets incorrectly classified as negative. The accuracy is computed as the ratio between the correctly classified tweets (TP+TN) and the total number of tweets (TP+TN+FP+FN). The precision metric quantifies the ratio between the correctly classified positive tweets (TP) and the overall number of tweets that have been classified as positive (TP+FP). The recall on the other hand, conveys the ratio between the correctly classified positive tweets and the total number of positive tweets.

The performance achieved by the base-line sentiment analysis algorithm us-ing the classical lexicons has been evaluated first. As it can be noticed in Ta-ble 6, the best results in terms of precision (75%) and accuracy (85%) have been achieved using the Sentiment140 lexicon. The best results in terms of recall (79%) have been obtained in the case of the Maxdiff lexicon. The worst values in terms of precision and recall have been encountered in the case of the VADER lexicon, while the accuracy was equally bad in compari-son to the Maxdiff lexicon. Arguably, since the automatically created Senti-ment140 lexicon contains far more tokens than the other two manually creat-ed lexicons, it has a better coverage of the words that appear in tweets and can provide a better classification.

3. Practical Application

Sentiment Analysis with Grey Numbers

Table 6. Classification performance using the initial lexicons

	Maxdiff	VADER	Sentiment140
True Positives	498	300	487
True Negatives	457	620	1238
False Positives	945	782	164
False Negatives	134	252	145
Precision	35%	28%	75%
Recall	79%	54%	77%
Accuracy	47%	47%	85%

Four grey sentiment lexicons have been created next that contain all the terms in the classical lexicons from which they have been created. As it can be noticed from Table 7, the grey sentiment lexicon created by keeping all the terms from VADER and Sentiment140 provides good results when con-sidering all the metrics. The accuracy for this lexicon (70%) is the best among the four considered grey sentiment lexicons and is only exceed by the one of the Sentiment140 lexicon (75%). The same holds true for accuracy (84%), which is the best among the considered lexicons, except for the Sentiment140 lexicon (85%). However, the recall of the lexicon (82%) is better than any of the ones achieved by the classical lexicons. The best recall value (88%) has been achieved by the grey sentiment lexicon created from all the terms in the classical lexicons, but this lexicon provides relatively low values for the pre-cision and recall metrics.

Additionally, it can be noticed that the lexicon created by combining all the tokens in the Maxdiff and VADER lexicons provides better performance in terms of precision, recall and accuracy than any of the two initial lexicons.

3. Practical Application

Sentiment Analysis with Grey Numbers

Table 7. Evaluation of the grey sentiment lexicons that contain all the tokens

	Maxdiff ∪ VADER	Maxdiff ∪ Sentiment140	VADER ∪ Sentiment140	Maxdiff ∪ VADER ∪ Sentiment140
True Positives	515	549	521	558
True Negatives	585	1010	1178	1017
False Positives	817	392	224	385
False Negatives	117	83	111	74
Precision	39%	58%	70%	59%
Recall	81%	87%	82%	88%
Accuracy	54%	77%	84%	77%

The case in which only the common terms in the initial lexicons are kept in the grey lexicon has also been analyzed, as highlighted in Table 8. It can be noticed that the results in terms of precision, recall and accuracy of all the lexicons in Table 8 is worse than those of the corresponding lexicons created by keeping all the terms from Table 7. Arguably, since these lexicons are cre-ated by keeping only the common tokens, they are smaller and thus have a reduced coverage of the words that appear in the analyzed tweets.

Table 8. Evaluation of the grey sentiment lexicons that contain only the common tokens

	Maxdiff ∩ VADER	Maxdiff ∩ Senti-ment140	VADER ∩ Sen-timent140	Maxdiff ∩ VADER ∩ Sentiment140
True Posi-tives	318	492	32	321
True Nega-tives	436	633	76	507
False Posi-tives	966	769	1326	895
False Nega-tives	314	140	600	311
Precision	25%	39%	2%	26%
Recall	50%	78%	5%	51%
Accuracy	37%	55%	5%	41%

3. Practical Application

Sentiment Analysis with Grey Numbers

In the following, the grey lexicon created by combining all the terms in the VADER and Sentiment 140 lexicon, that provides good result for all the con-sidered metrics, will be used to determine the polarity of several tweets.

For the tweet with the id 126491084854530049 "I love Ice Cream Sand-wich :) #Android #Google #Samsung" the overall polarity can be computed as shown in Table 9. The tokens that have been identified in the grey lexicon are "love" with grey sentiment polarity [0.2166, 0.8] and ":)" with a grey sen-timent polarity of [0.5, 0.5]. Using the addition operation for grey numbers, the overall polarity of the tweet is [0,8956, 1,479]. Given the fact that the po-larity is greater than 0, the tweet has a positive connotation. In the case of this tweet we can notice also a shortcoming of the base-line sentiment algorithm, which treats "Ice Cream Sandwich" as independent words and not as a group of words referring to a version of the Android operating system. However, this does not impact the results of the classification process, as even if we would disregard the values associated with the tokens "ice", "cream" and "sandwich", the tweet would still be recognized as positive, with an overall sentiment polarity of [0,7738, 1,3572].

