European Commission

JRC TECHNICAL REPORTS

Modelling for EU Policy support: Impact Assessments

Analysis of the use of models in European Commission Impact Assessments in 2003-2018

Acs, S., Ostlaender, N., Listorti, G., Hradec, J., Hardy, M., Smits, P., Hordijk, L.

2019

EUR 29832 EN

EU Science Hub
https://ec.europa.eu/jrc

#EUModelling

JRC117250

EUR 29832 EN

PDF ISBN 978-92-76-09671-9 ISSN 1831-9424 doi:10.2760/748720

Luxembourg: Publications Office of the European Union, 2019

How to cite this report: Acs, S., Ostlaender, N., Listorti, G., Hradec, J., Hardy, M., Smits, P., Hordijk, L., *Modelling for EU Policy support: Impact Assessments*, EUR 29832 EN, Publications Office of the European Union, Luxembourg, 2019, ISBN 978-92-76-09671-9, doi:10.2760/748720, JRC117250

Contents

Executive summary

The objective of this work is to **systematically analyse** how **models** are used in support to the **policy formulation phase** of the EU policy cycle. We focus on European Commission (EC) **Impact Assessments** (IAs).

The main framework of the EU regulatory policy, the **Better Regulation** (BR) **Agenda** (European Commission 2015), sets a clear commitment to a transparent and sound use of evidence for all EU policy making activities. The Better Regulation Guidelines (European Commission 2017), which complement the Agenda to provide concrete guidance throughout the policy cycle, recommend to quantify costs and benefits to the extent possible to support the policy formulation phase. In doing so, the EC makes extensive use of **models**. A better understanding of these models and how they are used can then contribute to a sound use of evidence in support to EU policies.

The Commission's **Competence Centre on Modelling** (CC-MOD) promotes a responsible, coherent and transparent use of models at the EC. As part of its activities, this analysis systematically investigates how models are used in support to the policy formulation phase, by looking at EC IAs which are publicly available. A total of **1063 IAs** carried out in the years **2003 to 2018** have been investigated to examine the frequency and characteristics of model use, by using text mining techniques complemented by manual post processing. The research is facilitated by and feeds back into **MIDAS**, the Commission-wide Modelling Inventory and Knowledge Management System developed and managed by CC-MOD (Ostlaender *et al.* 2019), which directly contributes to enhanced transparency and traceability of models used to support policies.

Our **results show that models** are **used in 16% of the total IAs** (173 out of 1063 IAs), with a positive trend over time, starting with only two IAs using models in 2004, to around **25-30% from 2015** onwards.

We identified **123 different models** contributing to IAs. These models have been developed or run by the EC or by third parties. **More than half (53%, or 65) of these models were used only once**, which leads to considerations related to the efficiency of model use and reuse in the EC, as well as on the scope for improved coordination of model related products and services. On the other hand, some models do dominate: the top **10 models contributed to 10 or more IAs**, and **were used in 66% (or 114)** of the total number of **IAs** using models.

Included among the top 10 models are also those used for the development of the series of **EU Reference Scenarios** ([1]), led by DG ENER, DG MOVE and DG CLIMA. Indeed, the consistent use of the same series of baselines in the areas of energy, transport and climate goes in the direction of increased consistency across policy areas and fields of analysis, though there is still room for strengthening the coherence in baselines across the whole EC (Marques *et al.* (2017). At the same time, given the dominance of these models to support the policy formulation phase, it is essential that they undergo careful quality scrutiny and that maximum transparency and traceability of results is ensured.

Policy areas with the highest number of IAs using models are **environment (including climate), internal market, transport and energy**. However, this could also reflect other factors, such as the frequency at which IAs are carried out in the various policy areas. In addition, it should be remembered that models can also be used in other policy relevant studies, or for internal notes and analyses which remain unpublished.

Finally, results show that **94% (116) of the total number of models** that were used in IAs were used for **ex ante assessment of policy options**. This is to be expected, since indeed the assessment of policy options is an extremely relevant task for quantitative analysis in IAs. As mentioned, most of the top 10 models were also used for baselines.

Transparency is crucial to understand how models work and to validate their behaviour, to encourage their sound and widespread use in support to policy making contributing also to a more effective and efficient use of resources for model development and use within the EC. The **requirement** introduced by the BR Agenda in 2017 for an annex on 'Analytical models used in the preparation of the impact assessment' contributed to better model descriptions and an increased transparency of the methodology used. There is, however, still room for improvement in better documenting models as the individual model descriptions in IAs are still of different quality. The information and reports generated by MIDAS on models and model use can be used in this respect.

([1]) https://ec.europa.eu/energy/en/data-analysis/energy-modelling/eu-reference-scenario-2016

We also identified some major challenges related to **referencing in IAs**, such as lack of harmonized and adequate references or outdated hyperlinks, which made the quantitative evidence untraceable in several IAs. Since 2017, the BR foresee that, when IA analysis relies on modelling or the use of analytical methods, the model should be documented in the corporate modelling inventory MIDAS (European Commission 2017). This represents a major step forward in terms of transparency. At the same time, it is also clear that **further action** is needed to use and promote best practices to ensure transparency and accessibility over time of the evidence base in support to IAs. In addition to the BR guidelines, the **JRC**, as the science and knowledge in house service of the Commission, can provide **additional assistance and support** to the Policy DGs.

To conclude, the results of our analysis **contribute to the BR Agenda** by highlighting aspects of transparency, coherence, traceability and accountability in the use of evidence for EU policy making.

Authors

Szvetlana Acs, Nicole Ostlaender, Giulia Listorti, Jiri Hradec, Matthew Hardy, Paul Smits, Leen Hordijk

1 Introduction

The Better Regulation Agenda (BR Agenda), adopted by the European Commission (EC) in 2015 (European Commission 2015), is the main framework for the current EU regulatory policy. The BR Agenda sets a clear commitment to a transparent and sound use of evidence for all EU policy making activities. The Agenda is complemented by the Better Regulation Guidelines (European Commission 2017), which provide concrete guidance throughout the policy cycle to ensure robust evidence based decision making.

The BR Guidelines on impact assessments **of the various policy options** recommend to **quantify** costs and benefits to the extent possible (European Commission 2017). In this respect, the EC makes use of **models** (see Box 1). A better understanding of these models and how they are used can lead to improving the transparency, coherence and efficiency of model development and use, thus contributing to a sound use of evidence in support to EU policies.

Box 1. Definition of model

For the present analysis, the following working definition is adopted: a model can be defined as an analytical representation or quantification of a real-world system, used to make projections or to assess the behaviour of the system under specified conditions ([2]).

Our objective is to systematically analyse how models are used in support to the EU policy cycle. We start from the policy formulation phase, namely EC Impact Assessments (IAs). A total of 1063 IAs carried out in the years 2003 to 2018 have been investigated by using Text Mining techniques complemented by thorough investigation of relevant documentation such as supporting studies.

The analysis has been carried out by the **EC Competence Centre on Modelling** (CC-MOD), which directly contributes to the BR Agenda and promotes a responsible, coherent and transparent use of modelling to underpin the evidence base for EU policies. The research is facilitated by and feeds back into **MIDAS**, the Commission-wide Modelling Inventory and Knowledge Management System which is under the responsibility of CC-MOD. MIDAS contributes to the BR Agenda by enhancing the transparency of the use of models in support to policy making and the traceability of their results.

This report presents the methodology and a first set of results of the analysis, namely concerning frequency and characteristics of model use in IAs. Results contribute to improving the efficiency of model development and use in support to EU policymaking.

We focus on IA reports which formally support the Commission policy formulation phase, usually published as Commission Staff Working Documents (SWD). To our knowledge, this report presents a first comprehensive analysis, using highly advanced tools. At the same time, we are aware that our analysis still provides a somehow partial perspective; future research will focus on the analysis of model use in other policy relevant documents for ex ante impact assessments, as well as on other phases of the policy cycle.

This report is organised as follows: in Chapter 2, background information and definitions are summarized. The methodology used for the analysis, including the preparation of the underlying knowledge base, is outlined in Chapter 3. The results of the analysis of model use in IAs are presented in Chapter 4, while discussion and concluding remarks are provided in Chapter 5. Finally, in Chapter 6 we provide an outlook on the ongoing further analysis for other phases of the policy cycle.

([2]) Please note we consider as 'model' only what correspond to this definition. We do not include in our analysis what could be indicated as a 'model', but does instead refer to a methodology or another meaning, such as for example 'standard cost model', 'lifecycle model', 'analytical model', and so on.

2 Background

This section provides background information, concepts and references which will be used in the following chapters. These refer in particular to:

- IAs in the context of the BR Agenda (section 2.1);
- aim and structure of the IAs reports, and main features of the EUR-Lex repository (section 2.2);
- MIDAS, the Modelling Inventory and Knowledge Management System of the EC (section 2.3);
- the Semantic Text Analysis Tool SeTA (section 2.4).

2.1 Impact Assessments for policy formulation

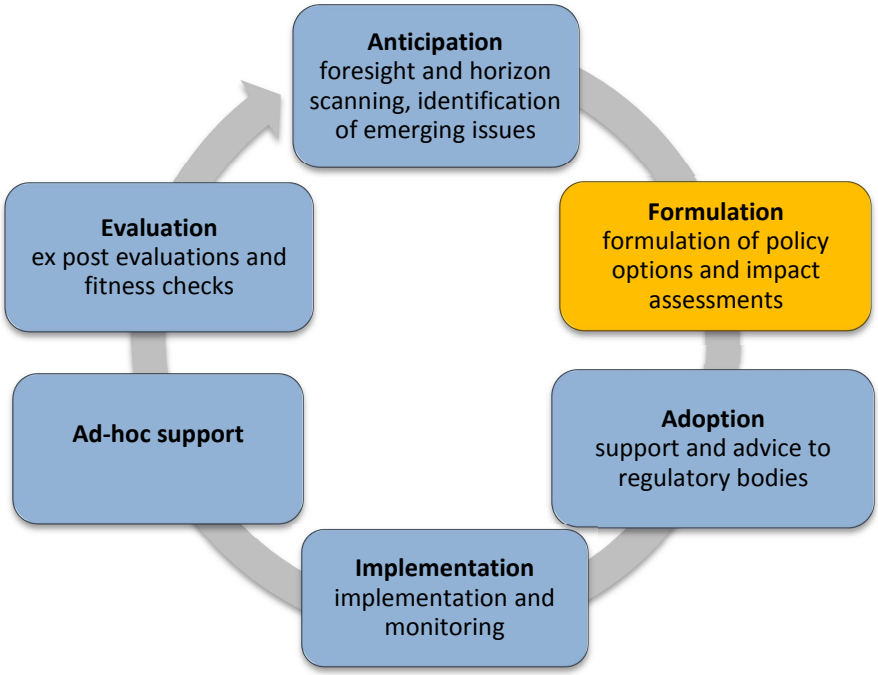
The Better Regulation Guidelines and Toolbox (European Commission 2017) complement the BR Agenda (European Commission 2015) ([3]) and provide concrete guidance and information on the tools to be used to ensure robust evidence based decision making throughout the policy cycle.

