Database on farm-level production and sustainability indices for assessing sustainable diets

Deliverable No. D4.7

SUSFANS DELIVERABLES

Christian Götz (UBO), Andrea Zimmermann (UBO), Adrian Leip (JRC)

Hannah van Zanten (WU), Sara Hornborg (SP), Friederike Ziegler (SP)

Version	Release date	Changed	Status	Distribution
V2	31/3/2017	--	Final	--

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 633692

SUSFANS Deliverable document information

Project name	SUSFANS
Project title:	Metrics, Models and Foresight for European SUStainable Food And Nutrition Security
Project no	633692
Start date:	April 2015
Report:	D4.7
Work package	WP4
WP title (acronym):	Drivers and data: primary agricultural and fisheries production
WP leader:	UBO, Andrea Zimmermann; JRC, Adrian Leip
Period, year:	2, 2017
Responsible Authors:	Christian Götz, Andrea Zimmermann, Adrian Leip
Participant acronyms:	UBO, JRC
Dissemination level:	Public
Version	V1
Release Date	30/3/2017
Planned delivery date:	30/03/2017
Status	Final
Distribution	Public

Dissemination level of this report

Public

ACKNOWLEDGMENT & DISCLAIMER

This project has received funding from the European Union’s Horizon 2020 research and innovation program under grant agreement No 633692. Neither the European Commission nor any person acting on behalf of the Commission is responsible for how the following information is used. The views expressed in this publication are the sole responsibility of the author and do not necessarily reflect the views of the European Commission.

Reproduction and translation for non-commercial purposes are authorised, provided the source is acknowledged and the publisher is given prior notice and sent a copy.

TABLE OF CONTENT

DELIVERABLE SHORT SUMMARY FOR USE IN MEDIA

The aim of this paper is to demonstrate the design, computation and application of the SUSFANS performance metrics and indicators. Experimental results are produced for the European Union and the project's four case study countries Czech Republic, Denmark, France and Italy. The results focus on environmental impacts of the food system and are intended to serve as a basis for the discussion and evaluation of the metrics and indicators in their further development and application. The procedure for designing the indicators and the decisions on how to use variables and data in order to calculate the performance metrics has a critical impact on the results that they deliver. By those decisions all assessments based on the indicators are centrally influenced.

The major question is if the developed metrics and indicators are functional and useful to assess the European agrifood sector's influence on the environment, namely (A) its impact on climate through the emission of climate relevant gases, (B) on ecosystems by the application of fertilizers and toxic substances in agricultural production, and (C) on biodiversity. In more detail the questions are (1) if it is possible to assess developments in the agricultural and food sector in terms of improvements or deteriorations for the environment, (2) what conditions and limitations need to be taken into account while using the performance metrics and indicators, and (3) which issues need to be taken care of in their further development and applications. Those issues are essential in the attempt to make the metrics significant and meaningful for their application in the assessment of policy measures and innovations that are intended to achieve sustainable food and nutrition security in the European Union.

The analysis is based on the Common Agricultural Policy Regionalised Impact Modelling System (CAPRI). The CAPRI system is developed for policy and market impact assessments at a global, regional and farm-type level and is focused on the European market. The production of metrics in this paper is based on its consolidated database and modelling output for the years 2010 and 2013.

The computation of metrics requires a definition of policy goals or visions. They are closely connected to the different stakeholders' interests and endeavour in aligning nutrition and food production in the European Union to foster public health aspects, environmental sustainability of production, and a prosperous food business sector. For the experimental production of metrics in this paper the policy targets are more oriented at a visionary ideal status in the future, for example zero greenhouse gas emissions from agricultural production in the year 2100. The policy target value needed for the computation of indicators is then derived from the linear trajectory pathway between the year 2010 and 2100.

The results show that the performance metrics and indicators may be effectively used to assess the impacts of agricultural production on the environment. They allow for qualitative assessments in terms of improved or worsened impacts and provide quantitative information about the magnitude of this change. Policy measures of potential innovations may be judged in a way that they lead to an improvement or deterioration of the situation in one or more of the environmental aspects by a certain percentage compared to a reference year.

The analysis also points at issues that are important to solve in the further process. Policy goals need to be translated into the target values used for the production of indicators. This should ideally be done in a consistent manner to allow for comparisons of developments across regions or countries and across metrics. If performance metrics comprise different indicators the way of aggregating them plays a crucial role for the result that the metrics deliver. The performance metric "Clean air and water" for example is an aggregate of four individual variables, namely (1) Reduction of nitrate air pollution, (2) Reduction of nitrate soil pollution, (3) Reduction of phosphate pollution, and (4) Reduction of toxic substances use. The performance metric might indicate a positive development even though some of the underlying indicators show a worsening of the situation for their subdomain. From this it follows that first, their weight in the aggregation plays an important role for the outcome and asks for special attention, and second, using the information from the underlying indicators allows for more differentiated assessments on the whole performance domain than sticking to the metric alone.

The metrics and indicators can be used to evaluate the change over time for a region and certain environmental aspect. However, the indicator does not say anything about the status quo or the level of impairment to the environment in the region. A region might show a stronger improvement concerning nitrate emissions compared to other regions even though its nitrate emissions are still at the highest level among all of them. This reveals that regions' geographic conditions and their differences in the status quo are important for the assessment. A precise judgment requires complementary information and comparisons across regions need to incorporate information on the local conditions as well. As the policy targets were set in an experimental way the results cannot be used to derive policy recommendations at this stage.

