COMPARE

CO2 Model for Operational Programme Assessment in EU Regions

Technical background and guidance for deployment in EU regions

Acknowledgement

Abstract

The CO$_2$MPARE model enables national and regional authorities to assess the carbon impacts of Operational Programmes co-financed through the European Regional Development Fund (ERDF). This document provides technical background information and guidance for deploying the model in additional EU regions.

Authors:
M. Hekkenberg **(ECN)**
S. Le Pierrès **(Énergies Demain)**
R. Del Ciello **(ENEA)**
I. Keppo **(UCL)**
J. Harnych **(Enviros)**
S. Papagianni **(CRES)**

Contact person:
M. Hekkenberg (ECN)
P.O. Box 1
1755 ZG Petten
hekkenberg@ecn.nl

ECN-O--12-032
March 2013

Contents

1

Introduction

1.1 The CO$_2$MPARE project

The CO$_2$MPARE model has been created to assess the carbon impacts of Operational Programmes co-financed through the European Regional Development Fund (ERDF). Through its regional policy, the European Union co-invests in regional structural development programmes in the EU Member States and regions. With this policy, the EU aims to support job creation, competitiveness, economic growth, improved quality of life and sustainable development. National and regional authorities will weigh their investment decisions on these criteria. CO$_2$MPARE is designed to assist the relevant authorities in these decisions by providing intelligence on their carbon impact.

The CO$_2$MPARE model is aimed at a target audience of policy advisors at national or regional authorities that are confronted with issues with regard to carbon impact of investment. As discussed in the user tutorial, user´s may operate the model in basic mode or in expert mod, depending on their needs and capacity.

The current version of the CO$_2$MPARE model provides a generic model template, that needs to be adjusted to regional needs before practical application by the target audience. The model template that has not been adjusted to a specific regional situation is referred to as the ´generic model´ in this document. The current version already contains one national and four regional applications that can be used as examples for model testing and model presentations. This technical background and guidance document refers to the model that has been adjusted to a specific national or regional situation as the ´regionalized model´.

1.2 Purpose of this document

This document is aimed at an audience that is interested in the CO_2MPARE model's working mechanisms from a model developer´s perspective. It should be regarded a technical elaboration, complementary to the user's tutorial and assumes that the reader already understands the fundamentals of the model described therein. It refers to the relevant sections of the tutorial where relevant.

The document specifically aims at providing a better understanding of the technical procedures and links between various model elements for those that want to deploy the generic model in new regions (other than the testing regions). Such better understanding can also be instrumental when to further tailor the performance of the regionalized model to regional needs beyond what is described in the user's tutorial.

Apart from discussing the functional model in more technical detail in chapter 2, the document provides some guidance in the process to deploy the model in new regions (chapter 3), based on the development of the model in the testing regions. Although this document provides more technical detail of the model, it should not be regarded a step by step guide to learn the procedures in the model code, nor a full documentation of the model's parameters. The model core is programmed in Microsoft's visual basic for excel language. Model developers able to read this programming language are best served to read the actual code. Much of the documentation of the model, including sources for parameter values and detailed calculation algorithms, are stored in the model files themselves.

This guidance document thus has two main purposes:
- Elaborate on the model architecture for model developers.
- Guidance for model developers that aim to deploy the model in new regions.

2

Technical model description

2.1 General model structure

The main principle of the model is to estimate the carbon emissions related to various investments by assessing the emission impact per euro spent for a given type of activity and multiplying this by the amount spent on the activity. The CO_2 emissions of the whole programme are calculated by summing over the emission impacts of all activities included in the model.

The model thus connects two types of data:

- Financial data that describes the amounts invested in various activities.
- Physical data that describes the emission impact of given activities.

On the most aggregate level, the model therefore consists of two parts: one that covers the financial disaggregation into activities (see section 2.2) and one that covers the calculation of physical impacts of these activities (see section 2.3). Figure 1 gives a graphical representation of this high level model structure.

The financial data in the model stems from the regional budget allocations over various investment themes pursued in a region. These data should be user defined, although the regionalized model provides default values based on historic distribution for reference. The financial data distribution is the principle model input that the user will adapt and adjust in order to evaluate their Operational Programme and compare it with possible alternative distribution options. The results and inputs of a particular financial distribution 'scenario' can be stored and compared later with alternative distribution scenarios, as described in the CO_2MPARE user tutorial.

The physical data used in the model stem from a database that is an internal part of the model. Although most of the data can be changed by the user, this functionality is available only in the expert-use mode of the model; this functionality is considered to require technical expertise on physical activities and emission calculations. As the physical data are considered to represent physical relationships that are mostly

unchanging, the database is part of the model core. Changes in the database will therefore be reflected across all available scenarios.

The activity specific algorithms that are used to calculate emission impacts per euro invested are an integral part of the model core and cannot be changed by the user.

The functional regionalized model structure effectively helps the users to disaggregate their regional financial budget over various activities and connects this to the emission impact data.
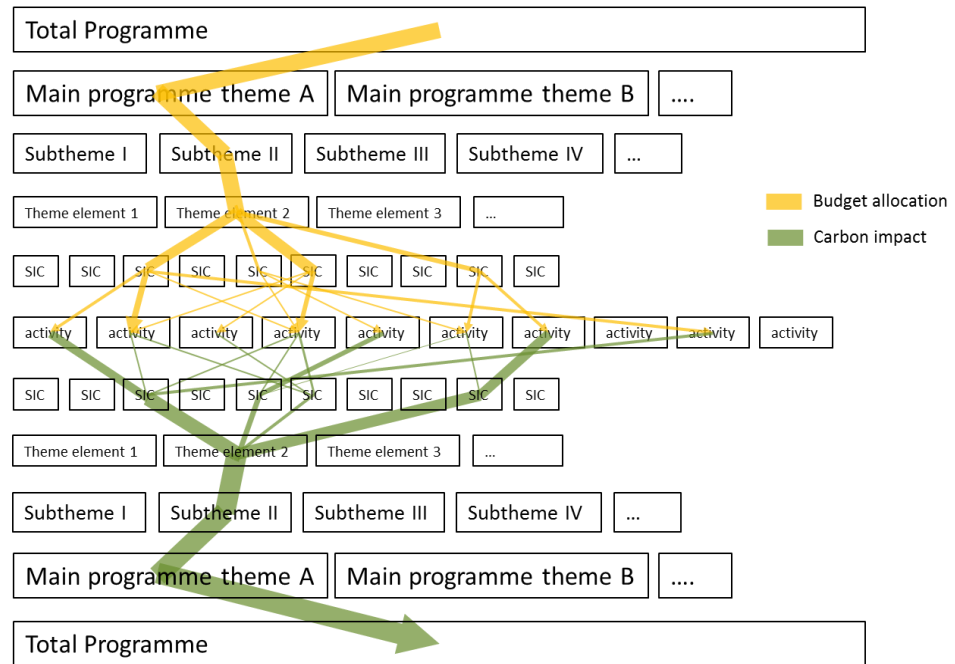
Figure 1:General model principle. [SIC = Standardized Investment Component]

To be functional for the anticipated target audience in a region, the regionalized model offers default data for both the financial and the physical inputs. This allows users to implement adjustments only for the areas of their own interest, without having to spend much time on specifying the details for areas with lower priority.

2.2 Financial disaggregation

2.2.1 Themes, subthemes, theme-elements

The structure of the model builds upon the architecture of the regional Operational Programmes. Regional programmes generally comprise a collection of activities in a large number of different themes. The activities within a programme may be grouped at different levels of aggregation. The way activities are grouped in a programme is called 'programme architecture' in this document. The highest aggregation level of a programme may consist of the main programme themes amongst which budgets are

divided. The themes in an architecture can then usually be further divided into the more descriptive, actual project-type activities that are invested in.

As programme architectures with either two or three levels are observed in practice (See Box 1), the model enables three levels of aggregation: main programme themes, subthemes and theme elements. For two-level architectures the subtheme level can simply be left out.

Box 1: Observed programme architectures

In Commission Regulation (EC) No 1828/2006, which targets the 2007-2013 programming period, the EC has defined a common European list of categories of intervention, to be used for information purposes. It categorises the types of actions financed into 86 'categories of intervention', which are also aggregated into 17 main themes.

The most recent proposal for categorization of funds assistance in the upcoming programming period (2014-2020) for the European Regional Development Fund, Cohesion Fund and European Social Fund uses a 3-level architecture, of which the most detailed level is now called 'intervention fields'.

Next to the EU architectures, regional or national authorities often use their own architecture, which may align better with internal priorities. Such regional or national architectures have been observed to consist of two or three levels.

2.2.2 Standardized Investment Components (SICs)

Generally, investment in a given theme element may fund a range of activities, with varying characteristics. The model is based on the idea that the investments on the level of theme elements can be further disaggregated into the actual physical activities that represent portfolios of projects. It works from the presumption that even though activities undertaken in different projects under the same theme element may differ in scope and specification, such activities are of essentially similar nature. Furthermore it presumes that with a sufficiently high number of individual projects, a representative activity will estimate the indicators (emissions etc.) relatively well. Thus, the model breaks down the activities within a theme element into a predefined set of standardized activities, called '*Standardized Investment Components*' (SICs).