Impact Assessments refer to the ex-ante analysis carried out in the policy formulation phase of the policy cycle (Figure 1) ([4]). According to the BR Guidelines, 'The impact assessment process is about gathering and analysing evidence to support policymaking. It verifies the existence of a problem, identifies its underlying causes, assesses whether EU action is needed, and analyses the advantages and disadvantages of available solutions.' (European Commission 2017: 15).

([3]) For a comprehensive review of the literature debate on the BR Agenda see Listorti *et al.* (2019).

([4]) Naundorf and Radaelli (2017) define regulatory impact assessment (RIA) as 'a systematic, comparative appraisal of how proposed primary and/or secondary legislation might affect stakeholders, society, economic sectors and the environment. In its ex post version, RIA is an appraisal of how existing rules have affected stakeholders, society, the economy and the environment. [...] "Systematic" means coherent-not episodic or random. "Comparative" means that more than one option is appraised, including the option of not altering the status quo (baseline)' (2017:189). In the present analysis, in accordance with the BR Agenda, we use the term 'impact assessment' to refer to ex ante RIA.

Figure 1. Scientific support within the EU Policy Cycle



Source: own elaboration, based on European Commission (2017:5, Figure 1).

The first IA was registered as a Commission SWD in EUR-Lex back in 2003 in the context of the Interinstitutional agreement on better law-making (European Parliament, European Council and European Commission, 2003) ([5]). Since then, IAs have been further extended and consolidated. In the analysis of the impacts of the various policy options, the BR Guidelines recommend quantifying costs and benefits to the extent possible (European Commission 2017). In this respect, the EC makes use of simulation models (see Box 1) to assess the environmental, economic, and social impacts of policies.

2.2 The Impact Assessment Report

Our analysis is based on IA reports, which present the final results of the impact assessment procedure and accompany the draft initiative through the Commission decision-making process.

2.2.1 Structure

Throughout the IA report, conclusions should be substantiated with evidence (e.g. data, estimations, scientific findings) together with appropriate citations. If not possible, the reason should be explained (European Commission 2017).

([5]) It emphasized that "the Commission will continue to implement the integrated advance impact-assessment process for major items of draft legislation, combining in one single evaluation the impact assessments relating inter alia to social, economic and environmental aspects" (European Parliament, European Council and European Commission, 2003:4)

Already since 2015, the BR Guidelines require that the models used are described in a mandatory Annex. Since the revision of the Better Regulation Guidelines and Toolbox in 2017, the report has to follow the specific structure detailed in Box 2.

Reports prior to 2015 or 2017, respectively, are less consistent in their structure.

Box 2. Structure of the Impact Assessment report according to the BR Guidelines

Section 1. Introduction: Political and legal context

Section 2. What is the problem and why is it a problem?

Section 3. Why should the EU act?

Section 4. What should be achieved?

Section 5 What are the various options to achieve the objectives?

Section 6. What are the impacts of the different policy options and who will be affected?

Section 7. How do the options compare?

Section 8. The preferred option

Section 9. How would actual impacts be monitored and evaluated?

Mandatory annexes:

Annex 1: Procedural information

Annex 2: Stakeholder consultation

Annex 3. Who is affected by the initiative and how?

Annex 4. Analytical methods used in preparing the impact assessment (incl. models)

Once completed, the IA report is published as a SWD of the EC, together with an Impact Assessment Summary, and the legislative proposal(s) or document(s) it accompanies. The Publications Office of the European Union (OP), whose mandate is to publish and disseminate the publications of institutions and other bodies of the European Union, makes these documents available on EUR-Lex. A copy is usually also made available on the Impact Assessment website of the Secretariat General ([6]).

2.2.2 Access through EUR-Lex

EUR-Lex is a website that provides free access to The Official Journal of the European Union, EU case law and other resources for EU law, starting from 1951. Each entry in EUR-Lex has a dedicated metadata record. The analysis described in this report relies in particular on the following subset of the available metadata:

- Title and reference
- Date of document: date when the document was published
- Descriptors: subject matter (Box 3)
- Author: institution or department that authored the document
- Department responsible (only available after 2015): this describes the lead DG or 'chef de file'
- Form: impact assessment

([6]) https://ec.europa.eu/info/law/law-making-process/planning-and-proposing-law/impact-assessments_en

Box 3. Subject matter

"The subject matter classification shows the concepts used for the indexation of notices published on EUR-Lex. The subject matter classification exists in parallel with two other classifications: the Directory of European Union Legislation and EuroVoc. Differently from these, the subject matter classification is aligned strictly to the evolution of European Union policies cited in the different treaties of the European Union that can be consulted at: http://eur-lex.europa.eu/collection/eu-law/treaties.html?locale=en. The content of the subject matter classification is managed by the GIL working group (Inter-institutional Group on Legislation)."

Source: *OP 2019: Subject matter (Asset). Name authority list. Last access: 2019-30.04. URI: http://publications.europa.eu/resource/dataset/subject-matter*

2.3 The Modelling Inventory and Knowledge Management System MIDAS[7]

In 2017, the EC launched the *Competence Centre on Modelling* (CC-MOD), run by the Joined Research Centre of the European Commission (JRC). Directly contributing to the BR Agenda, CC-MOD promotes a **responsible**, **coherent** and **transparent** use of modelling to underpin the evidence base for EU policies, and pools the Commission's competencies and best practices in building and using models. CC-MOD further helps identifying common approaches to quality and transparency of model use, and establishes a Community of Practice on Modelling ([8]).

Maintaining an overview of ongoing modelling activities, by documenting the models and model combinations in use across the EC, is an elementary first step for a more transparent and coherent use of models in the policy cycle. It is, however, also a major challenge: for example, in the EC more than 150 models are currently being used, with domains ranging from greenhouse gas emissions, energy consumption and economy, to agriculture and structural integrity assessment, to name but a few. In addition, the majority of these models are run in combination with other models, together with the related input datasets and assumptions. Thus, they form complex networks of dependencies and interaction.

In order to address these challenges, CC-MOD is responsible for the development and management of the **Modelling Inventory and Knowledge Management System of the European Commission,** MIDAS ([9]), which captures this knowledge and makes it available to modellers and policy makers.

MIDAS contains the descriptions of models in use by the EC which directly or indirectly support the policy cycle, independently of whether they were developed by the EC or by third parties, such as national or international research institutes, companies or consortia. MIDAS includes information on main purpose of the model, model type, ownership and licensing, as well as model structure and approach, inputs, model parametrisation and outputs, and information about *Model Quality Assurance* (MQA) ([10]), an activity which is supported by the CC-MOD Team SAMO ([11]). MIDAS furthermore collects details about the EC departments and external institutes that are developing or actively running the models on behalf of the EC, and the documents describing the models and their use.

MIDAS also describes the potential use of the model for policy making, including the types of impacts a model can help to assess when contributing to an IA. In addition, it also captures information about the actual use of these models in the context of IA and studies:

— **Impact Assessments**: MIDAS includes ongoing or completed IAs where modelling results were used. MIDAS collects information about the IA itself, and about the models contributing to the IA. The latter might be done also by referring to a study that describes the modelling exercise used by the IA.

([7]) This section is based on Ostlaender *et al.* (2019).

([8]) CC-MOD contributes to the implementation of the Better Regulation policy, the Interinstitutional Agreement on Better Law Making, and the Communication on the Management of Data, Information, and Knowledge at the Commission. Information on the Competence Centre is available at https://ec.europa.eu/jrc/en/modelling

([9]) MIDAS can be accessed by EC services at http://midas.jrc.cec.eu.int. From 2019 onwards parts of the system are open to the European Parliamentary Research Service under the umbrella of the *Interinstitutional Agreement on Better Law-Making* (European Parliament Council of the European Union and European Commission 2016).

([10]) For models and modelling exercises the quality itself may be assessed. This is known under the generic term Model Quality Assurance (MQA). MQA includes a large variety of methodologies as Uncertainty Analysis and Sensitivity Analysis.

([11]) SAMO – Sensitivity Analysis of Models. More details can be found on: https://ec.europa.eu/jrc/en/samo

— **Studies**: In MIDAS, 'studies' are analytical exercises that use modelling results to answer policy-relevant questions. Adding studies to MIDAS serves the purpose of increasing transparency on how (and in which combinations) models are used to answer policy questions. This is particularly important for studies that are frequently re-used in various impact assessments, such as the *EU reference scenario 2016 Energy, transport and GHG emissions: trends to 2050* ([12]).

MIDAS is used by the entire EC, enabling users to find models in use by the EC and to assess their use for specific policy purposes. Starting in 2017, MIDAS is also **integrated in the workflow for IAs**, since the Better Regulation Toolbox requests that **any model used in IAs has to be described in MIDAS**. This makes MIDAS an important corporate tool to use, reuse and document models in a proper way, leading to the propagation of sound methodology underpinning the EC's BR Agenda and potentially to significant efficiency gains in terms of financial and personnel outlays. Since 2019 parts of the system are also open to the European Parliamentary Research Service (EPRS) to provide support during the evaluation of IAs prepared by the EC ([13]).

A detailed description of MIDAS is available in Ostlaender *et al.* (2019).

2.4 The Semantic Text Analysis Tool SeTA[14]

The Semantic Text Analysis Tool SeTA has been developed by the JRC as an Artificial Intelligence (AI) digital assistant to policy analysts. SeTA was first designed and used in the context of this analysis to gather all IA from EUR-Lex ([15]), and, among these, to identify IAs that used modelling, and the models themselves (see section 3). In addition, SeTA constitutes one of the tools that support the CC-MOD in its activities, for example to analyse and answer questions related to the use of models in other phases of the policy cycle, the combined use of models for answering specific policy questions, and the development and use of models in the context of projects funded by the EU's framework programmes for research and innovation.

The system itself and this particular use case are described in detail in Hradec *et al.* (2019).

SeTA is based on a word embedding neural network, which has been trained on the European legislation and on technical/scientific reports in order to identify similarity patterns between terms. Terms are single words like 'fairness', combination of words like 'impact assessment' or model acronyms like 'GEM-E3' and 'PRIMES'. The neural network learns, in an unsupervised manner, the relations between these terms in a sentence, within a moving window, which means that it learns how the context in which the term is used can lead to its meaning, both in terms of synonyms and of the area of application or context in which it is used.

This particular neural network takes into consideration approximately 100 million sentences with 3.2 million unique words, and 700,000 phrases from 500,000 documents coming from the following sources: EU Publications ([16]), EUR-Lex, CORDIS ([17]), EU Open Data Portal ([18]) and JRC PUBSY ([19]).

The documents cover a wide range of time (starting from 1956 until today, June 2019), with an ever increasing volume per year. They were used to form a Language Model (LM), in this case the model of the EC, required for Natural Language Processing (details can be found in Hradec *et al.* (2019)).

([12]) https://ec.europa.eu/energy/en/data-analysis/energy-modelling/eu-reference-scenario-2016

([13]) The interinstitutional version of MIDAS was launched on the 7th February 2019 during the *Science Meets Parliament* event under the umbrella of the Interinstitutional Agreement on Better Law-Making(European Parliament Council of the European Union and European Commission 2016). It is accessible to all EC services and the Parliament at https://webgate.ec.testa.eu/midas-ii/. The implementation of a version of MIDAS accessible to the general public is currently under discussion.