TEASER FOR SOCIAL MEDIA

This paper demonstrates that SUSFANS metrics for assessing the environmental sustainability of the European food system can be effectively produced and applied to assess policy measures and potential innovations that aim at achieving sustainable food and nutrition security in the European Union. The analysis points at important issues that need to be taken into account for the further development and application of metrics.

Twitter

First SUSFANS metrics produced and applied to assess the environmental impacts of the European food system. Results point at issues to be solved for their further usage.

ABSTRACT

The paper demonstrates a first attempt of producing and applying the SUSFANS performance metrics. Their production is intended to reveal weaknesses and issues that need to be taken into account in their further development and usage. The discussion of metrics and indicators and the respective results are limited to the domain “Reduced environmental impacts of the food system” as their development is largely finalized. Results are based on the database and modelling output of the CAPRI system and focus on the European Union and the SUSFANS project’s four case study countries Czech Republic, Denmark, France, and Italy both at their country and their NUTS2 level. The results are experimental since their application requires a definition of policy goals and visions and is dependent on the purpose of their usage and the users’ requirements. The results demonstrate that the indicators and metrics are straightforward to implement and easy to interpret. They allow for an assessment of improvements compared to a reference situation and the defined policy target or vision. This makes it possible to compare developments in different domains of performance metrics or aggregate indicators and over Member States or groups of countries. Concerning WP5 this allows for identifying innovation pathways by assessing the impact of prospective innovations. The discussion of indicators and results reveals several essential issues that need to be taken into account for their further development and usage. The design of indicators and their weighing in the aggregation to metrics have an essential impact on the metrics’ significance and their quantitative outcome. The procedure for translating policy visions or goals into target values for indicators is essential and requires standardization to allow for comparability across countries and indicators. Evaluations based on metrics require complementary information and should incorporate information on the geographic conditions as well.

1 INTRODUCTION

The deliverable was meant to provide a database on farm-level production and sustainability indices for assessing sustainable diets ex-post. As the SUSFANS metrics are still under development and to ensure timely delivery, it will provide a database on those indicators that can already be calculated. The full set of metrics will be provided once this has been finally decided on.

The deliverable is of high importance to the whole project as it is the first attempt to actually apply the SUSFANS metrics. Since these metrics are still under development, the deliverable can only provide examples of metrics applications. It is therefore used in an explorative way, i.e. (1) to demonstrate the way to get to the metrics in terms of example calculations and (2) to provide a discussion of issues in calculation and actually applying the SUSFANS metrics. More specifically, the deliverable informs WP1 for the metrics development, WP4 for primary production indicators, WP5 for the operationalization of metrics in the case studies and WP9 for the other models.

The exact aims of the deliverable are:

1. Experimental attempt to produce the SUSFANS metrics discussed so far as an intermediate step for further discussion and improvement
2. Show and share what can be produced by the means of the CAPRI modelling system
3. Develop procedures for calculating aggregated performance indicators from individual variables
4. Reveal major problems with SUSFANS metrics that need to be solved

2 THE CAPRI MODELLING SYSTEM

CAPRI is a comparative static partial equilibrium model for the agricultural sector developed for policy and market impact assessments from global to regional and farm type scale. The core of CAPRI is based on the linkage of a European-focused supply module and a global market module. The following brief description of the Common Agricultural Policy Regionalised Impact Modelling System (CAPRI) is based on the most recent CAPRI documentation (Britz and Witzke, 2014). All available information around the CAPRI modelling system can be found on the CAPRI webpage (www.capri-model.org).

The supply module consists of independent aggregate non-linear programming models which cover the EU27, Norway, Western Balkans and Turkey. They represent all agricultural production activities and related output generation and input use at regional (280 NUTS2, Nomenclature of Units for Territorial Statistics) or farm type level (Gocht and Britz, 2011). The programming models feature a hybrid approach consisting of the combination between a Leontief-technology for variable costs covering a low and high yield variant for the different production activities and a non-linear cost function which captures the effects of labour and capital on farmers' decisions. The non-linear cost function allows for perfect calibration of the models and a smooth simulation response rooted in observed behaviour. Each programming model (at NUTS2 or farm type level) optimises income under restrictions relating to land balances, including a land supply curve, nutrient balances and nutrient requirements of animals and, if applicable, policy obligations. Decision variables are crop areas and total land use, herd sizes, fertiliser application rates and the feed mix. With respect to policy implementation, the different policy instruments of Pillar I and Pillar II of the Common Agricultural Policy (CAP) are depicted in detail for the EU. Prices are exogenous to the supply module and provided by the market module.

The global market module is a spatial, non-stochastic global multi-commodity model for about 50 primary and processed agricultural products, covering about 80 countries/country blocks, which are organised in 40 trading blocks. The market module is defined by a system of behavioural equations representing agricultural supply, human and feed consumption, multilateral trade relations, feed energy and land as inputs and the processing industry; all differentiated by commodity and geographical units. Land is not explicitly allocated to activities when the model is solving. But the land demand elasticities in the system imply certain yield elasticities that may be used to disaggregate the total supply response into contributions from yields and from areas and to estimate the land allocation in scenarios, starting from the baseline land allocation. On the demand side the Armington approach (Armington, 1969), assumes that the products are differentiated by origin, allowing the simulation of bilateral trade flows and of related bilateral and multilateral trade instruments, including tariff-rate quotas. This sub-module delivers the output prices used in the supply module and allows for market analysis at global, EU and national scale, including a welfare analysis.