A SIC comprises a set of physical activities that typically relate to a given investment, during its construction and operation phases; e.g. the SIC 'road building construction' contains activities that take place for building a road, such as earthworks, paving, construction, etc., and the additional transport that is generated through the improved road network. The model uses a set of 26 SICs to describe all the activities that may take place within the various theme elements. An overview of the SICs available in the generic model is given in Table **1**. As is explained in more detail in section 2.3, some SIC characteristics may be adjusted to align better with the given theme element. This

allows for instance to differentiate between the fact that roads built within one theme element would generally be highways, whereas in another theme element they would more often be local roads.

The regionalized model provides a default allocation of budget within a theme element over the various SICs, as described in section 3.2; however, this division can be adapted manually by the model user. Ultimately, through these steps, the budget of a programme is thus split into investments in various '*Standardized Investment Components*' (SICs).

Figure 2 graphically shows how the financial disaggregation leads to a given budget per SIC. This budget is used as an input for the physical impact calculation described in the following section.

Themes	Standardized Investment Components
Buildings	Building construction – extension
	Building rehabilitation
	Building demolition
Energy	Fossil energy plant
	Renewable centralised energy
	Renewable decentralised energy
	Energy efficiency
	Energy switch equipment
Transport	Airport
	Bicycle infrastructure
	Ports
	Public transportation
	Rail construction
	Rail electrification
	Rail renovation
	Road construction
	Road renovation
	Waterways
Waste & Water	Waste management infrastructure
	Wastewater treatment
	Water supply treatment
Others	Studies
	Civil engineering
	Equipment
	Reforestation
	Configurable SIC

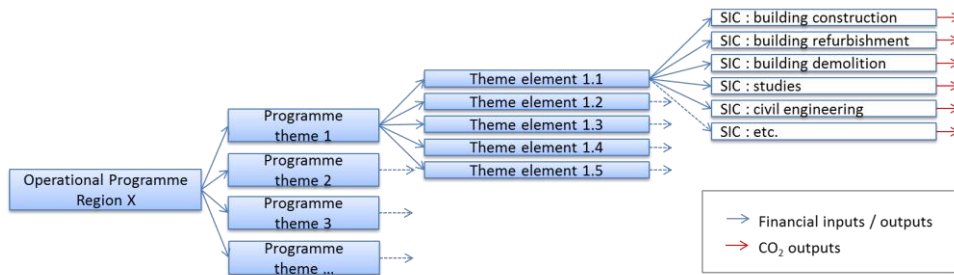Table 1: Overview of SICs included in the model

Figure 2: Breakdown of programme funds to SICs at theme element scale – case of a European architecture

2.3 Physical impact calculation

2.3.1 The SIC algorithms

The core of the model is formed by a set of algorithms that describe the carbon emission consequences of investments in each SIC. This section describes the general approach to the calculation algorithms. Section 2.4 provides a detailed example of the methodology. The full algorithms of each SIC can be retrieved directly in the model by clicking "SIC management" on the home page and then selecting the SIC.

The algorithms describe the various physical determinants that are linked to a SIC over its construction and operation phases, per euro invested. For instance, the SIC 'building construction' involves physical determinants such as material use, energy consumption of construction equipment, labour, but also the energy consumption during the time the building is in use. Each of these physical determinants may also have carbon emission consequences. The SIC algorithms consist of stepwise calculations of the determinants of a SIC. Descriptions of the main physical determinants covered in each SICs are given in Annex 1 of this report, and can also be retrieved in the model.

The parameter values ('*ratio values*') for many of the relations may be region specific. Therefore, the actual values are flexible elements in the algorithms and the relevant values are called from a database, based on their regional identification (see section 2.5). Using a set of regionalized values allows the model to generate regionally specified results. The use of a separate database allows the model to be more easily adjustable to new regions, because only the region specific parameter values need to be updated while the core algorithm remains the same.

Figure 4 below illustrates the calculation process from financial input (amount of euros invested in a given theme element) through physical determinants to carbon output (total CO_2 emissions) for the SIC 'Road construction'. In this example, ratio values – possibly region specific – are needed to convert monetary values (e.g., funds allocated to local road construction) into physical determinants (km of local road constructed) as well as to compute the resulting carbon impact of such physical determinants.

Total main determinant

As midpoints in the calculation between budget and carbon impact, the model uses central determinants that describe the result of the investment in physical units. For each SIC there is one central determinant, named *total main determinant* (TMD). For example, for the SIC 'road construction' the TMD is 'total length of road constructed (km)'. The TMD is not necessarily used in the model directly, but it is used as a method of control for model users; it provides a quick reference to gauge the order of magnitude of assumed physical activity related to the investment.

A simplified representation of the calculation procedure from theme element investment to carbon emission consequences is given in Figure 3.

SIC characterisation

SIC parameter values ('ratios') are generally fixed within a region. However, the model allows to differentiate the investment focus of the SIC across programme theme elements. The *characterisation* dimension represents parameters for which users can provide values to fine-tune the calculation of a SIC within a given theme element. For example, this allows the SIC ´road construction´ to use values typical for highways in one theme element, and values for local roads in another. In a way, the 'characterisation' could be regarded as an additional level of disaggregation in the financial model.

Different types of characterisations are used in the model:
- Characterisations with parameters in percentage. They are mainly used to further break down funds allocated to a SIC. For instance, in the SIC 'Road construction', the first characterisation is 'Investment per type of road' (in % of €). This characterisation is used to split funds allocated to road construction into local road construction, national road construction and highway construction.
- Characterisation with parameters in structural units (m², km, kg, kW, etc.). They are used to fine-tune the calculation by bypassing the part of the algorithm that would otherwise lead from the financial SIC specific input to the values desired for these parameters. For instance, for the SIC 'Road construction', one step of calculation consists in calculating the length of road constructed from an investment in Euros. If the length of the road is already known, by entering the length directly, the results for carbon emissions may be more accurate, as uncertainty concerning the algorithm has been reduced.

Figure 4 shows how the use of SIC characterisations leads to additional calculations in the calculation process.

Baseline activities

For properly evaluating the carbon emission consequences of investments, the emissions need to be calculated compared to a reference that describes the situation in absence of the investment. The SIC algorithms therefore account for the carbon emission impacts of the baseline activities. E.g. it is assumed that electricity produced from the SIC 'renewable centralised energy' would be produced through 'standard average' national power generation capacity in absence of the investment.

Although the activity levels and related carbon consequences of baseline activities could show trends in time, the current version of the model assumes a static activity level for baseline activities as a first approximation. Box 2 explains the reasoning behind this choice. The model structure, however, allows the development of time dependent

trends for the baseline activities and these could therefore be implemented in future versions of the model.

The SIC descriptions in Annex 1 shortly describe the baseline assumed for each SIC. The baselines considered relate mainly to the operation phase, and generally do not contain the carbon impacts of construction alternatives.
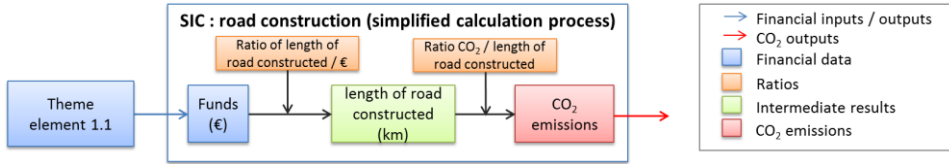
Figure 3: Simplified calculation process
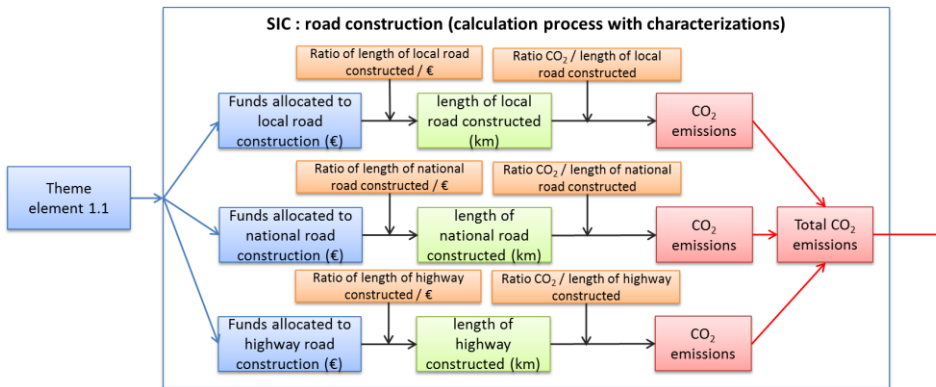
Figure 4: Calculation process including characterisations (for legend, see Figure 3)

2.3.2 Scope of evaluation

This section describes and discusses the model system boundaries. There are several dimensions for which the boundaries of the evaluation need to be defined.

- Temporal scope – which parts of the lifetime emissions related to investment are being evaluated?
- Financial scope – which investments are covered?
- Geographical scope – what geographical limitations are set with regard to evaluation results?
- Scope for impacts – which impacts related to the investment are being covered?

Temporal scope

Different choices in terms of how emission impacts across time are dealt with have different implications for, at least, uncertainties of the results, usability of the model, consistency of the methodology and model complexity.