([14]) This section is based on Hradec *et al.* (2019).

([15]) EUR-Lex provides free access to The Official Journal of the European Union, EU case law and other resources for EU law, with documents dating back as far as 1951. Accessible at https://eur-lex.europa.eu

([16]) The EU publications website provides access to reports, studies, information booklets, magazines and other publications from the EU institutions and other bodies. This was formerly known as the EU Bookshop. Accessible at: https://publications.europa.eu/en/web/general-publications/publications

([17]) CORDIS stands for Community Research and Development Information Service. It is the European Commission's primary source of results from the projects funded by the EU's framework programmes for research and innovation (Framework Programme (FP)1 to Horizon 2020). Accessible at: https://cordis.europa.eu

([18]) The EU Open Data Portal (EU ODP) is a catalogue of datasets from the EU institutions and other bodies. Accessible at http://data.europa.eu/euodp/

([19]) PUBSY is an online service giving access to data about research publications, technical reports, etc. produced by the European Commission's Joint Research Centre. It was established to assist with central storage, management and search to our research publications that go beyond the official publications stored in the EU Bookshop. PUBSY has a native API. The publicly available documents stored in PUBSY are accessible at: http://publications.jrc.ec.europa.eu/

3 Methodology

In this report, we present the first set of results of an in-depth, systematic analysis of model use in EC IAs since 2003 ([20]).

Based on previous work by Petrov *et al.* (2017), we compiled a first list of models mentioned in IAs in the years 2009 to 2014. To this, we added a number of models which were present in MIDAS at the time. However, this first picture was not yet complete, both regarding the set of IAs, and the list of models contributing to them.

Therefore, we decided to start a systematic analysis of all formal IAs carried out by the EC since 2003. This required answering the following questions:

1. Preparatory work
    a. Which IAs carried out since 2003 mention models?
    b. Which of these IAs (and models) are relevant for our analysis?
2. Analysis
    c. How are models supporting IAs?

In the remainder of the present chapter, the methodology will be presented in detail. This concerns the identification of IAs which mention models (question (a), section 3.1); the identification of IAs and models which are relevant for our analysis (question (b), section 3.2); model support to IAs (question (c), section 3.3). The results of the analysis will be presented in chapter 4.

3.1 Identification of IAs which mention models

The first step was to identify which IAs, amongst all those carried out by the EC since 2003, mention models. In fact, the models used to support IAs can be developed or run by the EC or by third parties, such as national or international research institutes, companies or consortia (see Annex 7.1).

The large number of IAs to be investigated required the adoption of a systematic approach. Therefore we decided to build a knowledge base that contained all IAs reports included in the EUR-Lex database. In fact, the IA reports contain the methodology and main findings of the IAs and should then provide a good indication on the models used (see section 2.2.1). Text mining algorithms and neural network techniques were then applied to this knowledge base to identify both the IA reports that mentioned models, and the models themselves. These were complemented by manual post-processing using additional sources.

In the remainder of this section, we present data preparation (section 3.1.1) and document search, representing the actual text mining process (section 3.1.2). This was followed by manual post-processing to identify false positives and false negatives in both the list of IAs and in the one of models (section 3.1.3).

The very demanding data preparation and document search in the context of this exercise lead to the development of SeTA, a text mining tool with a corpus of 500.000 documents, using several neural networks. This allows searching through all IAs included in EUR-Lex. Further detail can be found in section 2.4, and in Hradec *et al.* (2019).

3.1.1 Data preparation

To design the corpus to be mined, all IAs reports available from the **EUR-Lex database** were loaded and transformed into one single cleaned text corpus. This was performed by retrieving all SWD whose title contained the words 'Impact Assessment', and ignoring those that are titled 'Executive Summary' ([21]). This returned a total number of **891 IAs**. A more straightforward approach would have been to retrieve simply all documents tagged as IAs in EUR-Lex. However, this request returned a significantly lower number of documents (843 IAs), as the tagging of IAs documents in EUR-Lex turned out not to be complete.

([20]) Our analysis is limited to the policy formulation phase, and namely to IAs carried out by the EC. This means that models which are also used in the policy formulation phase, but in other types of documents which are not part of the formal EC Impact Assessment process, are not considered. In fact, future research will focus on the analysis of model use in other policy relevant documents for ex ante impact assessments as well as to other phases of the policy cycle (exploring databases such as the EU Bookshop). In addition, models can be used also for internal notes and analyses which remain unpublished.

([21]) This was important, as the majority of IA reports are accompanied by an executive summary for policy makers, which has a separate staff working document number, and would thus generate confusion about the real number of IAs.

We found out that most of the EUR-Lex metadata records provide links to multiple documents, available in various formats: mostly in PDF, but also MS Word, XML, and, in rare cases, as scanned originals. But what can be read by a human reader (e.g. several PDF/MSWord documents split into several parts) can be very inappropriate for machine processing.

Therefore, the ingestion, format conversion and text extraction took several weeks, and many critical issues had to be solved. These included, for example, footnotes becoming a part of a word: conversion mechanisms from all formats converted e.g. *Common Agricultural Policy[1] to Policy**1***. Also PDF conversions often yielded merged words or words split into characters separated by space (ENV I R ONMEN T A L P O L I CY). This particular issue was addressed by the development of a Bayesian spellchecker (Hradec *et al.* (2019).

3.1.2 Document search

The resulting cleansed text was the base for the actual document search or text mining activity.

To start with, the cleaned text was searched with a **first set of 76 models acronyms** which were fed into machine learning algorithms using both unsupervised and supervised learning (Hradec et al. 2019). This first set of models was based on Petrov et al. (2017), which performed an analysis of the use of models by the EC in IAs for the years 2009 – 2014 ([22]). During text mining:

— Frequency-based noun chunk identification was developed to distinguish model acronyms from terms that also represent common words, such as GAINS, IMPACT, SMART. For example, the phrase “GAINS model” identifies a model while "gains importance" does not.

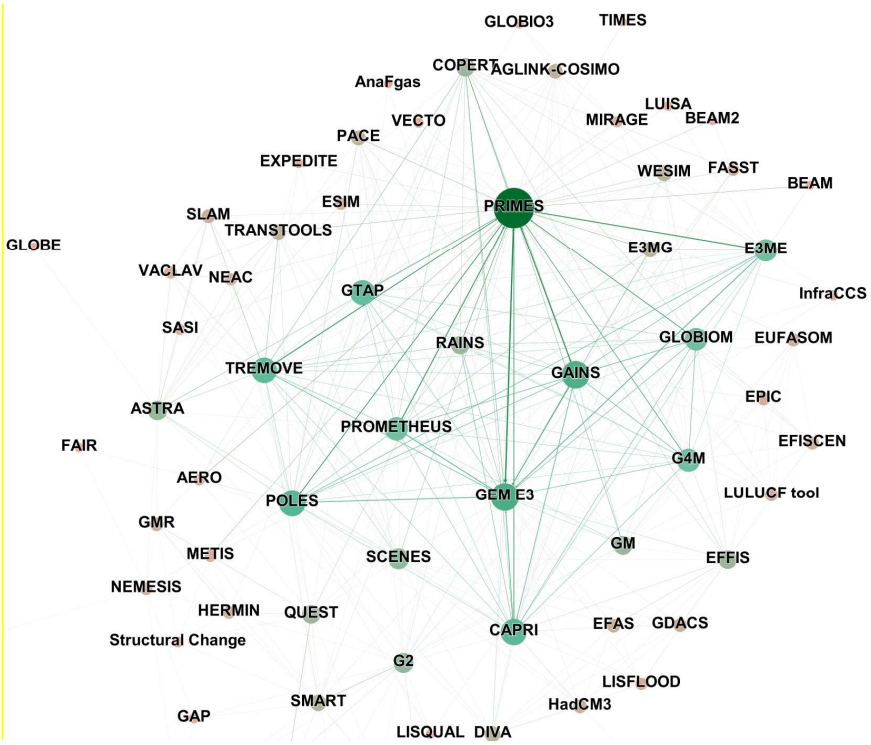
— A specific task was the utilisation of the neural word embedding to discover typos in the text of the IA reports (e.g. for GEM-E3 we got GEME3, GEM E3 and GEME 3).

Neural word embedding also allowed **discovering new models**, since deep recurrent neural network allowed clustering similar word vectors (e.g. PRIMES -> E3ME, GREEN-X, GEM-E3, TREMOVE). This allowed identifying models mentioned in the texts without any prior knowledge of their existence. This is a critical step that would otherwise have required the authors to read all IAs.

Statistics and graph analysis of occurrences (frequency of model mentioning in IA) and co-occurrences (how often model names are mentioned together in IA) was also elaborated (see Figure 2 for an example). Results provided useful insights also for the next steps.

([22]) Although the total number of models listed in the publication is larger (91), only 75 of these could be clearly identified through their acronym. The other models were indeed described through ad-hoc names describing either the area of application (like ‘biomass model’, ‘fishery model’, etc.), or the applied methodology (like ‘cge’, etc.).

Figure 2. Example of graph analysis of occurrences and co-occurrences of models mentioned in IAs as generated with the help of SeTA



3.1.3 Post processing

The results of text mining provided an initial overview of IAs mentioning models, of the frequency of these mentions in the IAs and of the co-occurrences of different models in the same IA.

To build a complete and consistent database of models and IAs, these results underwent further post-processing to ensure that the list of identified IAs and of the models they refer to was complete, and to eliminate false positives ([23]). The following additional sources were used:

— Impact Assessment website of the Secretariat General ([24]). This dedicated website lists all IAs carried out by the Commission, and was our main source to ensure that the complete list of IAs was included in our research, namely to include the IAs that our initial search in EUR-Lex had not revealed. These additional IAs had to be added manually to the knowledge base, resulting in a total of **1063 IAs** ([25]).

— LIAISE kit website ([26]), developed in the context of the FP7 LIAISE Network of Excellence with the aim to provide a library of models, publications, projects, good practices and experts to support policy impact assessments, and a community platform to collaborate in this field.

— Primary sources of the identified models, such as model websites.

Post processing also presented a number of challenges mostly due to missing or inadequate referencing, which were addressed as explained in Box 4.

Data preparation (section 3.1.1), document search (section 3.1.2) and post processing (section 3.1.3) were actually carried out iteratively, by progressively enlarging the number of IAs and models covered. Results

([23]) For example, checking whether the model name mentioned in the IA was really a model.
([24]) https://ec.europa.eu/info/law/law-making-process/planning-and-proposing-law/impact-assessments_en
([25]) On the SecGen Website we identified a total of 1069 IAs from the years 2003 to 2018. Annexes and progress reports that belonged to the same IA were not counted separately. From these 1069 we considered only 1063 IAs, as 6 IAs were not accessible (either because they were restricted or the file itself was not provided). The reasons for the gap between the returned 891 documents from EUR-Lex and the 1063 documents accessible from the IA website are manifold, and will be further investigated.
([26]) http://www.liaise-kit.eu/. Please note that the page is no longer maintained.

were recursively fed back into the text-mining process to complete the knowledge base and to improve the algorithms, which then provided new results that were again fed into the post-processing, and so on.