The main databases used in CAPRI are based on EUROSTAT, FAOSTAT, OECD and extractions from the Farm Accounting Data Network (FADN). The supply response of each NUTS2 or farm type in the European-focused supply module is estimated using time series data on land use and corresponding price and cost developments (Jansson and Heckelei, 2011). The parameters of the global market model are synthetic, i.e. to a large extent taken from the literature and other modelling systems.

Typically, CAPRI is used for simulations starting from a given baseline. For medium-term horizons, the price-quantity structure of the model is calibrated to commodity market outlook of the European Commission. To produce longer-term baselines, trend estimations constrained to account for technology restrictions and external prior information become more important. Some information supporting this process comes, for example, from the GLOBIOM model. This baseline alignment across models ensures a relatively similar starting point for subsequent scenario analyses.

A complete list of products considered in CAPRI is given in Table 1. Main outcome indicators are listed in Table 2.

Table 1 Outputs, inputs, income indicators, policy variables and processed products in the data base

Group	**Item**	**Code**
Outputs		
Cereals	Soft wheat Durum wheat Rye and Meslin Barley Oats Paddy rice Maize	SWHE DWHE RYEM BARL OATS PARI MAIZ

Group	**Item**	**Code**
	Other cereals	OCER
Oilseeds	Rape Sunflower Soya Olives for oil Other oilseeds	RAPE SUNF SOYA OLIV OOIL
Other annual crops	Pulses Potatoes Sugar beet Flax and hemp Tobacco Other industrial crops	PULS POTA SUGB TEXT TOBA OIND
Vegetables Fruits Other perennials	Tomatoes Other vegetables Apples, pear & peaches Citrus fruits Other fruits Table grapes Table olives Table wine Nurseries Flowers Other marketable crops	TOMA OVEG APPL CITR OFRU TAGR TABO TWIN NURS FLOW OCRO
Fodder	Gras Fodder maize Other fodder from arable land Fodder root crops Straw	GRAS MAIF OFAR ROOF STRA
Marketable products from animal product	Milk from cows Beef Pork meat Sheep and goat meat	COMI BEEF PORK SGMT

Group	Item	Code
	Sheep and goat milk Poultry meat Other marketable animal products	SGMI POUM OANI
Intermediate products from animal production	Milk from cows for feeding Milk from sheep and goat cows for feeding Young cows Young bulls Young heifers Young male calves Young female calves Piglets Lambs Chicken Nitrogen from manure Phosphate from manure Potassium from manure	COMF SGMF YCOW YBUL YHEI YCAM YCAF YPIG YLAM YCHI MANN MANP MANK
Other Output from EAA	Renting of milk quota Agricultural services	RQUO SERO
Inputs		
Mineral and organic fertiliser Seed and plant protection	Nitrogen fertiliser Phosphate fertiliser Potassium fertiliser Calcium fertiliser Seed Plant protection	NITF PHOF POTF CAOF SEED PLAP
Feedings tuff	Feed cereals Feed rich protein Feed rich energy Feed based on milk products Gras Fodder maize Other Feed from arable land	FCER FPRO FENE FMIL FGRA FMAI FOFA

Group	Item	Code
	Fodder root crops Feed other Straw	FROO FOTH FSTRA
Young animal Other animal specific inputs	Young cow Young bull Young heifer Young male calf Young female calf Piglet Lamb Chicken Pharmaceutical inputs	ICOW IBUL IHEI ICAM ICAF IPIG ILAM ICHI IPHA
General inputs	Maintenance machinery Maintenance buildings Electricity Heating gas and oil Fuels Lubricants Water Agricultural services input Other inputs	REPM REPB ELEC EGAS EFUL ELUB WATR SERI INPO
Income indicators	Production value Total input costs Gross value added at producer prices Gross value added at basic prices Gross value added at market prices plus CAP premiums	TOOU TOIN GVAP GVAB MGVA
Activity level	Cropped area, slaughtered heads or herd size	LEVL
Policy variables Relating to activities	Premium ceiling Historic yield Premium per ton historic	PRMC HSTY PRET

Group	Item	Code
	yield Set-aside rate Premium declared below base area/herd Premium effectively paid Premium amount in regulation Type of premium application Factor converting PRMR into PRMD Ceiling cut factor	SETR PRMD PRME PRMR APPTYPE APPFACT CEILCUT
Processed products	Rice milled Molasse Starch Sugar Rape seed oil Sunflower seed oil Soya oil Olive oil Other oil Rape seed cake Sunflower seed cake Soya cake Olive cakes Other cakes Gluten feed from ethanol production Biodiesel Bioethanol Palm oil Butter Skimmed milk powder Cheese Fresh milk products Creams Concentrated milk	RICE MOLA STAR SUGA RAPO SUNO SOYO OLIO OTHO RAPC SUNC SOYC OLIC OTHC GLUE BIOD BIOE PLMO BUTT SMIP CHES FRMI CREM COCM WMIO