The model takes into account all CO_2 emissions that are linked to the *construction* phase and *operation* phase of the evaluated investments. Emissions in these phases may be differentiated into 'direct' and 'indirect' emissions. Direct emissions are those that occur immediately as a result of combustion of fossil fuels or industrial processes; indirect emissions are derived through other processes that may occur (or have occurred) elsewhere, such as emissions related to electricity production or the embedded emissions of materials (emissions related to the production process and

freight of the materials). For further discussion see the discussion of the 'geographical scope' below.

Emissions related to the disposal phase are generally not included in the evaluation; usually, the disposal phase requires additional funding and as such could be covered as the emission impact of those investments separately.

The various SICs available in the model have different lifetimes. Assumed duration phases for the SICs can be retrieved in the model by clicking the button 'See/modify SIC parameters' in the 'SIC Management interface'. Evaluating the emission impact over the total lifetime of the investments therefore leads to a number that covers different timeframes for different parts of the results. Although this approach is useful and very straightforward for evaluating the total impact of investments, such a single number offers less guidance when interest is mainly in evaluating specific emission reduction targets in a given year. For such an evaluation, a result that presents indicators for a given fixed period is more useful. Also, as the actual lifetimes of investments are uncertain and the results show the relative impacts compared to a baseline which is also progressively uncertain in time, model users may be less interested in results for impacts that take place decades from the moment of investment. For above reasons, the evaluation period used in the model can be selected by the user. It can be set to cover the total lifetime of the SICs or a selected, fixed period (1, 2, …, n, number of years).

The model provides two types of results:
- Total cumulative emissions over the evaluation period, given separately for the construction and operation phase.
- An emission profile in time representing the cumulative emissions of construction and operation phase.

Although the model results are presented as an emission profile, the model does not specify an actual timeline that indicates the emission impacts for investment in a specific year. This is because investments in a regional Operational Programme may be started at any moment within its 7 years planning period; the model tracks emissions related to the investment from the start of the construction phase. For better overview in the results, all construction phase emissions are cumulated and presented in year 1, even though construction may sometimes cover a significantly longer period.

The model also provides emission indicators outside of the results section, that show a value for emission per euro invested in a theme element. These emission indicators can be used by the model user to get a quick indication of the relative consequences of investing in various theme elements. These indicators show the average emission impact over the full lifetime of the SIC. As this value consists of a combination of construction phase and operation phase emissions, it is an imperfect indicator for any evaluation that does not cover the full SIC lifetime.

Box 2: Considerations with regard to the evolution of the exogenous baseline factors and SIC parameters

A counterfactual baseline definition is essential for determining the emission impacts of the SICs. What complicates this task is that the further into the future one needs to look, the more difficult it is to have confidence in the projection that is being made. For Clean Development Mechanism, for example, this is reflected in that the baseline needs to be reassessed every seven years, if renewal for the crediting period is being sought.

In light of the above, a few options for the baseline definition exist. The simplest one is to assume a "frozen baseline", i.e. a scenario in which everything stays exactly as today for the relevant time frame. This approach is robust and transparent in the sense that it does not necessitate additional assumptions for how the indicators may evolve in the future. At the same time, however, it is likely to be accurate only for a relatively short period, for the same exact reason, because it does not take into account any additional knowledge that may exist about future trends that appear to be visible already today.

A second option would be to choose a set of key indicators (e.g. emission factors for cars, power generation) for which we can see clear trends and implement this projection into the analysis. This approach is likely to give better emission implications for the SICs that use these parameters, but would create an imbalance between the projects that use future projections and those that don't (e.g. if carbon impact of electricity generation is projected into the future, but the carbon content of material production is not, model would suggest that construction remains as carbon intensive as during the base year whereas electricity use would, in the future, be less harmful).

Finally, practically every factor can, and probably will, change, if a long enough time period is being considered; construction methods, process technologies for material production etc. are all likely to change and it may therefore be very difficult to determine scenarios that would cover the full (or adequate) range of the possible scenario space. Additionally, even if this was possible, any individual scenario combining a single, unique combination of parameters would, in practise, be so unlikely that its probability would be very close to zero.

In light of this it was decided to use a **frozen baseline**, at least as the first step. The model is, however, constructed in such a way that it is possible to add time dependent trends to parameters, even if they are not used in the current formulation of the model.

Financial scope

The model converts financial resources allocated to a SIC (e.g. road construction), into physical quantities (e.g. km of road constructed), which then lead to changes in CO_2 emissions, meaning that the model is very sensitive to financial inputs as these lead directly to physical activities and emission impacts. As the model is driven by the monetary inputs, it is important that the financial scope of evaluation is well-understood. Projects that receive funding through the European Regional Development Funds (ERDF) generally are co-financed by other public and private funds. The model

assesses the impact of the total activity financed by the combined "ERDF-funded" and "non-ERDF-funded" (all other) expenditures.

This is important for understanding the results by model-users as well as for being able to e.g. correctly update ratio values for model developers. If, for example, values that relate only to the ERDF-funded investments were used, emissions per euro would look different compared to if the assessment was considering the fact that usually also other (co-)funding sources contribute to the same projects.

Geographical scope

Investments made within a region are likely to have, in addition to direct emission effects (due to e.g. emissions from fossil fuel combustion at the site of construction) inside the region, spill over emission effects both within and outside the region. Which geographical system boundaries are the most relevant for the model depends on the main purpose of the model within the region.

For example, if the goal is to determine whether an investment is likely to make it more or less difficult to reach regional CO_2 emission targets, effects of the investment on emissions outside the region would be irrelevant. If the goal is related to national non-ETS targets, direct impacts and spill over effects that affect other sources of non-ETS emissions within the region and in other national regions would be relevant (but not, for example, sources within the region that are part of the ETS, such as the power sector). If the aim is related to a general environmental assessment of actions, e.g. to evaluate the global emission impacts of decisions, system boundary should be extended to cover any possible impacts, no matter where and how these take place. The choice on how system boundaries are placed is therefore essential for determining the types of questions the model may be able to address.

The feedback from the intended users in the model testing regions indicated that most of them are interested in the global emission effect rather than strictly regional effects. Therefore, total model results principally reflect the emission impact on a wide geographical scale, including those that are possibly, or likely, to take place outside the region/country.

However, in order to allow a wide range of uses for the model, the model provides some disaggregation of impacts, at least on a rudimentary level, that allows the user to focus on emissions that are most relevant for his analysis. The impact results are therefore disaggregated into:

- **direct emissions impacts (within the region/country):** physical emission changes originating directly from the project, e.g. switching from oil fired boiler into a bio fuel fired one.
- **indirect emissions impacts:** e.g. changes in electricity use, induced by the production/consumption of electricity by the project, or emissions 'embedded' in materials (i.e. emissions related to the material production process, including delving, processing and transportation to construction site).

This division provides the user with a set of building blocks that can be used to answer different types of questions. As the option of extending the system boundaries beyond the region/country greatly increases the data needs, it was decided that **the emphasis in terms of data collection and impact assessment is, at this stage, on the regional/national emission impacts**. The emission implications of e.g. material use are included, but the evaluation is kept simpler than for emissions that are more directly

attributable to the SIC (e.g. only the most relevant construction materials are included and material use during operation is largely omitted).

The benefit of the chosen approach is that it allows the user to draw the system boundaries of his evaluation. The approach also covers all the emission sources and should therefore enable scenario comparisons, no matter which emission mechanism is the main contributor. The drawback of this approach is that, compared to considering regional or national emissions alone, it requires a more complex structure, even if it is not filled with very complex data. Furthermore, it is important to keep in mind that although all emission sources are represented, some impacts may be based on detailed evaluation of all relevant production processes, replacement mechanisms and logistical chains, whereas others might be better described as proxies or place holders for the more detailed data collection, which has been considered outside the scope for the model development.

Impact scope
The model evaluates the investments on carbon impacts only. Investments will obviously have other impacts as well. Many of the investments intend mainly to strengthen the regional/national economic structure and any environmental considerations are, while important, still secondary in terms of determining the useful investments. Insight in terms of the impacts on economic growth and job creation is therefore indispensable when evaluating the expected consequences of a proposed Operational Programme. Information on CO_2 implications of these activities is likely to serve as secondary information that may however still enable optimization of decision making over multiple goals.

However, it should also be noted that the model's focus on CO_2 alone means that it is also incomplete from an environmental – or even climate change – perspective. Investments are likely to have also other, non-CO_2 emission impacts; those related to other environmental impact categories (e.g. acidification, eutrophication) as well as other greenhouse gas emissions that are not included. For most investments, other greenhouse gases contribute only to a relatively limited extent to the total climate impact. However, it should be noted that for some investments, such as waste and wastewater treatment, methane and nitrous-oxide emissions impacts may surpass the CO_2 related impacts.

2.4 SIC calculation procedure

2.4.1 Algorithms to calculate SIC specific effects

This section provides step by step details for a SIC algorithm. It shows how inputs, intermediate results, and ratios are used to calculate CO_2 emissions for each SIC. Throughout the section, the algorithm principles are illustrated using as example the SIC 'Road construction'.