Box 4. Inadequate or missing references

During the post-processing, a number of challenges had to be tackled.

First of all, it was not always possible to identify models. Especially in the earlier years, in some cases no model acronym or clear reference to a specific model or study was found in the IA. Instead, generic terms like "economic model" were used.

Referencing of models and of their results was also not always traceable. In the cases where only hyperlinks were used, these were subject to a phenomenon called link-rot ([27]). As a consequence, the information referred to was untraceable, thus making it impossible to check whether the study cited by the IA actually used models. In other cases, the hyperlinks still existed, however the content accessible through them had changed, misleading the reader and making the original information untraceable.

References in many cases were also found not to follow an appropriate citation standard: especially in supporting studies, the original work was then often not identifiable. For example, many references consisted of internal or ad-hoc names, or sometimes only the projects (and not the studies) were cited.

In all the cases mentioned above, where it was possible to acknowledge that models had indeed been used, we considered the IAs as model based, although the models themselves could not be fully identified.

In some other cases the IAs were incomplete, since the main document pointed to Annexes that are not publicly available. In these cases, we assumed that no model had been used.

3.2 Definition of the final set of IAs and models

After text-mining and post processing, the resulting list of IAs and models provided the initial base of the analysis. However, it became immediately evident that the importance of a model in an IA could vary widely, from a mere being mentioned, to being the main tool used to assess the various policy options.

Therefore, for each IA, and for each of the mentioned models, it was assessed if this model was indeed really used in the IA, and what was its role.

([27]) "Link rot" (or linkrot) is the process by which hyperlinks on individual websites or the Internet in general point to web pages, servers or other resources that have become permanently unavailable. This expression also refers to the effects of failing to update out-of-date web pages that clutter search engine results. Research shows that the half-life of a random webpage is two years." Source: https://en.wikipedia.org/wiki/Link_rot

We distinguished four different **roles** that a model can play in an IA:

1. Supporting the **problem definition**. This is usually found in the first part of the IA, where the context and the motivation for taking EU action in a specific area are described.
2. Providing evidence on **ex post evaluation** and fitness checks of existing policies that are relevant for the proposal the IA accompanies, and that are used in the IA.
3. Providing evidence for the **baseline scenario**. A baseline is a description of what may happen under a specific set of assumptions, which at the time of making the projections were judged plausible, and which provides a base for comparing the policy options ([28]).
4. Contributing to the **assessment of policy options**. This is the core of the IA, where different policy alternatives and their impacts are analysed.

Based on this, we retained in our analysis only models which were **actually** found to have **contributed to the evidence base** of an IA, that is model results were used in the IA ([29]) and played at least one of the roles described above (regardless of whether the model was actively run or not).

The evidence base can include models referred to:

— directly in the text;

— indirectly via another IA or via a study mentioned in the IA, if the study makes an important contribution to the IA, by providing quantitative evidence coming out of models used in the study ([30]).

Applying these criteria allowed us to compile a **final list** of **173** of the **1063** IAs to be **considered for further analysis**, using a total of **123** different models.

Amongst these 123 models, a total of **109 models** could be **explicitly identified** through an acronym which could be associated with a model documented online or in the literature. The remaining **14 models were assigned generic names** as reported in the IA. All models are reported in Annex 7.1.

3.3 Model support to Impact Assessments

The analysis of model use in IAs has been performed based on the body of knowledge generated as described in section 3.2. Before moving to the presentation of the results (chapter 4), in the remainder of this section we describe more in detail concepts and methods that were used for the analysis.

3.3.1 Definition of supported policy areas per IA

Each IA report published in EUR-Lex can be tagged with more than one **subject matter** (see Box 3 and section 2.2.2).

More specifically, the IA reports used for this analysis are associated in EUR-Lex with **56** different subject matters.

For our analysis we have grouped these subject matters into **19** broader **policy areas**, and used them to tag the respective IAs (see Annex 7.3).

However, not all IA reports are represented in EUR-Lex; in these cases, we needed to manually assign policy areas. In the majority of cases, we could make use of the subject matters that were assigned to the document(s) that accompanied the IA report (i.e. the proposal, Recommendation, Commission decision etc.). Only in three cases, no accompanied document could be found. For these, we investigated the IA report itself, and assigned the subject matter using similar IAs as a reference ([31]).

([28]) For alternative definitions of baselines see for example European Commission (2017), Marques *et al.* (2017).

([29]) For example, if the reference to the model is made when describing the methodology, terminology, definitions and concepts used, but without any numerical value that represents an output of the model.

([30]) Important contribution means that the study is either the baseline or reference scenario used, or considerable parts of the calculated impacts of policy options are taken from it. This excludes problem definition, and was decided on a case by case basis for ex post evaluation of existing policies. Note that, if a study or another IA had made an important contribution, then all models contributing to the study or IA have been considered as part of the evidence base.

([31]) SWD(2013) 184, SWD(2013) 68, SWD(2015) 289 by DG TRADE were assigned the subject matter *External relations*, following the example of SWD(2012) 209.

3.3.2 Lead DG per IA

In our analysis we also refer to the **lead DG** of the IA.

For IA reports published in EUR-Lex, with a completion date of **2015 or later**, this information can be automatically retrieved as one of the EUR-Lex metadata items (see section 2.2.2).

For IA reports published in EUR-Lex with a completion date **before 2015**, and for those that were **not available in EUR-Lex**, this information was added manually by using the sources that also served the post-processing step, such as the Impact Assessment website of the Secretariat General, and the websites of the various DGs.

The DG assigned to an IA is the one that was responsible for it when the IA was carried out. We did not consider changes in DG names and activities (e.g. before the existence of DG CLIMA the climate related activities were taken care of by DG ENV). It should also be noted that one IA can actually be led by more than one DG.

3.3.3 Role of models in IAs

Based on section 3.2, we distinguished **four different roles** that a model can play to support IAs:

- supporting the **problem definition** part of the IA;
- providing evidence in **ex post evaluation** of relevant existing policies;
- providing evidence for the **baseline**;
- contributing to the **assessment of the policy options**.

Please note that we assigned the role of 'baseline' only if the model was explicitly and exclusively used for the baseline. This happens regularly e.g. in case of the *EU reference scenario 2016 Energy, transport and GHG emissions: trends to 2050* ([32]), one of the European Commission's key analysis tools in the areas of energy, transport and climate action. At the same time, for all models described as 'contributing to the assessment of policy options', this of course also usually includes the generation of the baseline as a reference scenario to compare the policy options to.

3.3.4 Model contribution to IAs

Models can contribute to the same IA in various ways, e.g. if they play more than one role (see section 3.3.3).

This usually means that the model itself has been applied independently for each of these roles, and that these applications do form independent contributions.

Keeping this in mind, for our analysis we make use of the notion of **model contribution** to represent the different uses described above. In this way, the same model can contribute multiple times to the same IA. This concept was used notably for the analysis of the role of models in IAs (section 4.3).

([32]) https://ec.europa.eu/energy/en/data-analysis/energy-modelling/eu-reference-scenario-2016

4 Results

In this section, we present the main results of the analysis of model use in IAs over the period 2003-2018. The analysis is based on the set of IA reports and models identified in the previous chapter, as summarised in Table 1 below.

In total, we considered 1063 IAs, of which **173** turned out to have been supported by models. Of these, **161** are **inserted in EUR-Lex**.

Overall, **123 models** have been used in support of these IAs. Of these, only **109 could be explicitly identified**, while the remaining **14 were given generic names**.

In total, there were **557 model contributions** to IAs.

Table 1. Number of IAs and models considered in the analysis (years 2003-2018)

Total number of IAs considered	**1063**
- which use models	173
- which are included in EUR-Lex	161
Total number of models supporting IAs	**123**
- which could be explicitly identified	109
- which were given generic names	14
Total number of model contributions to IAs	**557**

4.1 Frequency of model use to support IAs

Our analysis reveals that a total of **173 (16%)** of the **1063** EC Impact Assessments carried out over the years 2003-2018 **were actually supported by models**, whose results contributed to the evidence base.

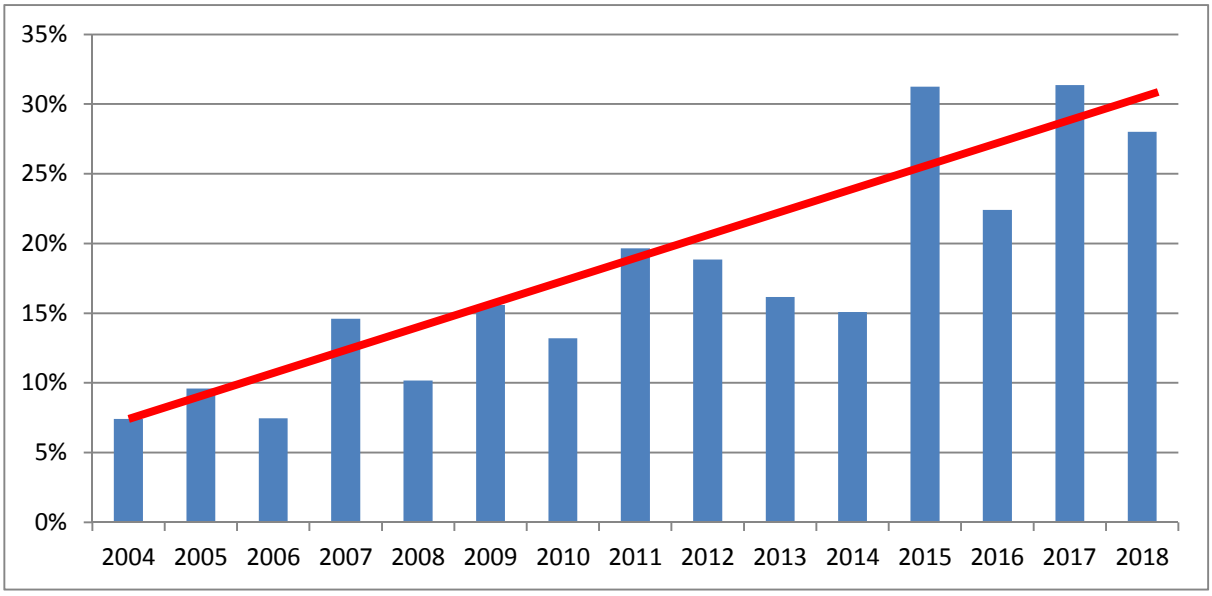
The use of models in support to IAs shows a positive trend over time.

In 2003, the first year analysed, no evidence of model use was found and only two cases emerged in 2004. Since 2005, model use to support IAs was more and more frequent starting from around 10% of the total IAs, up to around 25-30% from 2015 onwards (Table 2 and Figure 3).

Table 2. Number of IAs using models, per year

Year	Total number of IAs	Number of IAs with models	% of IAs with models
2003	21	0	0%
2004	27	2	7%
2005	73	7	10%
2006	67	5	7%
2007	89	13	15%
2008	118	12	10%
2009	77	12	16%
2010	53	7	13%
2011	117	23	20%
2012	69	13	19%
2013	99	16	16%
2014	53	8	15%
2015	16	5	31%
2016	58	13	22%
2017	51	16	31%
2018	75	21	28%
SUM	**1063**	**173**	**16%**

Figure 3. Share of EC Impact Assessments supported by models, per year (2004-2018)



In total, we found that **123 models** were used to support 173 IAs.