Group	**Item**	**Code**
	Whole milk powder Whey powder Casein and caseinates Feed rich protein imports or byproducts Feed rich energy imports or byproducts	WHEP CASE FPRI FENI

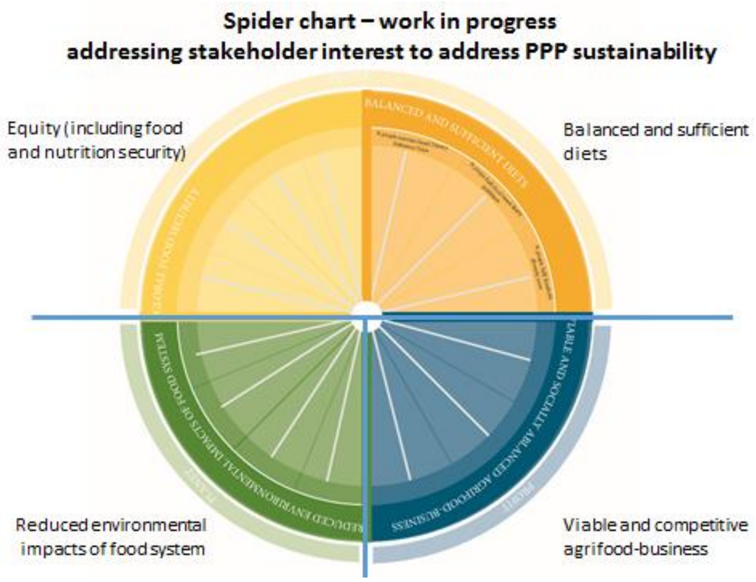
Table 2 Main outcome indicators

Module	Outcome indicator
European supply module	Supply Feed demand (in bulks) Areas Herds Yields Input demand Gross value added
Global trade module	Supply Final demand Feed Processing Prices (consumer/producer) Trade flows
Post model processing	Support of results analysis Aggregation over scales/products/activities Decomposition: yield response, behavioural functions market model Economics Farm income indicators Welfare analysis CAP budget, CAP instruments Environmental indicators Gaseous emissions N, P, K balances GHG inventories

Module	Outcome indicator
	Energy use in European agriculture Diet related Calories Macronutrients Spatial downscaling

3 SUSFANS METRICS

3.1 Spider diagram



3.2 From individual variables to SUSFANS performance indicators – Metric Hierarchy

Societal goals[1] are the overarching societal challenges

Each policy goal is composed of various *'areas of concern'* where society wants to improve the situation. This is measured with ***performance metrics***, which indicate how far society has come in one point in time for reaching the desired endpoint of development against a reference point in time. Performance metrics should be named 'positively' thus referring to the situation where the policy goals are met.

[1] Note the difference to D1.2 – the term 'societal goal' is used where we formerly used 'policy goal'. This has two reasons: (i) we need another term for the 'vision' which is quantified at the level of the aggregated variables/indicators, and (ii) the four big concerns are really societal concerns, while policy tackles smaller issues: nitrate directive, NEC directive etc. thus policy goals would be final objective where policy targets set concrete objectives for a certain point in time. But the thematic level of both is the same.

The performance metrics themselves are composed of one or more **policy goals**. Policy goals are linked to a certain time frame and link to measureable data.

Aggregate variable*[2]*s V_g combine one or various derived variables to the level of the policy goals g. It is measured (or transformed) to the same unit as the policy targets which quantify the policy goals. **Policy targets** V_g^t for a policy goal are linked to a certain point in time t_1 and might be different for different countries or at EU level. Simple case of a policy target is the level of GHG emissions for a country in a target year t_1. Both the V_g and T_g for the target year t_1 (predicted by models) are compared to the situation of the aggregate variable V_g^R in the reference period t_0 (e.g. present).

Aggregate indicators I_g evaluate aggregate variables with respect how much of the path to go from the reference V_g^R to the level of desired level of the policy goal V_g^G is already achieved. The name of the aggregate indicators is usually a 'reduction of a gap to optimum or an undesired fact (e.g. emissions)'

$$I_g = \frac{V_g - V_g^R}{V_g^G - V_g^R}$$

The aggregate indicator can also be calculated for the policy target:

$$I_g^t = \frac{V_g^t - V_g^R}{V_g^G - V_g^R}$$

I_g can assume values between zero and one if there is an 'improvement' towards reaching the policy goal, but it can also be negative if the situation is worsened

I_g^t can assume values between zero and one. Thereby, the higher I_g^t the more ambitious are the policy target. It could be interpreted such that in such case the policy goal is judged to be more urgent as compared to policy target with a lower I_g^t. Only in very rare cases it is possible that I_g^t assumes values >1, for instance if a world with zero emissions is ideal, but a world with 'negative' emissions is thinkable.