2.4.2 Basic algorithm rules

The model uses a stepwise calculation approach for each SIC. Technically this means that values are calculated one after the other, using output of calculation *n* as an input for calculation *n+1*. A calculation uses three types of input:

- The funds allocated to the SIC.
- The value of the 'characterisation' entered by the user (it should be reminded that this data entry is optional and the model uses default data when no value is entered).
- A set of specific ratios dedicated to each SIC and each region/country (some ratios may be common to several SICs and/or regions).

The general logic of the algorithm used by the model is generic and thus common to all SICs. Table 2 shows a simplified example of how the calculation proceeds for the SIC 'Road construction'. The model calculates determinants line by line, following the step defined in the ´Step´ column from the smaller value the higher. Several ratio values required for this calculation are given in Table 3.

Step	Input determinant		Ratio		Output determinant
0	Investment in road construction (€)	x	Share of local road (%)	=	Investments in local road (€)
0	Investment in road construction (€)	x	Share of national road (%)	=	Investments in national road (€)
0	Investment in road construction (€)	x	Share of highway (%)	=	Investments in highway (€)
1	Investments in local road (€)	x	Ratio of km / € (local road)	=	Length of local road (km)
1	Investments in national road (€)	x	Ratio of km / € (national road)	=	Length of national road (km)
1	Investments in highway (€)	x	Ratio of km / € (highway)	=	Length of highway (km)
2	Length of local road (km)	x	Bituminous concrete requirement for local roads (t/km)	=	Bituminous concrete requirement (t)
2	Length of national road (km)	x	Bituminous concrete requirement for national roads (t/km)	=	Bituminous concrete requirement (t)
2	Length of highway (km)	x	Bituminous concrete requirement for highway (t/km)	=	Bituminous concrete requirement (t)
3	...	x	...	=	...

Table 2: Simplified calculation chain for the SIC 'Road construction'.

Ratio	Value
Share of local road (%)	100 %
Share of national road (%)	0%
Share of highway (%)	0 %
Ratio of km / € (local road)	1.9 E-6 km/€
Ratio of km / € (national road)	4.1 E-7 km/€
Ratio of km / € (highway)	1.4 E-7 km/€
Bituminous concrete requirement for local roads (t/km)	1 490 t/km
Bituminous concrete requirement for national roads (t/km)	4 040 t/km
Bituminous concrete requirement for highway (t/km)	6 200 t/km

Table 3: Default values of the ratios used in the first steps of the algorithm for the SIC 'Road construction'

STEP = 0
At the first line of calculation, the model uses as input the "Investment in road construction (€)". From this value, the model calculates:
1. Investments in local road (€).
2. Investments in national road (€).
3. Investments in highway (€).

To perform those calculations, three ratios are used:
a) The share of local road (%).
b) The share of national road (%).
c) The share of highway (%).

The 'investment per road size' is defined as a 'characterisation' in the SIC road construction. Therefore the shares invested in each road size can be manually put in by the model user. If no manual input has been provided, the model will use its default values that are stored in the model database to execute this calculation step.

Thus:
• Investments in local roads = Investment in road construction x share of local roads.
• Investments in national roads = Investment in road construction x share of national roads.
• Investments in highways = Investment in road construction x share of highways.

The values calculated in step 0 are available for further calculations in the next steps of the algorithm.

STEP = 1
On the basis of the investments in local/national road or highway calculated in the previous step, the model calculates:

4. The length of local road (km).
5. The length of national road (km).
6. The length of highway (km).

To perform those calculations, three ratios are used:

a) A ratio of km / € (local road).
b) A ratio of km / € (national road).
c) A ratio of km / € (highway).

Thus:
- Length of local road = ratio of km / € (local road) x Investments in local road.
- Length of national road = ratio of km / € (national road) x Investments in national road.
- Length of highway = ratio of km / € (highway) x Investments in highway.

Like in the step 0, users have the possibility to directly enter the value of the length of road constructed per type, because this is defined as a characterisation. Values put in manually for characterisations will 'overwrite' the value calculated by the model.

STEP = 2

The third step in this example differs compared to the two previous steps: the output determinant is equal in all lines of this step. If, within a calculation step, output determinants are equal, the algorithm recognizes this and it will sum the values calculated in each line rather than treat it as individual values.
Thus:

Bituminous concrete requirement =	Length of local road (km) x Bituminous concrete requirement for local roads (t/km) + Length of national road (km) x Bituminous concrete requirement for national roads (t/km) + Length of highway (km) x Bituminous concrete requirement for highway (t/km)

Note that further steps are used to calculate emissions from bituminous concrete requirements.

2.4.3 Further algorithm rules

After explaining the general methodology of the algorithms through the simplified algorithm in the previous section, this section presents the algorithm rules as they are used in the model.

The SIC algorithms in the model are structured as in Table 4 below.

Step	Determinant A	Determinant B	Output determinant	Ratio	Operation
0	Investment in road construction (€)		Investments in local road (€)	Share of local road (%)	0
0	Investment in road construction (€)		Investments in national road (€)	Share of national road (%)	0
0	Investment in road construction (€)		Investments in highway (€)	Share of highway (%)	0
1	..	..	..		

Table 4: Structure of the table that contains SIC algorithms

There are two major differences compared to the simplified algorithm presented before:

- To enable combining two output determinants from earlier lines in the algorithm, the algorithm uses a determinant A and a determinant B, in addition to calling a ratio from the database.
- To enable multiple mathematical operations, the column 'Operation' is introduced. Based on the value in the cell of this column, the model determines the mathematical operation used to calculate the output determinant. The operations available in the model are described in the Table **5**.

Operation	Formula to calculate the output determinant
0	Output determinant = Ratio x Determinant A x Determinant B *(if Determinant B is empty, Determinant B =1)*
1	Output determinant = Ratio
2	Output determinant = Ratio x (Determinant A / Determinant B) *(if Determinant B is empty, Determinant B =1)*
3	Output determinant = Ratio / (Determinant A + Determinant B) *(if Determinant B is empty, Determinant B =0)*
4	Output determinant = Ratio / (Determinant A x Determinant B) *(if Determinant B is empty, Determinant B =1)*

Table 5: Value of 'Operation' and calculation of the output determinant

Thus, for each line in the table, there are:
- Two inputs: determinant A and determinant B,
- One output: output determinant,
- One parameter value: ratio,
- One type of operation: operation.

One should be reminded that the algorithm does not contain any parameter values directly. Instead, each individual parameter is represented by an individual identifier. The value of the ratio is stored in a database that contains for each ratio:
- The ratio identifier (ID).
- The ratio name.
- The region/country the ratio is associated to including its geographical (NUTS) level.
- The ratio value.

Thus, depending on the region/country the model is used in, the model automatically uses the value of the ratio related to the region/country. If the regional value for the ratio is not available, the model will use the national value if available or default/global/EU value. Also, this allows adjusting the ratio when required without intervening in the algorithm.

2.4.4 Emissions calculation

Emissions are the final outputs of all algorithms. Since emissions are categorized into phases (construction, operation) and scopes (direct and indirect), there are four different types of emission results:

- Construction, direct emissions (tCO_2).
- Construction, indirect emissions (tCO_2).
- Operation, direct emissions (tCO_2/y).
- Operation, indirect emissions (tCO_2/y).

Operation emissions are estimated on a yearly basis. Therefore, if the user wishes to analyse cumulative emissions over the lifetime of projects, emissions must be multiplied by the duration of operation phase. Expected operational lifetime is set for all SICs in the model. For instance, for the SIC 'Road construction' lifetime duration is set at 30 years. Default values for SIC lifetimes are set in the model database, but can be adjusted by the model user through the SIC Management interface.

2.5 Ratios

The strength of the model lies in both the algorithms and the database of ratios used by the algorithms. There are numerous types of ratios used in the model, such as emission factors, distribution of investment, quantity of material per km of road, traffic flow generated by km of constructed road, etc.

The algorithms of the generic model use 902 individual parameters. As many ratios may be defined for each EU region, a complete database for all EU regions would contain a multitude of this.

As mentioned in section 2.4 all ratio values are stored in the model's ratio database and are called upon by the algorithm through unique identifiers (ID numbers).

2.5.1 Common ratios vs. SIC specific ratios

Some ratios are common for multiple SICs. All ratios with an ID between 0 and 999 are common ratios. For instance, emission factors are often common ratios.

ID ratio	Ratio's name
110	Emission factor of electricity (tCO$_2$/kWh)
408	Embedded emission factor of concrete (tCO$_2$/t)
202	Emission factor of freighting construction materials (tCO$_2$/t)
…	…

Table 6: Example of common ratios used in the model

All ID ratios over 1000 relate to ratios which have specifically been created for this SIC. The next table introduces some examples of specific ratios used for the SIC 'Road construction'.

ID ratio	Ratio's name
1000	Part of total investment in local road construction (%€)
1003	Length of local road constructed per € invested (km/€)
1006	Quantity of bituminous concrete per km of local road (t/km)
…	…

Table 7: Example of specific ratios used for the SIC 'Road construction'

2.5.2 Spatially differentiated ratios – EU, national and regional

As the model is regionalised, some ratios are specific to a geographical area. Depending on available information and relevance, they can be specific to a region (NUTS1, NUTS2) or to a country (NUTS0). Those ratios are called 'territorialized ratios'. Ratios that are not geographically dependent and thus valid for all the countries of the European Union are called 'global scale ratios'.