To identify which models are frequently used to support IAs, we looked at the number of IAs each model contributed to ([33]).

We found that, from the 123 different models that have been used in IAs, **65, more than a half** (or **53%**) were **used only once** (i.e. in a single IA, see Table 3)**.**The remaining models contributed to more than one IA, with 10 models (8% of the total) contributing to 10 or more IAs.

Table 3. In how many IAs the same model is used

Models used in	**Number of models**	**% of the total number of models**
- only one IA	65	53%
- 2 to 3 IAs	29	24%
- 4 to 10 IAs	19	15%
- 11 or more IAs	10	8%
Total	123	100%

The full list of models and the total number of IAs they support are listed in Annex 7.1.

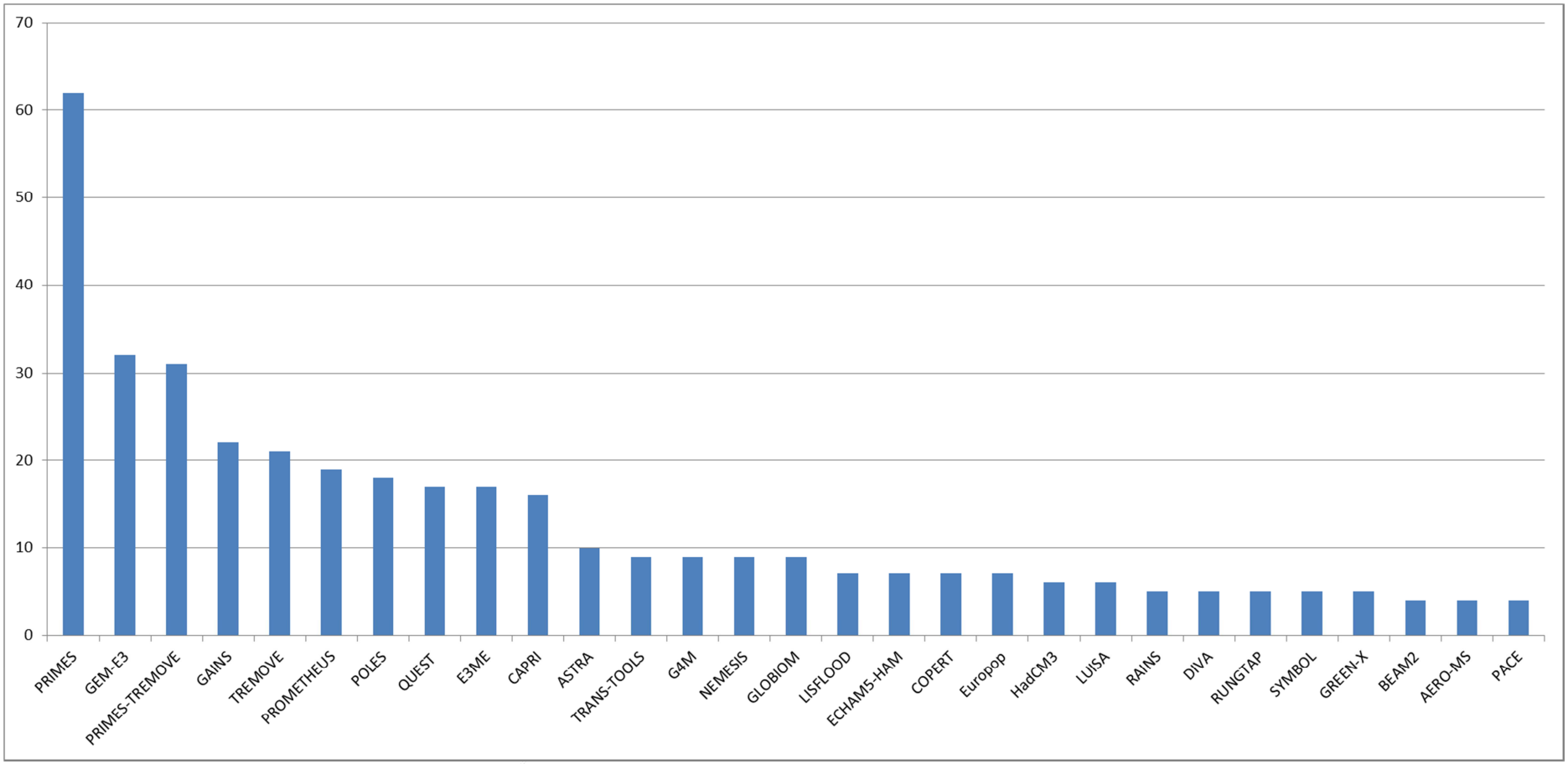
Figure 4 shows the frequency of model use in IAs for models that were used in at least four different IAs. We can see that **some models do dominate** (e.g. PRIMES, GEM-E3 and GAINS). Most of them have been used in energy-climate analyses.

The most frequently used model, **PRIMES**, was used in **62 individual IAs**, more than one third (36%) of the total number of IAs using models.

The **top ten models** (PRIMES, GEM-E3, PIRMES-TREMOVE, GAINS, TREMOVE, PROMETHEUS, POLES, QUEST, E3ME and CAPRI, see Figure 4) were used in **two thirds** (114, or 66%) of the **total number of IAs** supported by models (see also Annex 7.1)

([33]) Note that we counted each IA only once per model, irrespectively of the number of contributions (i.e. roles) the model made.

Figure 4. Number of IAs supported by model, for models used in at least four IAs (2004-2018)



Note: The full list is reported in Annex 7.1.

4.2 Policy areas of the IAs supported by models

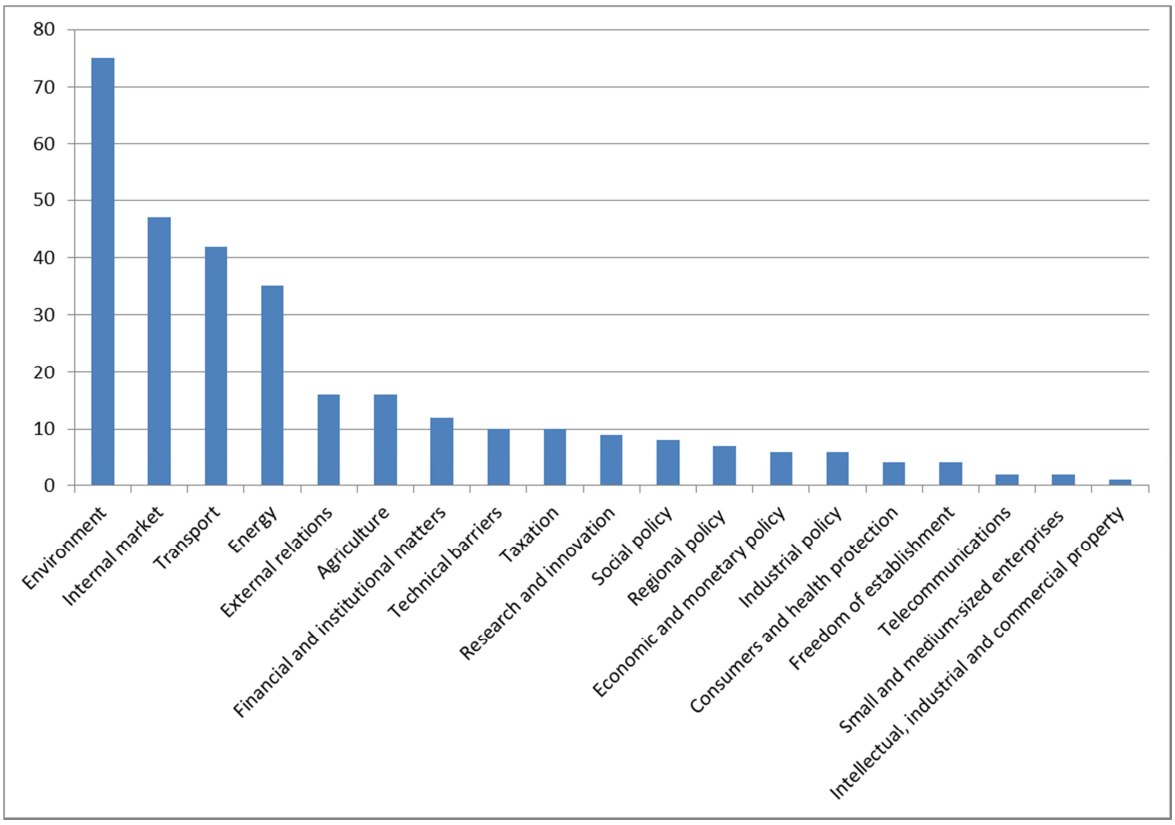
In this section we look at which **policy areas were** covered by the **173 IAs** that use modelling.

The IA reports are associated in EUR-Lex with **56** different subject matters, which have been grouped into **19** broader **policy areas** (section 3.3.1 and Annex 7.3).

Figure 5 shows the results of this analysis (the full set of the underlying data is reported in Annex 7.3). Please remember that a single IA can be assigned to more than one subject matter. Consequently, this means that IAs are considered more than once if they refer to more than one subject matter assigned to different policy areas (therefore the total of the IAs shown in Annex 7.3 is larger than 173).

The most frequent areas assigned to IAs which are supported by models are **environment, internal market, transport, energy**. It should be noted that climate action, where extensive modelling has been carried out, belongs to the environment subject area. Thorough analysis reveals that, in fact, many cases of model use related to environment actually refer to assessing impacts related to climate.

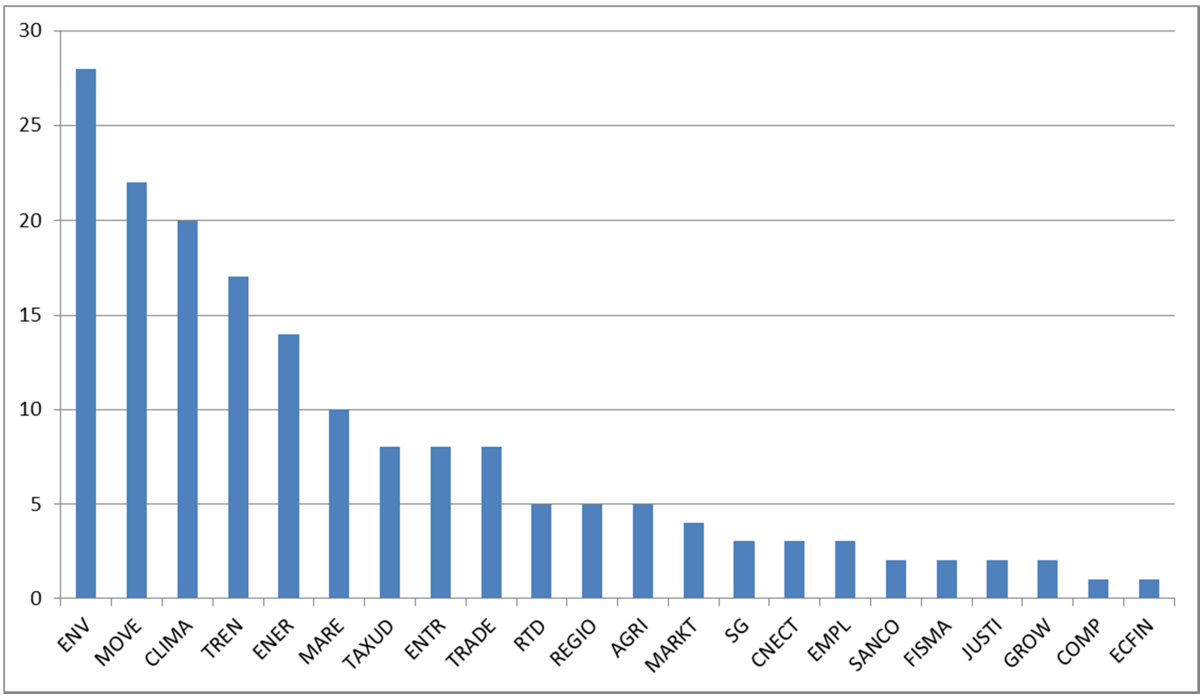
Figure 5. Main policy areas of IAs that use models (2004-2018)



A somewhat different perspective emerges when looking at **which DGs** were responsible for the 173 IAs (Figure 6). As explained in section 3.3.2, please note that we do not consider changes in DG names and activities (e.g. before the existence of DG CLIMA the climate related activities were taken care of by DG ENV). All IAs were led by one DG only.