Performance metrics aggregate indicators into meaningful number that shows how well the 'scenario' performed for each of the dimensions defined for the

[2] Aggregate variables are 'new' and have not been introduced before. Basically the definition of 'aggregate indicator' is split into two steps: first aggregation to aggregate variables and then evaluation against objectives to aggregate indicator.

overarching societal goals. To do the aggregation, weighting factors *w* must be defined for each policy goal within one of the dimensions (=> performance metrics). Those weighting factors usually are 1 unless the predominance of one policy goal over the others can be justified. Weighting factors can be different from one to consider correlations between policy goals. For instance, the policy area 'nutrient surplus' is overlapping with the policy area 'air and water pollution' and 'GHG emissions' and thus has a weighting factor of zero to avoid double counting.

$$M = \frac{\sum_g \{I_g \cdot w_g\}}{\sum_g \{w_g\}}$$

In analogy, the policy targets can be aggregated to the same dimensions:

$$M^t = \frac{\sum_g \{I_g^t \cdot w_g\}}{\sum_g \{w_g\}}$$

Aggregation of the performance metrics to the societal goals can be done in many ways and it `could be the use of the target metrics M^t as a proxy for importance, based on the following reasoning:

The more important a dimension of a societal goal is considered, the higher the level of ambition is sought for setting the targets, thus the targets are closer to the policy goal (vision) and the higher the score of M^t.

Thus one option to calculate the overall score for the societal goal could be summing over all performance metrics *m*:

$$S = \frac{\sum_m \{M^t \cdot M\}}{\sum_m \{M^t\}}$$

Spider diagram

The spider diagram is made of

- an inner circle indicating the reference situation (e.g. current situation) with the values 0
- an outer circle indicting the ideal situation (vision, paradise) with the values 1
- a 'dot' for the performance of a scenario (projection, forecast or other) M. This 'dot' can be within the inner circle if the situation is projected to worsen, between inner and outer circle (in most cases), indicating

progress towards the policy goal, whereby the closer to the outer circle the ‘dot’ is located the better is the performance of the scenario.

- a ‘star’ for the level of ambition of the policy target M^t.
- Within each ‘quarter’ of a societal goal, the segment for each performance metrics can be varied to indicate the urgency of the different dimensions. One could distinguish
    - Urgency in the reference situation: the higher the level of ambition the more urgent the dimension is regarded. The area A_m of the segment m is thus proportional, for example, to the share of all target metrics n:

$$\frac{A_m}{A} = \frac{M_m^t}{\sum_n M_n^t}$$

    - Urgency in the predicted situation: the higher the residual ‘gap’ between scenario performance metrics and target metrics the more urgent are additional efforts. The area A_m of the segment m is thus proportional, for example, to the share of the total ‘performance gaps’:

$$\frac{A_m}{A} = \frac{M_m - M_m^t}{\sum_n (M_n - M_n^t)}$$

- Colors could be used to visualize additional information
- Next to the spider diagram a ‘thermometer’ diagram can indicate the ‘heat’ of the societal goals and of the overall sustainability. Aggregation of societal goals to overall sustainability could in principle be done following the same principles laid out above, but are dependent on user’s preferences.
- Below the spider diagrams trend lines etc. could be displayed …

Data sources

V_g model prediction

V_g^t legislative texts

V_g^G ‘vision’ (status which is assessed optimal with respect to the policy goal)

w_g scientific considerations

Everything else is derived from it.

3.3 Metrics stratification

Since the selection and definition process of indicators is not yet finalized for the domains (1) Equity, (2) Balanced and sufficient diets, and (3) Viable and competitive agrifood-business, Chapter 3.3 treats only indicators and performance metrics of the domain (4) Reduced environmental impacts of the food system that could be produced based on the CAPRI modelling system.

3.3.1 Reduced environmental impacts of the food system

The societal goal comprises four areas of concern where performance metrics are designed and computed for, namely the areas of (1) Climate stabilization, (2) Clean air and water, (3) Biodiversity conservation, and (4) Preservation of natural resources. The corresponding aggregate indicators are described in the following together with their constraints related to their design or underlying data and measurement.

3.3.1.1 Climate stabilization

The societal goal of climate stabilization is measured by the indicator "Reduction of total GHG emission caused by the agri-food chain". Total GHG emissions are measured as the global warming potential of climate relevant gases caused by agricultural production activities. This comprises emissions from both the cropping and animal sector. Three climate relevant gases – carbon dioxide (CO2), methane (NH4), and nitric oxide (N20) – are aggregated based on their specific global warming potential and measured as CO2 equivalent. The CO2 equivalent is the amount of a combination of those gases that would result in the same global warming effect as the corresponding amount of CO2 alone.

The societal goal of climate stabilization is politically reflected in various goals where most of them either refer to a global warming target or to concentration targets for the climate relevant gases. Both are still controversial issues at the international level and in an ongoing discussion. The European Union has a two degrees Celsius target on its agenda and officially declared that a concentration target of 400ppm CO2 equivalent issued by the 350.org campaign and discussed at the Tällberg forum is in line with this agenda.[3] However, both warming and concentration targets can be vague and confusing.

Concentration pathways are usually formulated as relative cuts compared to a reference period. However, such vaguely formulated targets may be met in different ways. The decline may be linear or there might be a peak first and a sharp decline afterwards. The amount of climate relevant gases emitted to the atmosphere resulting from the chosen pathway might differ tremendously and thereby also its effect on the climate. Less vague than a relative target is an emission trajectory precisely defining how much greenhouse gas emission is allowed for each year going forward. At the same time the emission target needs to be measurable and controllable.