In terms of algorithm, when the model is implemented in a region, the model looks up the value with the most regional detail. Each time a ratio is used in the algorithm, the model searches the ratio database for a value corresponding to the NUTS2 code of the region. If a NUTS2 value is not available, it will search the ratio at NUTS1 scale, then at NUTS0 scale or finally at European scale.

The next table presents the values for two ratios used in the SIC 'Road construction'.

ID ratio	Ratio's name	Territory	Value
1006	Quantity of bituminous concrete per km of local road (t/km)	EU	1490
303	Emission factor of personal car (tCO$_2$/pass.km)	CZ	0.00014004294
303	Emission factor of personal car (tCO$_2$/pass.km)	DK	0.00009863412
303	Emission factor of personal car (tCO$_2$/pass.km)	DE	0.00009682000
303	Emission factor of personal car (tCO$_2$/pass.km)	IE	0.00012820941
303	Emission factor of personal car (tCO$_2$/pass.km)	EL	0.00009160294
303	Emission factor of personal car (tCO$_2$/pass.km)	ES	0.00010843529
303	Emission factor of personal car (tCO$_2$/pass.km)	FR	0.00010978941
303	Emission factor of personal car (tCO$_2$/pass.km)	LV	0.00011888529
303	Emission factor of personal car (tCO$_2$/pass.km)	HU	0.00015304235

ID ratio	Ratio's name	Territory	Value
303	Emission factor of personal car (tCO$_2$/pass.km)	NL	0.00012590059
303	Emission factor of personal car (tCO$_2$/pass.km)	AT	0.00013800412
303	Emission factor of personal car (tCO$_2$/pass.km)	PL	0.00009098824
303	Emission factor of personal car (tCO$_2$/pass.km)	PT	0.00010956294
303	Emission factor of personal car (tCO$_2$/pass.km)	RO	0.00013799412
303	Emission factor of personal car (tCO$_2$/pass.km)	SI	0.00011729059
303	Emission factor of personal car (tCO$_2$/pass.km)	SK	0.00024763176
303	Emission factor of personal car (tCO$_2$/pass.km)	FI	0.00008956235
303	Emission factor of personal car (tCO$_2$/pass.km)	SE	0.00010686706
303	Emission factor of personal car (tCO$_2$/pass.km)	UK	0.00010537882
303	Emission factor of personal car (tCO$_2$/pass.km)	CZ	0.00014004294
303	Emission factor of personal car (tCO$_2$/pass.km)	DK	0.00009863412

Table 8: Example of two ratios used for the SIC 'Road construction'

In the example in Table 8, ratio 1006 is defined at European scale, whereas ratio 303 is territorialized at NUTS0 scale. Details on the ratios used in each SIC algorithm are available in the model file. Meaningful use of the model requires the availability of a complete set of regionally applicable values in the database. The generic version of the model has datasets available for the 4 EU regions and one country in which the model was applied as part of the model development process. How to prepare the model for other regions/countries is described in Chapter 3.

2.5.3 Data sources

As in any model, the data used for its parameters, as much as the calculation procedure, is crucial for the representativeness of the results. Data between different model elements need to be consistent as much as possible, represent the same geographical and temporal scope and, if relevant, aggregate similar components. In a model that covers a scope of activities as wide as the CO$_2$MPARE model, data consistency may be regarded as a challenge. Generally speaking, data for the different activities covered by the SICs can only be collected by consulting and combining data from multiple sources. As such, discrepancies in scope and representativeness are virtually unavoidable.

Given the high aggregation level of the model results, some inconsistency of the model seems acceptable. After all, as it will be discussed in more detail below, the model approach groups a large variety of project activities into the same SIC. The model results should be interpreted in the scope of their emerging overall picture rather than the exact individual numbers. Parameter values therefore need to reflect the proper order of magnitude rather than provide details to six decimals.

The data sources used to populate the generic model and its regional applications are available in the 'Ratio management interface'. Data sources include national and European statistics, but also scientific papers, or reports describing results in LCA databases. The issue of consistency suggests that any further data collection required should focus on using data from available national or European statistics as much as possible, although it is recognized that using incidental data from other sources cannot be avoided.

2.6 Uncertainties and Sensitivities

The complexity and heterogeneity of projects covered by the SICs, the long evaluation horizons and uncertain parameter values result in high uncertainty surrounding the model results. Additional uncertainty in the approximations can be caused by small sample sizes, i.e. if it is assumed that the SIC describes an 'average' project, a number of projects is needed before the emission impact is likely to approach the expected value. Sub-regional characteristics of individual projects, as well as developments in the costs of projects can also lead to a wide range of emission outcomes per euro invested. First and foremost, the model results should therefore be interpreted as estimates.

Due to the long time frames and wide scope of activities and sectors covered by the SICs, capturing and adequately qualifying all relevant structural and parametric uncertainties in the model, including their correlations, is practically infeasible. Moreover these uncertainties are likely to vary per region or per country and would therefore need to be redefined for each regional/national application. Actually calculating the expected uncertainty ranges of the outcomes therefor is not possible in the model.

However, the model does provide an indication of model sensitivities on a more general level of the model. This indication can guide the user to theme-elements or SICs for which uncertainties are high and that contribute considerably to the results. It guides the user to the model parameters that require most scrutiny in order for the model to provide most relevant results.

This approach for the sensitivity analysis is based on defining uncertainty ranges for the emission indicators (tCO2/€) of individual SICs level and focus on how these SIC specific uncertainties are reflected on the programme/theme level. The approach can therefore be described as a sensitivity analysis in a probabilistic framework.

Quantifying SIC specific uncertainties
As stated above, the approach to identifying and quantifying individual uncertainties is highly stylized and the analysis focus is on model sensitivity. Following this, the uncertainty for the emission impact (in CO2/€) of a given SIC is simply determined as **1) High 2) mid or 3) low**. This qualitative default assessment of the SICs can later be altered by the user of the model. Each of the three uncertainty classes links to a specific 95 % (two standard deviations) confidence interval for the emission indicator: Low ± 10 %, Mid ± 50 % and High ± 100 %. It is also assumed that the uncertainty is normally distributed and that the SIC specific distributions are independent. These assumptions are simplifications of reality, but little data exists to point out more appropriate distributions and correlation factors. Also, the assumptions greatly simplify the process of summing across the SIC specific distributions.

Assessing the uncertainties for the outputs
Normally distributed variables can be summed easily and the resulting distribution is always also normally distributed. More precisely, the mean of the resulting distribution is the sum over the means of the SIC specific distributions. The variance can be summed

similarly, for independent variables. These two parameters define the new normal distribution.

The above summation process can be done for any level of aggregation deemed relevant, i.e. one can sum over all SICs (for total emissions), over the SICs related to a specific budget line or just focus on individual SICs and their emission ranges. The result will always be precisely defined normal distributions, which means that it is easy to define, for example, confidence intervals around the aggregated emission impacts.

2.6.1 Uncertainty presentation in the result interface

The uncertainty results on a sheet 'Uncertainty' will be accessible if the user is in the expert mode. All the results presented on this sheet use the same user defined length of evaluation as the rest of the evaluation does and only cumulative results are presented.

Graphs and tables are drawn for total emissions and for emission per theme. The user can decide the confidence interval for the results (same interval applies to all emission indicators, no matter the aggregation level) and the upper and lower emission estimates of the interval will be given in the tables and shown in figures. An additional table is given for SIC specific variance and standard deviation.

Also, top five SICs contributing to uncertainty are highlighted based on two alternative criteria: 1) by their contribution to the variance of the total emission distribution or 2) by the relative contribution, in comparison to the funding an individual SIC receives. The former highlights SICs that contribute significantly to uncertainty, because of high uncertainty, emission impact or financial input. The latter removes the impact of financial input and is therefore better for identifying the SICs that are uncertain by character.

Implicit assumptions concerning the uncertainties
Such a stylized sensitivity approach makes a number of assumptions, some of which have clear implications for what uncertainties this approach is implied to capture. These implications are briefly discussed here.

Firstly, it is assumed that the SIC specific uncertainty is unaffected by the volume of investments into a SIC. In reality a higher number of projects would imply greater heterogeneity across the projects and would therefore reduce SIC specific uncertainty. The fact that this is not considered means that it is implicitly assumed all the individual projects related to a SIC are perfectly correlated and have identical standard deviations. This means that uncertainty is likely to be overestimated on this level[1]. From the above, it also follows that the uncertainties related to the heterogeneity of projects under a given SIC are not captured, but only those related to parameter values and structure of the one generic project that underlies a SIC (i.e. the model operates on the aggregated level and uncertainties can therefore only be represented on this level).

---

[1] Assuming that the SIC level uncertainties would describe accurately also project level uncertainties.

Secondly, since project heterogeneity is not accounted for, the expected emission impact of a SIC does not depend on the sample size either, i.e. no additional assumptions concerning the sample size are needed as the assumptions concerning sub-SIC project correlation and standard deviation already lead to this.

Thirdly, all the SICs are assumed to be independent, so that if, for example, emissions for a given SIC would be above the expected values, this knowledge does not in any way change the emission distributions of the other SICs. This is also inaccurate, as many SICs rely on same parameter values, reflecting, for example, emission coefficients for materials or electricity use. This assumption therefore leads to underestimate the aggregated uncertainties, as introducing new SICs always works as a diversification measure, no matter how much the emissions of the new SIC depend on parameters shared with SICs already in the portfolio.