Again, the policy areas of **Environment**, represented by DG ENV and DG CLIMA, **Energy**, represented by ENER, and **Transport**, represented by DG MOVE and DG TREN, provide for the majority of entries. However, using this perspective the extensive model use by DG **CLIMA** becomes apparent. Given the fact that DG CLIMA only started to exist in 2010, the dominance of DG CLIMA in extensive model use becomes even more evident.

Figure 6. DGs responsible for IAs that use models (2004-2018)



It should be noted that these results also depend on the number of proposals and IAs carried out in the period under analysis in a certain policy area or led by a certain policy DG. This, in turn, is also likely to be related to the duration of the policy cycle for a specific policy area.

Therefore, results do not necessarily reflect the importance of model use by policy DGs or in a certain policy area, but rather the frequency of the IAs using models carried out by that DG or in that policy area.

At the same time, it should also be remembered that, although not reflected in IA reports, models could also be used on a continuous basis for policy formulation. For example, in agriculture, models are regularly used for the yearly EU agricultural outlook and for the assessment of impacts of trade policy scenarios and of other specific policy measures. However, they are used in IAs especially during the assessment of regular reforms of the Common Agricultural Policy, which occurs every seven years and which is supported by a range of different models.

4.3 Role of models in IAs

We finally investigate how models are used to support IAs, by looking at the frequency of the various roles they perform.

For this, we use the notion of **model contribution** (see section 3.3.4), precisely because a model might contribute to the same IAs more than once, if used for different roles. The full detailed list of model contributions can be found in Annex 7.2.

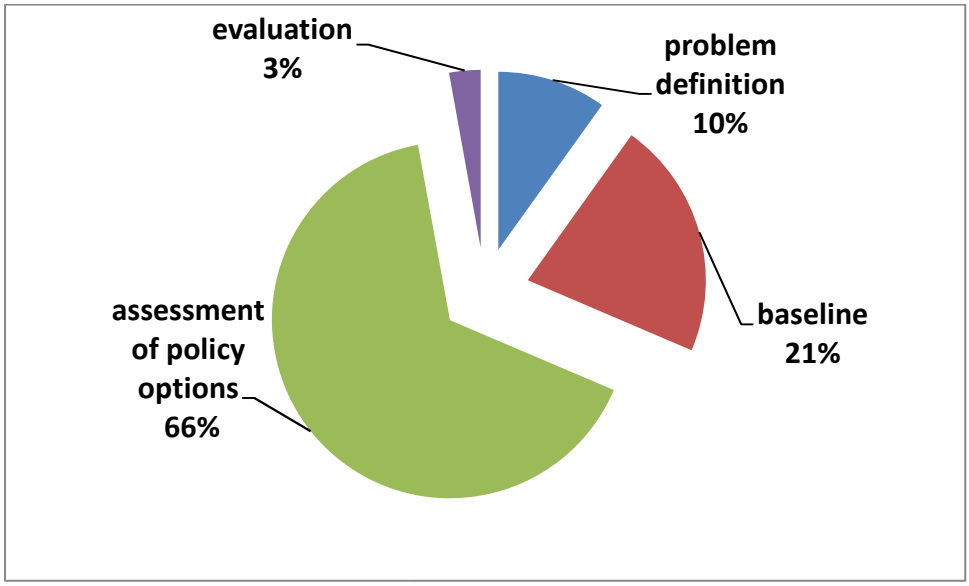
Over the **173** IAs which are supported by models, there are a total of **557 model contributions**. This implies that the **123** models found usually perform more than one role for the same IA.

The analysis reveals that, in general, there is a **predominant use of models for assessment of policy options** (**366** model contributions, or **66%**) and for the baseline (**120** model contributions, or **21%**). Model results are also used in several cases as evidence for the problem definition of the IA (**55** model contributions, **10%**), while so far there were only a few cases (**16** model contributions, **3%**) where the model contributed to IA providing quantitative evidence in relation to the ex post evaluation of policies related to the IA (see Table 4 and Figure 7).

Table 4. Model contributions to IAs (totals and % of IAs with models, 2004-2018)

Role	Total number of model contributions per role	Percentage of total
Problem definition	55	10%
Ex post evaluation	16	3%
Baseline	120	21%
Assessment of policy options	366	66%
Total	557	

Figure 7. Model contributions to IAs (% of IAs with models, 2004-2018)



These findings are even clearer if **we look at which roles were performed by individual models** (Table 5). It can be observed that basically all models, namely **116 out of 123, or 94%**, are used for the **assessment of policy options**. This is to be expected, since indeed the assessment of policy options is an extremely relevant task for quantitative analysis in IAs.

Table 5. Number of different models used for each role in Impact Assessment (2004-2018)

Row Labels	Models performing this role in any of the IAs	% of the total number of models used in IAs
Problem definition	31	25%
Ex post evaluation	13	11%
Baseline	25	20%
Assessment of policy options	116	94%
Total	**185***	**123**

**The Total of 185 is considerably larger than the list of 123 individual models, as some models performed several roles. E.g. the model PRIMES was used in the problem definition phase, the baseline, and the assessment of policy option, so it is counted 3 times.*

In fact, also the majority of the **models that are used only in one IA** (65 or 53%, see Table 3) almost exclusively contribute to the assessment of policy options.

The **top 10 models** which contributed to IAs (see Figure 4) were predominantly used both for the **assessment of policy options** and for the **calculations of baselines**.

This can be explained by looking more in detail at the role of model use in baselines in IAs. In fact, we assigned the role of 'baseline' only if the model was explicitly and exclusively used for the baseline (section 3.3.3). The "baseline only" model use occurs regularly, e.g. in the case of the *EU reference scenario 2016 Energy, transport and GHG emissions: trends to 2050* ([34]) of DGs ENER, MOVE and CLIMA. The report is the latest in a series of modelling exercises aiming to generate a harmonised and updated reference scenario, and it forms one of the European Commission's key analysis tools in the areas of energy, transport and climate action. This series is among the most quoted in IAs and it involves the majority of the frequently used models: PRIMES, PRIMES-TREMOVE, PROMETHEUS, GAINS, GLOBIOM, G4M, GEM-E3, CAPRI, and Europop. In fact these models represent the majority of the model contributions found in the category "baseline" see (Annex 7.2).

([34]) https://ec.europa.eu/energy/en/data-analysis/energy-modelling/eu-reference-scenario-2016

5 Concluding remarks

This report presents the results of a first comprehensive analysis of model use in support to the policy formulation phase of the EU policy cycle, using highly advanced tools. In particular, we focus on IA reports that were published as Staff Working Documents of the EC.

Our results are relevant in two main respects. Firstly, since they provide an overview of how models are used in IA reports published by the EC. To our knowledge, our study represents the first comprehensive attempt in this direction. Our main findings are presented in section 5.1. Secondly, our research is facilitated by and feeds back into MIDAS, the modelling inventory of the European Commission. MIDAS makes it possible to retrieve information on model use in support to IAs in a comprehensive and consistent way. The contribution of MIDAS to the BR Agenda is outlined in section 2.3.

5.1 Main results of the analysis

We analysed how models are used in the **1063 EC Impact Assessments** published **between 2003 and 2018**, by using text mining techniques (the Semantic Text Analysis Tool, SeTA) complemented by manual post processing.

Overall, we found that **models were used in 173 IAs (or 16% of the total)**. **Model use in IAs** started in 2004 with two IAs. In 2005, around 10% of IAs were using models, and this share increased to around 25-30% from 2015 onwards. This clearly indicates an increase in model use over time to contribute to quantification in IAs.

These **IAs were found to be supported by 123 different models**. However, **more than half of these models (65 or 53%) were used only once**. The remaining models contributed to more than one IA, with **10 models contributing to 10 or more IAs**. In fact, **the top ten models were used in two thirds (114, or 66%)** of the total number of **IAs** using models. It is, therefore, essential that these models undergo careful quality scrutiny and that maximum transparency and traceability of results is ensured.

Policy areas where models have been used more frequently are **environment (including climate), internal market, transport and energy**. However, it should be noted that this also reflects the frequency at which IAs are carried out in the various policy areas.

Results show that **94% (116) of the total number of models** used to support IAs were used for **ex ante assessment of policy options**. This is to be expected, since indeed the assessment of policy options is an extremely relevant task for quantitative analysis in IAs. However, the top 10 models were predominantly used also for **the calculations of baselines**. In fact, there is a consistent use of the same series of baselines in the areas of energy, transport and climate, which goes in the direction of increased consistency across policy areas and fields of analysis. However, according to Marques *et al.* (2017) there is still some room for strengthening the coherence in baselines across the EC. Model results are also used in several cases as evidence for the **problem definition** of the IA, while so far there were only a few cases in which models contributed to the **ex post evaluation of policy options**.

5.2 Transparency of model use in support to policies

The results of our analysis **contribute to the BR Agenda** by highlighting aspects of transparency, coherence, traceability and accountability in the use of evidence for EU policy making.

Transparency is crucial to understand how models work and to validate their behaviour. This in turn encourages trust as well as their sound and widespread use in support to policy making. In addition, by providing an overview of models used in support to policies, transparency contributes to a more effective and efficient use of resources for model development and use within the Commission. Enhanced transparency in models and model use can also increase trust and contribute to the responsible and sound use of modelling, improving quantification in IAs (RSB SG and JRC Working Group 2018).

Impact Assessments have a crucial role in the policy formulation phase of the policy cycle. Over time, the quality of EC IAs is found to have improved (EP cited in Golberg, 2018). However, it is argued that there is still a lack of synopsis of data, findings and methodologies such as modelling (Impact Assessment Institute 2017). **More transparency has been recommended** in these respects (RegWatchEurope cited in Golberg 2018;

Impact Assessment Institute 2017). Otherwise, the complexity of IAs could also represent an obstacle to their use (Nowag and Groussot 2018).