The European Union's emissions trading system (EU ETS, EC 2003) may serve as a starting point and prototype for an emission trajectory. It defines the overall volume of greenhouse gases that are allowed to be emitted by different industries. The overall emission is limited by a ceiling on the number of emission allowances. Within these Europe-wide caps, companies receive or buy emission allowances which they can trade as needed. The ceiling is cut linearly by a reduction amounting to 1.74% of the average total quantity of allowances issued annually over the period 2008-2012.

Applied to the computation of the indicator the GHG emissions from the agricultural sector of the European Union in the reference year 2010 are taken to calculate the linear reduction amount. By applying the same relative cut of 1.74% this amounts to a yearly reduction of 6954 tons of CO2 equivalent for the whole European agricultural sector where emissions in the reference year were at 399670 tons of CO2 equivalent. By following this trajectory path emissions would be around 20% lower in 2021 compared to the reference year and the idealistic point of zero emissions would be reached in 2068.[4]

3.3.1.2 Clean air and water

The performance domain "Clean air and water" comprises five aggregate indicators, namely (1) Reduction of N surplus, (2) Reduction of N emissions to the atmosphere (air pollution), (3) Reduction of N emissions to the hydrosphere (water pollution), (4) Reduction of P surplus, and (5) Reduction of Toxic

[3] See for example: http://www.eea.europa.eu/highlights/climate-change-targets-350-ppm-and-the-eu-2-degree-target, 2017-03-01.

[4] The new Effort Sharing Regulation 2021-2030 proposal of the European Union includes sectors that were so far not covered by the ETS regulation, among them agriculture as well. See for example https://ec.europa.eu/clima/policies/effort/proposal_en, 2017-03-01.

substances use. The indicators (1) to (4) are described together in the following section since their related issues are overlapping.

3.3.1.2.1 Nitrate and phosphate leaching and emission

The computation of the aggregate indicators (1) to (4) is based on the total amount of leaching and emission of nitrate and phosphate across all agricultural activities in a region. This reflects the crucial role that the respective chemical compounds play in the nutrient cycle for natural ecosystems. Depending on the farming activities and intensity the related emissions and leaching may critically disturb the equilibrium of the life cycles and whole ecosystems. They are the cause of ground water pollution, eutrophication and acidification of the soil and water bodies.

At the European level this issue is addressed in the European Union's Water Framework Directive (EU 2000), the Groundwater Directive (EU 2006), and the Nitrates Directive (EC 1991). The implementation of the Directives at the Member State level comprises various measures and mechanism like intense monitoring of nitrate levels, regular reporting on the implementation by the European Commission, the establishment of codes of good practice for farmers, designation of nitrate-vulnerable zones, compulsory action programmes for farmers located in vulnerable zones, and limits for the application of nitrogen from manure. The Groundwater Directive defines a limit for nitrate concentrations and Member States may set their own tighter limits.

The policy vision for the operationalization of the aggregate indicators might be zero emission and leaching in the far ideal future. However, it is more difficult to define a target value for the policy goal. The impact of pollution through nitrate and phosphate is critically dependent on the geographic conditions in regions, including climate, soil conditions and inclination, presence of surface waters and natural reserves. This is the reason why the European Directives on the issue involve the creation of vulnerable zones that are considered to require specific attention and stricter regulations.

In view of the intricacy expressed above the policy target is oriented at an ideal state of zero emissions in the far future and pollution targets are interpolated over the chosen time frame. The experimental results are based on the policy target value of zero emissions and leaching in the year 2100 and the policy target V_g^G for 2013 is based on a linear interpolation over the time span from the reference year 2010 to the year 2100.

Since (1) is overlapping with (2) and (3), (1) Reduction of N surplus is assigned zero weight in the aggregation for the performance metric to avoid a double counting of the nitrate emission part. The remaining aggregate indicators are assigned a weighting factor of one under the assumption that they are all equally important for the assessment of the domain. This seems appropriate as long as there is no information available that nitrate or phosphate or the leaching category of nitrate is more harmful compared to the other so that it should receive a higher weight in the aggregation.

3.3.1.2.2 Reduction of Toxic substances use

The aggregate indicator is derived from the individual variable on the usage of substances in the scope of plant protection and crop growth regulatory measures. It is, first, not differentiated between the different measures and agents applied, and second, there is no information available on both the efficiency and harmfulness of the different agents used. All used crop protection measures comprising the application of herbicides, fungicides, insecticides, as well as growth regulatory measures are aggregated on a monetary basis with data stemming from the Economic Accounts of Agriculture provided by EUROSTAT. The data is the sum of expenditures for all non-mechanical plant protection measures and is corrected by the inflation rate to account for price changes between the reference and the target year.

There are two restrictions that need to be taken into account when using the indicator. First, there is no differentiation of the expenditure over the shares of used agents and their harmfulness for nature. The indicator might show an improvement of the situation in terms of less expenditure for toxic agents even though the overall toxicity of a changed composition of used agents over the whole sector has increased. Second, the overall use of toxic agents is undoubtedly related to the variety and intensity of cropping activities and thereby also to a region’s natural conditions for high intensity farming. The geographic conditions in connection with the established farming practices and crops may have an influence on the reduction potential.

The second limitation makes it difficult to define a political target as the target value ought to vary over regions. As a consequence, it is oriented at the policy vision of an ideal future state of zero pollution from toxic substances use. The year 2100 is chosen for the ideal state and the pollution target for the year 2013 is based on linear interpolation between the reference year and 2100.