Generally, the approach captures (and overestimates) the value of diversification across the SICs, but does not capture heterogeneity on the sub-SIC level (focusing instead on parametric uncertainty for a typical project). As the former leads to an underestimation and the latter to overestimation of uncertainty, these two inaccuracies caused by the assumptions at least partially counteract each other.

2.7 Comprehensive overview of the model

As discussed in the previous sections, the model consists of multiple elements that communicate with each other. Figure **6** (at the end of this section) gives a comprehensive overview of the interaction between the various elements.

The communication procedures in the model are written in the programming language Microsoft Visual basic for Excel. For specific programming details of the model, the source code of the procedures is available via Excel's visual basic editor[2].
Figure 5 simplistically presents the core calculation procedure. It shows that the model runs the various SIC algorithms in a loop for each theme element. For each theme element, it retrieves the relevant values per SIC from database or user interface. It does so for each SIC in the theme elements. When all SICs in the theme element are calculated, it stores the results for this theme elements and proceeds to the next theme element. This results in a result database that contains results for each SIC in each theme element. Different cross sections through this result database then provide the results; per SIC, per theme element, per theme, or for the total programme.

---

[2]   There seems to be limited value in trying to translate the programming language to text and explain these procedures in detail in this document.

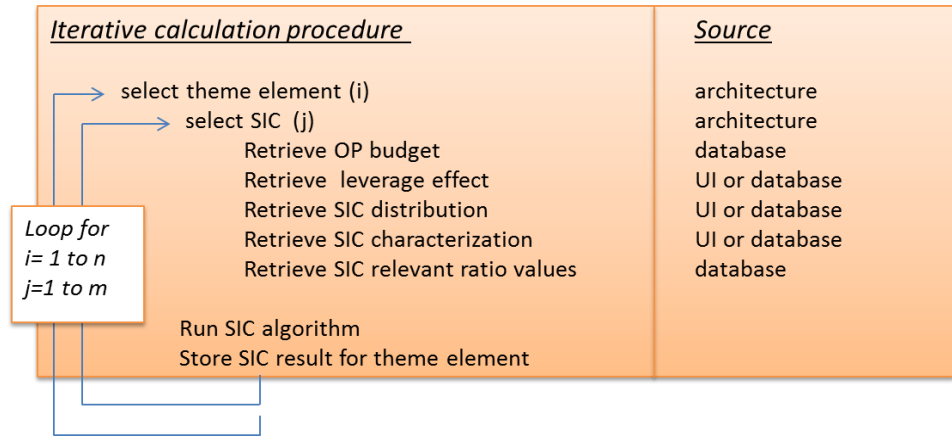| Iterative calculation procedure	Source
select theme element (i)	architecture
select SIC (j)	architecture
Retrieve OP budget	database
Retrieve leverage effect	UI or database
Retrieve SIC distribution	UI or database
Retrieve SIC characterization	UI or database
Retrieve SIC relevant ratio values	database
Run SIC algorithm	
Store SIC result for theme element	

Loop for
i= 1 to n
j=1 to m

Figure 5: Iterative calculation procedure

n = number of theme elements

m = number of SICs in theme element
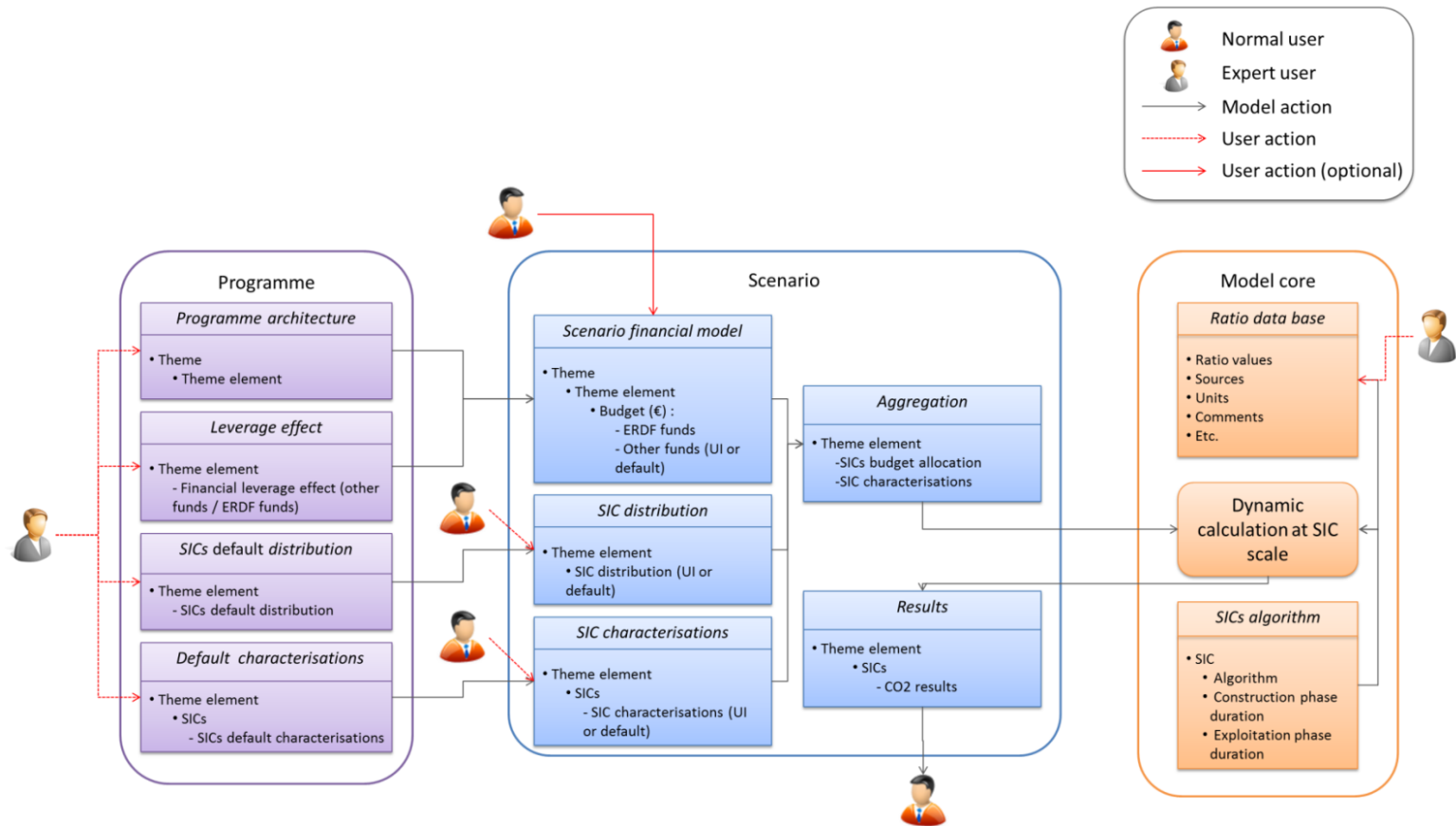
UI = user interface

Figure 6: Comprehensive overview of the model elements

3

Deploying the model in new regions and countries

3.1 Introduction

The generic model has been developed to function as a template that enables deploying functional regional and national applications for every region and country in the EU. The test applications have been developed in close cooperation with five testing-regions. A functional application has been developed for each of them. This chapter will discuss the approach model developers could take to deploy the model in a region/country for which no regionally/nationally functional application yet exists.

There are many differences between the various regions and countries of the EU. Differences relevant for the scope of the model include indicators such as population density, geography, climate, price level, income, economic structure, share of various transport modes, portfolio of electricity generation capacity, etc. Regional or national differences may cause different priorities in investment programmes as well as lead to different climate impacts from the same type of investment. Application of the model in new regions/countries requires that the values used in the model correctly represent such specific conditions. Moreover, application in a new region or country requires that the available programme architecture in the model reflects actual priorities, and that the SICs connected to the various theme elements optimally represent the actual projects financed in the Operational Programme.

To deploy the model in new regions or countries, therefore requires
1. Preparing a programme architecture, including the link between the relevant SICs to its theme elements
2. Regionalizing the ratio values in order for the model to provide regionally/nationally relevant results.

Section 3.2 and 3.3 discuss these steps. Section 3.4 then discusses the process of collecting the required data.

3.2 Creating a programme architecture

For a regional/national application of the model to function properly, the specific Operational Programme architecture should be reflected appropriately in the model. This step involves:

1. Making sure the appropriate themes, sub-themes and theme elements to represent the formal budget architecture.
2. Linking the relevant SICs to each of the theme elements.
3. Setting default values for the share of budget allocated towards the various SICs in each theme element.

As the approach to link the relevant SICs to the theme elements almost simultaneously provides default values for the share of budget to be allocated towards the various SICs, these elements are discussed in combination.

3.2.1 Preparing the financial architecture

The default architecture for Operational Programmes in the generic model reflects the list of ´priority themes´ and ´categories of expenditure´ provided in Commission Regulation (EC) No 1828/2006 (see box 2). The list of theme elements eligible for ERDF financing in the future programme periods is likely to differ from this list. Moreover, national and regional authorities frequently use a customized architecture for an Operational Programme that differs from the architecture used by the European Commission. Therefore, a model developer deploying the model in a new region or country should make sure that the appropriate themes, sub-themes and theme elements are available in the model to represent the formal budget architecture used in the Operational Programme. The approach to adapt or create a new architecture is presented in the user tutorial, and will not be further discussed here.