In this respect, the **requirement** introduced **by the BR Agenda in 2017** for an **annex** on 'Analytical models used in the preparation of the impact assessment' contributed to better model descriptions and an increased transparency of the methodology used. This effectively contributes to keep IAs understandable and useful for decision-makers and scrutinising bodies like the Regulatory Scrutiny Board of the EC and the European Parliament.

There is, at the same time, **still room for improvement** in better documenting models as the individual model descriptions in IAs are still quite **heterogeneous**. In addition, we also identified some major challenges related to **referencing in IAs**, such as lack of harmonised and adequate references or outdated hyperlinks, which made the quantitative evidence untraceable in several cases (see Box 5). More than once, it was difficult to trace whether a model was used to provide the quantitative information present in the text, which model was it exactly, and what its role had been. The quantitative evidence base was therefore untraceable in several IAs. Next to text mining many person-hours also had to be used to identify the models, and even in this case it was not always possible to identify the source. This means that, in fact, some models may have been missed.

MIDAS, the Commission-wide knowledge management tool for modelling, directly contributes to enhanced **transparency** and **traceability** of models used to support policies. Starting in 2017, the BR Toolbox foresees that, when IA analysis relies on modelling or the use of analytical methods, the model should be documented in the corporate modelling inventory MIDAS (European Commission 2017). This represents a major step forward in respect to transparency. In fact, MIDAS contributes to the Better Regulation policy in various respects:

- the majority of the analysed model-based **IAs have now been stored** and organised in MIDAS. Before 2015, there was no central register on model use in IAs. To collect data for our analysis, we had to rely on text mining techniques; future analyses of model use in IAs can instead be more easily and efficiently based on information stored in MIDAS.
- MIDAS can contribute to **better referencing in IAs**. From now on, information related to model use in an IA will be stored in MIDAS in a consistent manner, and the IAs can use this information for references in case a model is used in another IA.
- the **information and reports** generated by MIDAS on models and model use can be used to generate the mandatory annex on 'Analytical models used in the preparation of the impact assessment'. This could effectively contribute to keep IA understandable and useful for decision-makers, as recommended in the academic debate (Golberg 2018).
- organised evidence gathering, like the one achieved thanks to the MIDAS modelling inventory, contributes to **learning** (Smismans 2015). This includes also facilitating the examination process by the EU internal scrutinizing bodies like the Regulatory Scrutiny Board of the EC and the European Parliament.

However, further action is needed to use and **promote best practices** to ensure transparency and accessibility over time of the evidence base in support to IAs. **CC-MOD supports the EC strategy**, promotes good practices in data management and facilitates the use of the EU Open Data Portal and the JRC Data Catalogue by modelling teams. In addition to the BR guidelines, the **JRC**, as the science and knowledge in house service of the Commission, can also provide **additional assistance and support** to the Policy DGs.

Box 5. Best practices for referencing in IAs

Based on the challenges we encountered during the analytical process, the following best practices can be recommended for IAs:

1) Hyperlinks should be used as the only source only if they are permalinks with persistent identifiers; in all other cases, hyperlinks should be accompanied by full citation of the source.

2) For studies which are not yet published, citation standards should be followed, including number and funding scheme in case of projects, and by using permalinks wherever possible.

3) If published datasets are used, then the datasets themselves should be cited.

A final consideration concerns external studies, and the fact that many of them, also in recent years, are neither published in the EU Bookshop nor equipped with persistent unique identifiers like DOIs. This is hindering their traceability and should be addressed.

6 Outlook

As explained above, the present study only focuses on the use of models in support to the policy formulation phase of the policy cycle, and more specifically on IAs reports of the EC that are published as SWD.

Other types of model used by the EC are still to be covered. Namely, this includes models used in other documents or processes to support the policy formulation phase, as well as other phases of the policy cycle.

These interesting **future directions for the analysis** could be performed by using the SeTA tool to explore existing data sources from the Commission. These include over 500,000 documents, such as legislative and preparatory documents available in EUR-Lex, and studies and other publications from the EU publications website and the Repository of JRC publications, PUBSY.

In addition to this, the analysis could explore in more depth the characteristics of model use to support policymaking; as an example, it could be studied whether the same models are used for ex post and ex ante analysis (in the terminology adopted in the Better Regulation, evaluation and impact assessment), an aspect which is extremely relevant both in terms of the application of the 'evaluate first principle' and of consistency in the methodology. Another, interesting aspect to explore could be whether existing models in the EC cover the modelling needs, and to identify gaps. This could be valuable input to contribute to planning medium and long term modelling activities within the EC.

References

European Commission (2015) 'Better Regulation for Better results - An EU agenda', *COM (2015) 215 final, Brussels, 19 May.*

European Commission (2017) 'Better Regulation Guidelines', *SWD(2017) 350 final, Brussels, 7 July*.

European Parliament Council of the European Union and European Commission (2016) 'Interinstitutional Agreement on Better Law-Making', *13 April 2016, OJ L 123/1*.

European Parliament European Council and European Commission (2003) 'Interinstitutional Agreement on Better Law Making', *31 December 2003, OJL 321/01*.

Golberg, E. (2018) '"Better Regulation": European Union Style', *Harvard Kennedy School, Mossavar-Rahmani Center for Business and Government, M-RCBG Associate Working Paper Series* (98).

Hradec, J. et al. (2019) 'Semantic Text Analysis Tool', *EUR 29708 EN, Publications Office of the European Union, Luxembourg, ISBN 978-92-76-01518-5, doi:10.2760/577814, JRC116152*.

Impact Assessment Institute (2017) 'A year and a half of the Better Regulation Agenda : what happened ?', *Final study*.

Listorti, G. et al. (2019) 'The debate on the EU Better Regulation Agenda: a literature review', *EUR, Publications Office of the European Union, Luxembourg, 2019, ISBN 978-92-76-00840-8, doi:10.2760/46617, JRC116035*.

Marques, A., Hradec, J. and Rosenbaum, E. (2017) 'Baseline Assumptions in EC Impact Assessments', *EUR 28951 EN, Luxembourg: Publications Office of the European Union, 2017, ISBN 978-92-79-77153-8, doi: 10.2760/982695, PUBSY No. JRC109839*.

Naundorf, S. and Radaelli, C. M. (2017) 'Regulatory Evaluation Ex Ante and Ex Post: Best Practice, Guidance and Methods', in U. Karpen and H. Xanthaki (eds). *Legislation in Europe. A Comprehensive Guide for Scholars and Practitioners*. Hart Publishing, pp. 187–214.

Nowag, J. and Groussot, X. (2018) 'From Better Regulation to Better Adjudication? Impact Assessment and the Court of Justice's Review', in S. Garben and I. Govaere (eds). *The EU Better Regulation Agenda: A Critical Assessment*. Bloomsbury Publishing, pp. 186–202.

Ostlaender, N. et al. (2019) 'Modelling Inventory and Knowledge Management System of the European Commission (MIDAS)', *EUR 29729 EN, Publications Office of the European Union, Luxembourg, 2019, ISBN 978-92-76-02852-9 (online), doi:10.2760/900056 (online), JRC116242*.

Petrov, L. O. et al. (2017) 'Analysis of the use of models by the European Commission in its Impact Assessments for the period 2009 – 2014', *EUR 28628 EN, Publications Office of the European Union, Luxembourg, ISBN 978-92-79-69283-3, doi:10.2760/096340, JRC105329*.

RSB SG and JRC Working Group (2018) *Quantification in Commission Impact Assessments and Evaluations*, Brussels, available at https://ec.europa.eu/info/publications/report-rsb-sg-jrc-working-group-quantification-commission-impact-assessments-and-evaluations_en.

Smismans, S. (2015) 'Policy evaluation in the EU: The challenges of linking ex ante and ex post appraisal', *European Journal of Risk Regulation* 6(1): 6–26.

List of abbreviations

BR	Better Regulation
CC-MOD	Competence Centre on Modelling
EC	European Commission
EPRS	European parliamentary Research Service
IA	Impact Assessment
JRC	Joint Research Centre
MIDAS	Modelling Inventory and Knowledge Management System
SeTA	Semantic Text Analysis Tool
SWD	Staff Working Document

List of figures

List of tables

List of boxes

7 Annexes

7.1 Models used in IAs

Note: the list below contains models developed or run by the EC or by third parties, such as national or international research institutes, companies or consortia.

Model Acronym	Full title	number of IAs supported
a4a	Assessment for All	3
AERO-MS	Aviation Emissions and evaluation of Reduction Options Modelling System	4
AGLINK-COSIMO	AGLINK-COSIMO partial equilibrium model for global agriculture	3
AIDS	Almost Ideal Demand System	1
AIDSK	Model based on EU-FADN used to assess distribution of direct payments	2
AnaFgas	Analysis of Fluorinated greenhouse gass in EU-27	1
ArcGIS Network	Geospatial model for biomass transport chains	1
ASTRA	ASsessment of TRAnsport Strategies	10
BEAM2	Built-Environment-Analysis-Model	4
BEMTOOL	Bioeconomic model	1
BioMA	Biophysical Model Applications modelling platform	1
CAPRI	Common Agricultural Policy Regional Impact Analysis	16
CARIS	Centre for Analysis of Regional Integration at Sussex model	1
CBM-CFS3	Carbon Budget Model of the Canadian Forest Sector	1
CENTURY	Agroecosystem dynamic model	1
CETM	Copenhagen Economics Trade Model	1
COPERT	Computer model to calculate emissions from road traffic	7
CORTAX	Corporate Tax Model	3
CropSyst	CROPping SYSTems simulation model	1
DIONE	Road Transport Fleet Impact Model	2
DIVA	Dynamic Interactive Vulnerability Assessment model	5
DSSAT	DSSAT crop models	2
E3ME	Energy - Environment - Economy Model for Europe	17
E3MG	Energy - Environment - Economy Model at the Global level	2
ECHAM5-HAM	European Centre Hamburg Model (Global Aerosol Climate Model)	7
Economics of Tobacco Toolkit	Economics of Tobacco Toolkit	1
EFFIS	European Forest Fire Information System	2
EFISCEN	European Forest Information SCENario	1
EmIO-F Europe	Employment Input-Output Model for Analysis of Policies and Measures for the European Union	1
EPIC	Environmental impact calculator	3
ESIM	European Simulation Model	3
EU waste model	European Reference Model on Municipal Waste Management	1
EUCS100	EUClueScanner	2