3.3.1.3 Biodiversity conservation

3.3.1.3.1 Reduction of the contribution of the agrifood chain to loss of MSA

The production of indicators under this subdomain is currently not feasible with the CAPRI modelling system.

3.3.1.3.2 Agricultural land use diversity

Shannon's entropy index is used for the assessment of agriculture's contribution to countries' or regions' biological diversity. The index is computed based on the agricultural sector's land use variety, i.e. more different land use types[5] and crop varieties and their more uniform distribution over a region's total agricultural area is supposed to be associated with a higher biological diversity in the sector. Changes in the land use variety is likely to have an impact on the variety of species in the agricultural area and thereby also in a region's total area. It may thereby serve as a proxy for the agricultural sector's influence on biodiversity conservation in a region.

The index is computed as

$$H_r = -\sum_{i=1}^{N} p_i ln p_i \,,$$

where p_i is the share of land use type i in region r's total arable land and N is the number of all land use types in the region. H_r increases with the number of different land use types N and if their shares in the total agricultural area p_i are more equal. The index is zero if there is only one land use type p_i in the sector.

The land use diversity index has several constraints that need to be taken into account when using it. First, the share of a region's agricultural area in its total agricultural area is not reflected by the entropy index. In a region where the agricultural sector covers a large part of its total area, changes in its land use structure is likely to show a stronger impact on the overall biological diversity than in a region where the agricultural sector has a rather modest share in the total area.

Second, the growth of agricultural area over time is not taken into account. In regions where the non-agricultural area is rather natural and untouched[6] it is likely the case that a growing agricultural area would lead to a decrease in the

[5] The term "land use type" in this context is used in the meaning of cropping activity like soft wheat, durum wheat, barley, etc. An overview of the different activities used to differentiate land use for computing the index is given in table 1 and comprises the activity groups cereals, oil seeds, other annual crops, fruits, vegetables and other perennial crops, and fodder activities.

[6] Finland, Norway and parts of Sweden might be examples for such a scenario.

overall biodiversity. However, the index might show an improvement nonetheless if just the number of different cropping activities and the uniformity of their shares increased at the same time. From this perspective it seems advantageous to complement the entropy index with a measure on the overall share of agriculture in a region's total area.

Third, a growing number of different crop activities in a region is not necessarily associated with higher biodiversity. The richness of species is likely to vary across different cropping activities, i.e. some crops might be associated with a higher biological diversity than others. In regions where low intensity grazing and pastoral activities as well as low intensity farming is predominant[7], an increase of the number and share of higher intensity cropping activities might likely result in a decreased biological diversity for the whole region even though the index would indicate an improvement. In general, the farming intensity shows an important impact on the biological diversity of cropping activities which cannot be taken into account by the index.

The critical issue is the definition of both the political target as well as the vision. At the European level there is the Union's Commitment taken within the International Convention on Biodiversity (CBD) in 2010 which was reaffirmed and expanded in 2011.[8] The strategy defines targets and actions to halt the loss of biodiversity in the EU by 2020. Furthermore, there is evidence on the negative impact that different aspects of intensified farming show on biological diversity (Benton et al. 2002, Donald et al. 2001, Reidsmaa et al. 2006, Petit et al. 2001). However, the question arises what the ideal state of land use entropy is supposed to be in this context. In order to make the indicator usable the target value has to be increased compared to the reference value by a certain percentage. The target value is here arbitrarily set to be 20% greater than the reference value. This corresponds to a vision that land use diversity should increase by 20% compared to the reference year.

3.3.1.3.3 Marine biological diversity

Seafood from capture fisheries represents the only large-scale food production based on a wild resource. It is sometimes argued that fisheries is be a good alternative to produce food with less impacts and resource use than many land-

[7] Examples for such a case might be parts of Scotland, Ireland for the most part and the mountainous regions of Austria.
[8] Information on the European Union's strategy concerning issues of biological diversity and its commitment within the CBD is online available at: http://ec.europa.eu/environment/nature/biodiversity/strategy/index_en.htm#stra, 2017-03-10.

based protein production systems, as fisheries do not require inputs like feeds, fertilizers or pesticides. However, there are limits to natural production, and many stocks are overexploited and thus produce less than optimal. Direct and indirect ecosystem effects from over-exploitation include feedback such as altered ecosystem functioning (Howarth et al. 2014). This is manifested in the form of depletion of predatory fish (Christensen et al. 2003), collapse of major fish stocks (Pinsky et al. 2011), altered seafloor structure and function (Tillin et al. 2006) and biodiversity loss of target and non-target species (Dulvy et al. 2003; Lewison et al. 2004). From an ecosystem production perspective, it has been estimated that current global fisheries exceed levels of sustainable exploitation, and have to decrease considerably to avoid risk of impaired function (Coll et al. 2008, Watson et al. 2014); the full effects of fisheries on marine ecosystems are still largely unknown.

Environmental pressures from aquaculture include: some species and farming practices require high level of feed input based on capture fisheries and may release invasive species, cause eutrophication, conversion of ecologically sensitive coastal land, and transmit diseases to wild fish (Diana 2009).

Indicators to address marine biological diversity in models are difficult, and seafood modules are still under development for CAPRI. At present, including marine biological diversity is not feasible with CAPRI.