3.2.2 Linking SICs to theme elements

The SICs connected to the theme elements may differ widely across regions/countries. Very different types of projects can be funded within the same theme element. This section discusses how to assess ex-ante which SICs should be connected to each of the theme elements. It simultaneously provides default values for the share of budget allocated towards the various SICs in each theme element. Examples from the process followed to develop the model in the five test regions are used.

This step of model deployment involves assessing which type of activities (and therewith SICs) are likely to be financed through each of the theme elements. As the model aims to provide ex-ante evaluation of the emission impact of future Operational Programmes and the model would preferably be deployed in a region/country even before that, the physical characteristics of projects to be financed through the Operational Programme are often largely unknown at the time of deployment of the model in a new region/country. Usually, the best information available is based on the realizations of existing or past Operational Programmes. Therefore, linking SICs to

theme elements requires extracting this information from current or past Operational Programmes.

In a nutshell, the approach elaborated below assesses which activities are financed in the current or previous Operational Programme by evaluating the projects it financed, and linking SICs to these projects.

Breaking down the financial allocation of a theme element to SICs requires analysing the regional/national Operational Programme and projects they fund, and thus gathering a set of information for each OP (referred to as "OP related data"). The breakdown involves several stages:
1. Selecting a representative sample of projects for each OP's theme element (further detailed below),
2. Manually link each project of the sample to SICs,
3. Estimating the financial weight of each SIC for each theme element.

In order to perform this breakdown, specific data about the architecture of the OP and about projects funded need to be collected. Objective and needed data are described in Table 9.

Objective	Needed data
1. Select a representative sample of projects to launch the analysis (data needed for each project of the OP)	Database of funded projects in OPs containing, for each project : • ID number of the project • Description/name of the project • ID and name of the theme element
2. Associate each project of the sample to SICs	• Budget allocation (eligible and total expenditures of projects) • Qualitative and quantitative indicators • Any other significant information about the project

Table 9: Definition of the OPs related data needed according to the different methodological objectives

Each project that is part of the projects sample is linked to one or several SICs. SICs are then gathered at theme element scale and the relative financial weight (share of the funds allocated to each theme element) of each SIC within a theme element is calculated, based on projects' financial weight. The diagram in Figure 7 sums up the process.
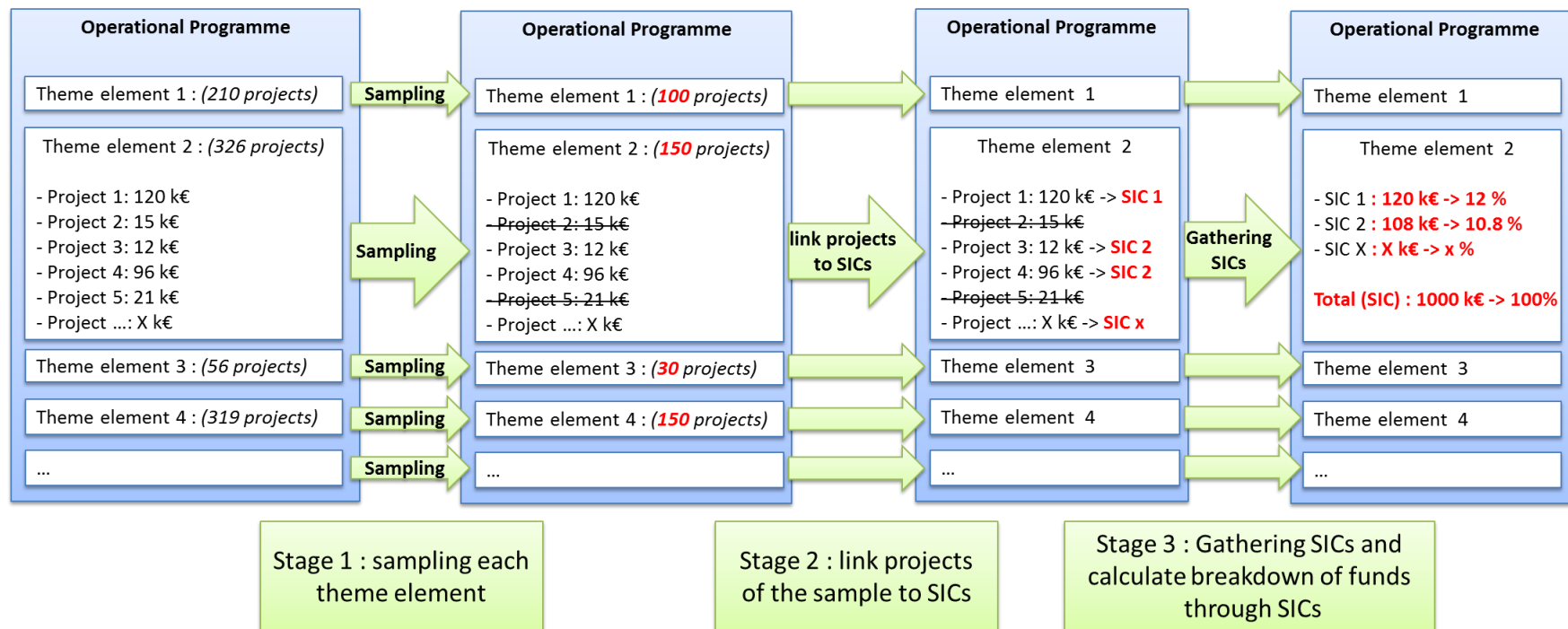
Figure 7: Data treatment process for OPs related data

Selecting a representative sample of projects

Projects need to be linked to SICs on a case-by-case basis. Moreover, it is not possible to automate this process. This step in the process therefore requires significant effort. For programmes with large numbers of projects, looking at every single project would be very time consuming. It is therefore generally preferable to select a representative sample of projects, to decrease the workload. This sample should be representative not only with regard to amount of funds, but also with regard to diversity of projects. For programmes with fewer projects, sampling may not lead to representative results and should therefore not be performed.

The following methodology is proposed to perform sampling:

First, the projects in an OP are grouped per theme element. For each theme element, a sample should be taken.

To create a relevant and representative sample, sampling is performed according to the following principles:

- As small projects are usually quite different from bigger ones, the sample should include projects with large as well as projects with smaller budgets.
- In order to cover a large part of the financial budget while keeping the total number of projects in the sample relatively small, projects with a larger budget should have a higher chance of appearing in the sample.
- Projects should be spread over the theme elements in a programme.
- Apart from these principles, samples should be selected randomly.

In order to be able to follow these principles, the following procedure has been adopted:

- Projects have first been classified on costs. Ten classes have been defined so that the total cost of each class is equal to one tenth of the total funds of the theme element.
- Then, the sampling is done in each class so that we are sure to get a real diversity of projects in terms of typology. For this sampling, projects are given a probability to be selected based on the size of their budget. This probability is defined as the cost of the project divided by the total of the costs of the projects within a theme element. This principle leads to select a financially representative sample of projects.
- To "spread" the sample over the different theme elements, the number of projects drawn per theme element should be weighted according to the weight (in terms of number of projects) of the theme element over the total programme and its financial weight over the total budget, using the following formula: $\%_{\text{of drawn projects}} = (\%_{\text{number of projects}} + \%_{\text{total cost}}) / 2$.
- As a general rule, about one third of the projects should be included.
- If an Operational Programme consists of less than 500 projects, all projects should be included.
- To reduce workload, no more than ~1000 projects should be included.

Thus, if there are e.g. 600 projects to select among 2000, those 600 projects will be equally selected between the ten classes (thus, 60 per class), but the bigger projects within each class have a higher chance of being selected.

Figure 8 shows projects repartition per class with this system of equal financial weight of classes.
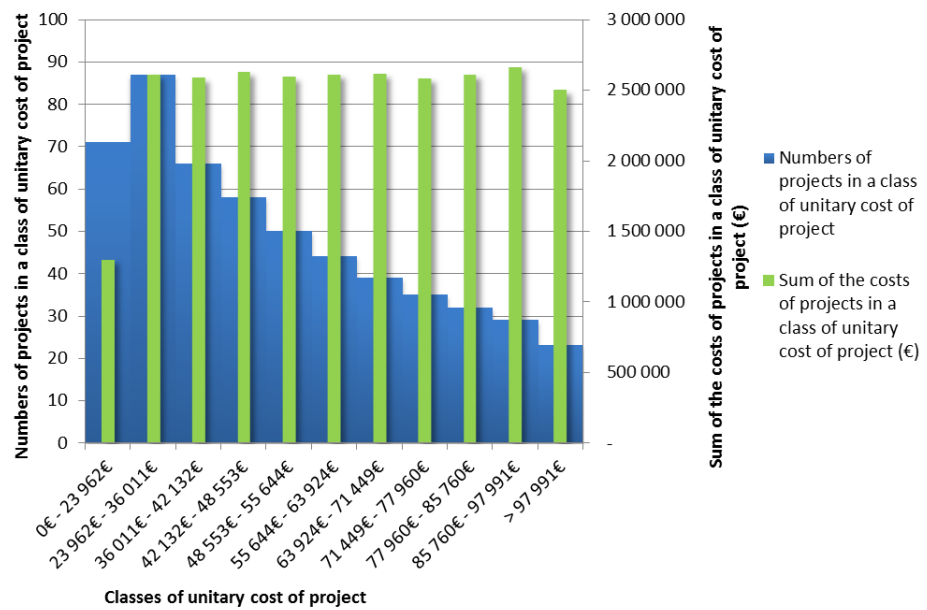

Figure 8: Number and total cost of projects classified per class of unitary cost of projects in a theme element (region Puglia). Classes are defined as financially equal.