3. Practical Application

Sentiment Analysis with Grey Numbers

Table 9. Computing the polarity for the tweet with the id 126491084854530049

Token	**Operation**	**Grey Sentiment Polarity**
i		
love	+	[0.2166, 0.8]
ice	+	[0.067, 0.067]
cream	+	[0.0574, 0.0574]
sandwich	+	[-0.0026, -0.0026]
:)	+	[0.5, 0.5]
#android	+	[0.058, 0.058]
#google	+	[-0.0008, -0.0008]
#samsung		
Overall Polarity	=	[0.8956, 1.479]

In the case of the tweet with the id 126519329025040384 , “Ice Cream Sandwich to stop carriers bullying smartphone users #google #android http://t.co/BZNy74Nn”. Several tokens have been found in the grey sentiment lexicon, most of them having a negative signification associated. Among them, we can recognize “bullying”, which is profoundly negative, with a grey sentiment polarity of [-0,725, -0,2022]. The overall polarity of the tweet is [-1,0956, -0,3456], as shown in Table 10, which indicates a negative tweet.

Table 10. Computing the polarity for the tweet with the id 126519329025040384

Token	Operation	Grey Sentiment Polarity
ice	+	[0.067, 0.067]
cream	+	[0.0574, 0.0574]
sandwich	+	[-0.0026, -0.0026]
to	+	[-0.033, -0.033]
stop	+	[-0.3, -0.0728]
carriers	+	[-0.2226, -0.2226]
bullying	+	[-0.725, -0.2022]
smartphone	+	[0.0298, 0.0298]
users	+	[-0.0238, -0.0238]
#google	+	[-0.0008, -0.0008]
#android	+	[0.058, 0.058]
http://t.co/BZNy74Nn		
Overall Polarity	=	[-1.0956, -0.3456]

The tweet with the id 126292335540699136 has the content "@Apple iTunes is the worst program ever. For such a great phone, you make some awful software.". Many of the tokens in the tweet are found in the lexicon, some with a positive polarity and others with a negative one, such as "worst" and "awful". The overall grey polarity of the tweet is [-1.0634, 0.0172], as shown in Table 11, which captures the fact that even though the tweet is es-sentially a negative one, it also mentions some positive aspects.

Table 11. Computing the polarity for the tweet with the id 126292335540699136

Token	**Operation**	**Grey Sentiment Polarity**
@apple		
itunes		
is	+	[-0.023, -0.023]
the		
worst	+	[-0.775, -0.375]
program	+	[0.0192, 0.0192]
ever	+	[-0.0024, -0.0024]
for		
such	+	[0.0054, 0.0054]
a		
great	+	[0.2354, 0.775]
phone	+	[-0.1542, -0.1542]
you	+	[0.1352, 0.1352]
make	+	[0.014, 0.014]
some	+	[0.077, 0.077]
awful	+	[-0.5, -0.359]
software	+	[-0.095, -0.095]
Overall Polarity	=	[-1.0634, 0.0172]

For the tweet with the id 1258400390317383368 , having the content "@apple your simply the best.", the overall grey polarity is positive one, as highlighted in Table 12, given the fact that all the matched tokens have a positive connotation.

Table 12. Computing the polarity for the tweet with the id 1258400390317383368

Token	Operation	Grey Sentiment Polarity
@apple		
your	+	[0.161, 0.161]
simply	+	[0.1316, 0.1316]
the		
best	+	[0.1574, 0.8]
Overall Polarity	=	[0.45, 1.0926]

Given the fact the overall polarity of the tweets is expressed using grey numbers, one can compare them. This aspect might be useful in real-life ap-plications, as it might offer a better overall view of the people's opinion re-garding certain topics. The comparison approach will be showcased using the tweets in Table 9 and Table 12.

As the tweet in Table 9 (noted tweet1) has a polarity score of [0.8956, 1.479], while the one in Table 12 (noted tweet2) has an overall polarity of [0.45, 1.0926], the probability that the positivity of tweet1 is greater than that of tweet2 is approximately 0.95, as shown in Fig . 38.

```
a = GreyNumber(0.8956, 1.479) # change these values
b = GreyNumber(0.45, 1.0926) # change these values
# a > b
probability_greater(a, b)
```

```
b.low < a.low < b.up < a.up
0.9482398756935448
```

Fig . 38. Comparison between the polarity of tweet1 and that of tweet2

While the creation of grey sentiment lexicons has been evaluated by Cotfas et al. [3] using the Sanders dataset, in the present chapter we have chosen to perform the evaluation using the STS-Gold dataset [23]. Even in the case of this dataset it has been shown that the grey lexicons can exceed the performance of the initial lexicons.

The grey lexicons have afterwards been used in the context of a base-line sentiment analysis algorithm to perform polarity analysis, during which the tokens in the analyzed text are compared with the ones in the lexicon. The values in the lexicon for the tokens that appear in the text are added using the grey addition operation. The algorithm can be improved by implementing heuristics, such as the ones proposed in the VADER algorithm. Booster words can be taken into account and their impact can be quantified using the grey scalar product operation. Finally, the comparison between grey numbers has been used to evaluate whether a tweet has a more positive connotation than another one.

Interreg

Co-funded by the European Union

GOVERNMENT OF ROMANIA

HUNGARIAN GOVERNMENT

Romania - Hungary

Name: **Prof.univ.PhD. Camelia Delcea**

Institution: **Bucharest University of Economic Studies**

E-mail address: **camelia.delcea@yahoo.com**

Thank you for your attention!