EU-FASOM	European Forest and Agricultural Sector Optimization Model	2
EUGas	EUropean Gas assessment	1
EUROMOD	Tax-benefit microsimulation model	1
Europop	EUROSTAT Population Projections	7
EXIOMOD	EXtended Input-Output MODel	1
FAIR	FAIR climate policy model	1
FLBEIA	Bio-economic Impact Assessment using Fisheries Library in R	1
FLR	Fisheries Library in R	2
Francois GTAP model	GTAP Model modified by Prof. Francois	3
G4M	Global Forest Model	9
GAINS	Greenhouse Gas and Air Pollution Interactions and Synergies Model	22
GEM-E3	General Equilibrium Model - Economy, Energy, Environment	32
GLOBE	The GLOBE model	2
GLOBIO3	A global biodiversity model	2
GLOBIOM	Global Biosphere Management Model	9
GMR	Geographic macro and regional modeling	1
GREEN-X	EU renewable energy model	5
GTAP	Global Trade Analysis Project model	1
HadCM3	Hadley Centre Coupled Model, version 3	6
HERMIN	Macro Modeling Framework to Investigate Structural Policies	2
IAM	Impact Assessment Model for Fisheries Management	1
IFM-CAP	Individual Farm Model for Common Agricultural Policy Analysis	1
IMAGE	Integrated Model to Assess the Global Environment	2
InfraCCS	InfraCCS - Infrastructure for Carbon Capture and Storage model	1
INIA	pan-European eutrophication model	1
LEITAP	Landbouw Economisch Institute Trade analysis Project model	1
LISFLOOD	LISFLOOD hydrological model	7
LISQUAL	The LISQUAL model	2
LUISA	LUISA Territorial Modelling Platform	6
LULUCF tool	Land Use, Land Use Change and Forestry tool	1
MAGNET	Modular Applied GeNeral Equilibrium Tool	1
MARE Nest model	Swedish Baltic Nest Institute MARE model	1
MARWAS	Maritine Waste gap model	1
MC-GENERCIS	"model of the European gas network"	1
MERGE	Integrated Assessment Model for Global Climate Change	1
METIS	Markets and Energy Technologies Integrated Software	2
MICUS	MICUS Management Consulting GmbH model	1
MIRAGE	Modeling International Relationships in Applied General Equilibrium model	2
MITERRA-EUROPE	Integrated Assessment of nitrogen flows in agriculture of EU-27	1
MULTIREG	Multiregional Input-Output model	3

NEAC	European Transport Forecast Model	2
NEMESIS	New Econometric Model of Evaluation by Sectoral Interdependency and Supply	9
PACE	Policy Analysis based on Computable Equilibrium	4
PAGE	PAGE Integrated Assessment Model	1
PESERA	Pan-European Soil Erosion Risk Assessment	1
Phoenix	population user support system	1
POLES	Prospective Outlook for the Long term Energy System	18
PRIMES	PRIMES Energy System Model	62
PRIMES-TREMOVE	PRIMES-TREMOVE Transport Model	31
PROMETHEUS	PROMETHEUS model	19
PyCSIS	Passenger Car fleet emissions SImulator	1
QUEST	Macroeconomic model QUEST III	17
RAINS	Regional Air Pollution Information and Simulation model	5
REMIND	Regional Model of Investments and Development	1
RHOMOLO	Regional Holistic Model	3
RunDynam	RunDynam - Recursive Dynamic Model	3
RUNGTAP	RUNGTAP model - Static CGE	5
RURAL EC MOD	Ex ante Spatial Policy Impact Analysis of the Rural Development Policy	1
RUSLE2015	Revised Universal Soil Loss Equation	1
SASI	Spatial and Socio-economic Impacts of Transport Investments and Transport System Improvements	2
SCENES	Scenarios for European transport	3
SLAM	logistics noise model	1
SMART	Single Market partial equilibrium simulation model	1
SWOV	road safety model	1
SYMBOL	SYstemic Model of Banking Originated Losses	5
TIMES	The Integrated Markal Efom System	1
TM5-FASST	Fast Scenario Screening Tool for impact analysis of air pollutant emissions on air quality and short-lived climate pollutants	1
TRANS-TOOLS	TOOLS for TRansport forecasting ANd Scenario testing	9
TREMOVE	TREMOVE economic transport and emissions model	21
TRUST	TRansport eUropean Simulation Tool	3
VECTO	Vehicle Energy Consumption calculation TOol	1
WaterGAP	Water - Global Analysis and Prognosis	3
WESIM	Whole-electricity System Investment Model	1
WOFOST	WOrld FOod Studies generic crop growth model	1
WORLD	World Linear Programming model	1
World Scan	World Scan computable general equilibrium model	1
aquaculture growth model	ad hoc name as given in the IA	1
bio-economic model	ad hoc name as given in the IA	1
biomass model	ad hoc name as given in the IA	1

CGE model 1	ad hoc name as given in the IA	1
CGE model 2	ad hoc name as given in the IA	1
combined transport model	ad hoc name as given in the IA	1
crop simulation models	ad hoc name as given in the IA	1
fishery model	ad hoc name as given in the IA	1
fishing gear model	ad hoc name as given in the IA	1
modelling exercise	ad hoc name as given in the IA	1
population model	ad hoc name as given in the IA	1
property pricing model	ad hoc name as given in the IA	1
single use of plastic model	ad hoc name as given in the IA	1
water balance model	ad hoc name as given in the IA	1
TOTAL		**513**

7.2 Model roles in IAs

	Model role				
Model Acronym	**problem definition**	**ex post evaluation**	**baseline**	**assessment of policy options**	**Total**
a4a				3	3
AERO-MS	1			3	4
AGLINK-COSIMO			1	2	3
AIDS				1	1
AIDSK				2	2
AnaFgas				1	1
aquaculture growth model				1	1
ArcGIS Network				1	1
ASTRA	1		1	8	10
BEAM2				4	4
BEMTOOL				1	1
bio-economic model				1	1
BioMA	1			1	2
biomass model				1	1
CAPRI	1	1	5	13	20
CARIS				1	1
CBM-CFS3				1	1
CENTURY				1	1
CETM				1	1
CGE model 1				1	1
CGE model 2				1	1
combined transport model				1	1
COPERT	2		1	4	7
CORTAX				3	3
crop simulation models				1	1
CropSyst	1			1	2
DIONE				2	2
DIVA	2	2		2	6
DSSAT	1			1	2
E3ME	2			15	17
E3MG				2	2
ECHAM5-HAM	4	1		3	8
Economics of Tobacco Toolkit				1	1
EFFIS	2				2
EFISCEN				1	1
EmIO-F Europe				1	1

EPIC			1	2	3
ESIM				3	3
EU waste model				1	1
EUCS100		1	2		3
EU-FASOM				2	2
EUGas				1	1
EUROMOD				1	1
Europop			5	2	7
EXIOMOD				1	1
FAIR				1	1
fishery model				1	1
fishing gear model				1	1
FLBEIA				1	1
FLR				2	2
Francois GTAP model				3	3
G4M			3	8	11
GAINS	2		6	15	23
GEM-E3	3	2	9	23	37
GLOBE				2	2
GLOBIO3				2	2
GLOBIOM			3	8	11
GMR				1	1
GREEN-X			1	5	6
GTAP				1	1
HadCM3	3	1		3	7
HERMIN				2	2
IAM				1	1
IFM-CAP				2	2
IMAGE		1		1	2
InfraCCS				1	1
INIA				1	1
LEITAP				1	1
LISFLOOD	4	2	2	2	10
LISQUAL			1	1	2
LUISA	1		3	3	7
LULUCF tool				1	1
MAGNET				1	1
MARE Nest model	1				1
MARWAS				1	1
MC-GENERCIS				1	1
MERGE				1	1
METIS				2	2
MICUS				1	1
MIRAGE	1			1	2
MITERRA-EUROPE				1	1

modelling exercise				1	1
MULTIREG			1	2	3
NEAC				2	2
NEMESIS	1		1	7	9
PACE				4	4
PAGE	1				1
PESERA				1	1
Phoenix				1	1
POLES	1		1	18	20
population model				1	1
PRIMES	6	1	32	29	68
PRIMES-TREMOVE	1		24	16	41
PROMETHEUS	1		9	9	19
property pricing model				1	1
PyCSIS			1		1
QUEST	2	1		14	17
RAINS	2			3	5
REMIND				1	1
RHOMOLO				3	3
RunDynam				3	3
RUNGTAP		1		4	5
RURAL EC MOD				1	1
RUSLE2015				1	1
SASI	1			1	2
SCENES	1		1	1	3
single use of plastic model				1	1
SLAM				1	1
SMART				1	1
SWOV				1	1
SYMBOL		1		4	5
TIMES				1	1
TM5-FASST				1	1
TRANS-TOOLS			1	8	9
TREMOVE	2		5	14	21
TRUST				3	3
VECTO				1	1
water balance model				1	1
WaterGAP	2	1			3
WESIM				1	1
WOFOST	1				1
WORLD				1	1
World Scan				1	1
Total of model contributions	**55**	**16**	**120**	**366**	**557**

7.3 Policy areas and corresponding subject matters of IAs using models

	Number of IAs
Agriculture	**16**
Agricultural structural funds	1
Agricultural structures	1
Agriculture and Fisheries	2
Common organisation of agricultural markets	1
European Agricultural Guidance and Guarantee Fund (EAGGF)	1
Fisheries policy	9
Veterinary legislation	1
Consumers and health protection	**4**
Consumer protection	1
public health	3
Economic and monetary policy	**6**
Conjunctural policy	2
Economic and Monetary Union	1
Economic policy	3
Energy	**35**
Energy	32
Investments	1
Nuclear common market	2
Environment	**75**
Environment	71
Pollution	2
Waste	2
External relations	**16**
Association Agreement	2
Commercial policy	3
Development cooperation	1
External relations	9
Preferential systems	1
Financial and institutional matters	**12**
Budget	1
Financial provisions	6
General provisions	1
Provisions governing the Institutions	4
Freedom of establishment	**4**
Freedom of establishment	4
Industrial policy	**6**
Industrial policy	1
Industry	2
Technology	3
Intellectual, industrial and commercial property	**1**

Intellectual, industrial and commercial property	1
Internal market	**47**
Approximation of laws	23
Internal market - Principles	18
Joint undertakings	1
Trans-European networks	5
Regional policy	**7**
Cohesion Fund	2
economic, social and territorial cohesion	2
European Regional Development Fund (ERDF)	1
Regional policy	2
Research and innovation	**9**
Information and verification	2
Research and technological development	5
Research and training	1
Scientific and technical information and documentation	1
Small and medium-sized enterprises	**2**
Small and medium-sized enterprises	2
Social policy	**8**
Employment	2
European Social Fund (ESF)	1
Free movement of capital	1
Safety at work and elsewhere	1
Social policy	1
Social provisions	2
Taxation	**10**
Taxation	9
Value added tax	1
Technical barriers	**10**
Technical barriers	10
Telecommunications	**2**
Telecommunications	2
Transport	**42**
Transport	42

GETTING IN TOUCH WITH THE EU

In person

All over the European Union there are hundreds of Europe Direct information centres. You can find the address of the centre nearest you at: https://europa.eu/european-union/contact_en

On the phone or by email

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

- by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),

- at the following standard number: +32 22999696, or

- by electronic mail via: https://europa.eu/european-union/contact_en

FINDING INFORMATION ABOUT THE EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website at: https://europa.eu/european-union/index_en

EU publications

You can download or order free and priced EU publications from EU Bookshop at: https://publications.europa.eu/en/publications. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see https://europa.eu/european-union/contact_en).

KJ-NA-29832-EN-N

The European Commission's science and knowledge service

Joint Research Centre

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub

ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub - Joint Research Centre

EU Science, Research and Innovation

EU Science Hub

Publications Office of the European Union

doi:10.2760/748720

ISBN 978-92-76-09671-9