For each test region, a sample of projects was determined. Table 10 presents the characteristics of the samples.

| | Crete | Emilia-Romagna | Puglia | Zuid-Holland | Czech Republic | |
					OPPI	OPZP
Number of projects in the programme	235	1 350	6 480	34	5 746	6632
Number of projects in the sample	235	441	1 029	34	1026	1079
Representativeness of the sample (% coverage of total projects analysed)	100%	32 %	16 %	100%	19 %	16 %
Financial representativeness of the sample (% coverage of total funds)	100%	81 %	39 %	100%	66 %	49 %

Table 10: Characteristics of the samples of projects determined for the test region

Processing the selected samples
Once a sample of projects had been defined, each project of the sample needs to be connected to a SIC. If it seems that no suitable SIC is available, the `configurable SIC´ can be configured to cover the desired investment component[3]. Through the sampling approach, it is also possible to calculate the relative distribution of each SIC to each

---

3   Ultimately, it is also possible to define a new SIC to be part of the model core by adding a new SIC algorithm to the model files and adjusting for this in the model visual basic code core.

theme element[4]. If needed, this procedure also provides information that allows refining the available set of SICs.

3.3   Regionalizing the ratio values

In order for the model to produce regionally or nationally relevant results, the values it uses in its calculation should represent the regional or national conditions. However, the model uses several thousand values in its calculation. Regionalizing each of these requires significant effort and resources. This section gives some guidance on how to approach model regionalization most efficiently.

Some ratios may be expected to show larger variation across the regions of the EU than others. This is why the model uses data on differentiated levels of geographical scale:
- European
- National
- Regional

As noted before, each data value in the model is connected to a regional identifier that enables the model algorithm to select the value with the most appropriate scale. The algorithm checks first if a regionally specific value is available, and if not, checks for a national value. If neither regional nor national value is available, the algorithm will use the value defined at the EU level. In principle, if limited variation is expected, using data with a higher geographical scale may be well applicable. However, for values with more significant variation across regions, national scale level or regional level data are more appropriate.

National scale data are often more easy to obtain than regional scale data; statistics are more often available for national indicators. Moreover, as more projects of similar nature are generally available on a national scale, national scale may allow assessing an average value for a specific indicator with higher robustness than basing it on one or two concrete examples on a regional scale (which could be outliers). Therefore, unless regional circumstances are thought to differ significantly from national average, using national scale indicators may be preferred.

Next to geographical representativeness, some ratios have a larger impact on the model result than others. To obtain regionally relevant results, regionalization should be considered more urgent for such ratios than for others. The ratio values for which regionalization is considered most urgent by the model developers are indicated in the model. The need for regionalization (high / medium / low / none) is one of the available indicators in the list of ratios, available through the ratio management interface or the SIC management interface. The approach for changing ratio values in the model database and for introducing new values for a specific region is described in the user tutorial.

---

[4]   It has been assumed that the financial disaggregation of theme elements to SICs was the same as for the sample.

Lastly, national and regional authorities may be more interested in some outcomes of the model than in others. Focussing the regionalization of ratio values to those model parts, while keeping other parts on a more aggregate level, could provide most value for the model user while saving effort where it is not really needed.

In order to most efficiently use available time to deploy the model in new regions and countries it is thus advised to

- First regionalize the ratios that have a large impact on the model result and that are expected to vary significantly throughout the EU, indicated in the model.
- Use national scale data for ratios that do not absolutely require regional scale.
- Focus further regionalization efforts on model parts with special interest from national and regional authorities.
- Further regionalization may be effectuated to obtain results that are even more customized.

3.4 Data requirements and collection

3.4.1 Data requirements and availability

To deploy the model in a new country or region, two types of data need to be collected:
1. Data related to the Operational Programme
2. Data related to the physical and financial properties of activities

Data related to the Operational Programme can only be collected in close cooperation with the management authorities. The management authority can provide the architecture that it is actually using, and an overview of projects that are being financed within each of its theme elements.

Data related to the physical and financial properties of activities are usually found in national or regional statistics. In specific cases, project case studies or (partial) programme assessments may provide the required data.

Missing data
In any given country or region, it is likely that not all data may be available at the required scale. In such cases, it may be appropriate to look for the required data in neighbouring regions or countries or regions in other countries that may have similar characteristics. In case relevant data appear to remain unavailable, expert judgement on adjustment of a higher scale ratio for regional application could potentially offer a solution[5].

---

5   A final solution would be to re-programme the given SIC algorithm so that it uses data that are available on regional scale. Re-programming a SIC algorithm goes beyond the scope of this guidance document - it requires adaptation of the model's operating core, and adjusting for this in the visual basic code.

3.4.2 Involving the future user

It is important to involve the intended users of the model in various stages of deployment. Firstly, without the consent of intended users, there would be small purpose in deploying the model in a region or in a country at all. Early involvement increases the feeling of ownership of the model and therewith the probability that it will actually be used for programme evaluations.

Secondly, involving intended users can facilitate obtaining data that would otherwise be difficult to obtain. Data related to, e.g., regional Operational Programmes need to be obtained from regional authorities. If the future user (at the regional authority) is involved in making this information available and aware of its purpose, it is more likely that information of sufficient quality is produced.

Thirdly, intended users may indicate which themes may be of more interest than others, and therefore help to prioritize the data collection process. This may result in significant effort saved.

In order to benefit from future users' cooperation, it would be beneficial to start by clarifying the model's scope, purpose, and functionalities. A model presentation based on a functional model version for one of the test regions could initiate interest in deploying the model in a new region or country.

An interactive session during which the intended user could play around with the model could further increase the user´s understanding of the model. This will generally lead to more concrete ideas about the exact way a user may wish to use the model. As discussed above, this may be helpful in prioritizing which parts of the model to regionalize.

After this, iterative contact with the future user serves best in order to optimally align its regional/national functionality with the users' purpose.

3.4.3 Bottlenecks and potential solutions

Number of projects funded by OPs

Some Operational Programmes may only finance a very limited number of projects. E.g. the ERDF OP of Zuid-Holland consists of only 34 projects. Even though all of these projects were included in the sample to link SICs to theme elements and calculate their relative financial contributions, such sample may not provide a very robust description of future programmes. Rather than working through the architecture in themes and theme elements, it may be better in such case to work directly at the SIC scale. Implementing this would imply to define one theme element in the model for each single SIC. This would allow a model user to directly define financial contributions to each SIC.

Delay for data collection

OPs related data collection can be a very slow and long process. For some of the test regions, it required 3 months, where a couple of weeks were anticipated. This may be viewed as a bottleneck for the spreading and implementation of the model in other European regions and countries.

3.5 Estimated workload

The workload to deploy the model to a new region or country is highly dependent on the operational detail required and the capacities of the personnel involved. The table below gives a rough estimation of time requirement for the model developer.

Data collection	30-60 days
SIC sample selection and distribution calculation	5-20 days
Adaptation of model parameters	2-10 days
Training of intended users	2-4 days' workshop + preparation

Table 11: Estimation of time requirement for the model developer

3.6 Further model development

The generic model and its five applications have been developed as a testing model. The first purpose of the process was to prove the principle of being able to assess ex-ante the carbon impact of different Operational Programmes. Secondarily, the process of developing the model served the purpose of clarifying the user requirements for the model so that it would be valued useful by the authorities that are its intended user group.

Although great care has been taken in developing the model to make it as functional as possible, the current model still should still be considered a beta version. The version is intended to be further developed while expanding its use to other EU regions and countries. Further development could still lead to various improvements of model functionality.

Further development could be in:
• Extending the presentation of results.
• Additional indicators.
• Further detailing existing SICs.
• Adding additional SICs.
• Visual improvements.
• User interface improvements.
• Monitoring functionalities.

Any additional functionality to make the model more user-friendly and connect to further user needs should be developed when the model is picked up and such needs arise.

Annex 1: Technical description of SICs

Annex 1 provides technical details of the algorithms and parameters used in each of the SICs. For reasons of volume, this annex is provided as a separate document.

CO2MPARE project team

Michiel Hekkenberg
Arjan Plomp
Bronia Jablonska
Marc Londo

Johan Schram
Charlotte Vincent-Genod
Steven le Pierrès
Antoine Saglio

Andrea Forni
Oscar Amerighi
Cecilia Camporeale
Roberto Del Ciello
Ivano Olivetti
Pasquale Regina
Maria Velardi

Ilkka Keppo

Michael ten Donkelaar
Jan Harnych
Jiri Spitz
Vladimira Henelova

Stavroula Papagianni