3.3.1.3.4 Red List Index (RLI)

The Red List Index (RLI) is a state indicator for trends in the extinction risk of terrestrial, freshwater and marine sets of species.[9] It is included in the CBD's list of indicators. At this point, it is not feasible to use it with the CAPRI modelling system.

3.3.1.4 Preservation of Natural Resources

The production of Indicators under this domain is currently not feasible with the CAPRI modelling system.

[9] Information on the index is online available at https://www.bipindicators.net/indicators/red-list-index, 2017-03-07.

4 EXPERIMENTAL RESULTS

Results are shown and discussed for the societal goal "Reduction of environmental impacts". The related performance metrics are (A) Climate stabilization, (B) Clean air and water, (C) Biodiversity conservation, and (D) Preservation of natural resources.[10] The outcome discussed in this chapter is in the experimental stage since, first, revealed weaknesses in the design of the indicators and data availability issues might call for amendments, and second, the used methods for deriving policy targets is merely experimental. Elaborated and agreed-on methods for deriving the policy target values of the indicators is the essential task in making the performance metrics usable and meaningful.

Results are shown for the European Union at the aggregate, and for the four case study countries Czech Republic, Denmark, France and Italy both at the Member State and at the NUTS2 level.

The policy target values are calculated for the different aggregate indicators as proposed in chapter 3. For the aggregate indicator "CO2 equivalent" the policy target value is based on the yearly cut on emission allowances of the European Emission Trading System (ETS). The yearly cut is a constant amount over time and corresponds to 1.74% of average yearly emission allowances over the time span 2008-2012. We oriented the yearly reduction amount at the value of CO2 emission equivalents in 2010 provided by the model output. The resulting trajectory path is linear with a constant reduction amount of around 0.017% and zero emissions in the year 2068. This results in a policy target value V_g^G that is around 0.052% lower in the measurement year 2013 compared to the reference year of 2010.

A corresponding procedure was used for defining the policy target values for the aggregate indicators under the performance domain (B) Clean air and water, namely (1) Reduction of N surplus, (2) Reduction of N emissions to the atmosphere (air pollution), (3) Reduction of N emissions to the hydrosphere (water pollution), (4) Reduction of P surplus, and (5) Reduction of toxic substances use. They have a linear trajectory as well leading towards the policy

[10] The performance metric (D) Preservation of natural resources is not included and discussed. The computation of its underlying aggregate indicators "Sustainable water use", "Overexploitation of wild-caught seafood resources" and "Maintenance of soil fertility" is currently not feasible with CAPRI. Also, (3) Biodiversity conservation comprises only one aggregate indicator since the discussed indicator "Reduction of the contribution of the agrifood chain to loss of MSA" is currently not feasible with the system as well.

vision of zero emissions in the year 2100. This results in policy target values V_g^G that are 0.033 percent lower than the 2010 reference value.

For the Shannon land use entropy index underlying the performance metric of the domain (C) Biodiversity conservation a growth factor of 20 percent was defined as policy vision. This means that the policy target value V_g^G is 20 percent higher compared to the reference year 2010.

Table 3 shows the outcome for the seven aggregate variables and related three performance metrics at the European level, i.e. for the EU28 aggregate.

Table 3: Results for aggregate indicators and performance metrics at the European level (EU28)

Aggregate indicator	(1) Reduction of N surplus	(2) Reduction of N air pollution	(3) N water pollution	(4) Reduction of P surplus	(5) Toxic substances
Indicator value	-0.33	1.28	-1.46	0.00	-3.23
Reference value	12010.67	3409.80	848.22	2423.37	11420973
Target value	11610.32	3296.14	819.94	2342.59	11040274
Measured value	12142.00	3264.87	889.38	2423.43	11956925
Min	3.37	1.09	0.22	-28.56	810.34
Max	2008.26	552.13	131.69	409.18	3265910.90
Mean	430.01	122.08	30.37	86.76	408177.59
SD	672.26	196.31	46.58	145.82	809006.15
Weight	0.00	1.00	1.00	1.00	1.00
Performance metric			(B) Clean air and water		
Metric value					-0.85

Aggregate indicator	(6) CO2 equivalent	(7) Shannon land use entropy
Indicator value	-0.60	-2.27
Reference value	399670.24	2.76
Target value	378807.46	3.31
Measured value	412284.32	2.73
Min	66.25	1.47
Max	67134.71	3.25
Mean	14029.91	2.79
SD	22577.31	2.34
Weight	1.00	1.00
Performance metric	(A) Climate stabilization	(C) Biodiversity conservation
Metric value	-0.60	-2.27

The first row shows the result for the aggregate indicator and the second row presents the policy target value for the European Union as computed based on the procedures explained above. The third rows shows the EU value of the measurement year 2013, and the fourth to the seventh row show minimal, maximal, and mean value of the policy target in European Member States.

Target and measured value and the statistics for indicators (1) to (4) are given in 1000 tons, for (5) the use of toxic substances in 1000 Euros, for (6) CO2 equivalent in tons and for the (7) Shannon land use entropy as index value. The eighth row presents the weighting factors that were applied in the aggregation of the indicators to the performance metric as explained in chapter 3.2. In the ninth and tenth row the performance metric and its value derived from the aggregate indicators above is given.

The sample statistics refer to the target value, i.e. the minimal target value for the use of toxic substances in the European Union is around 810.340 Euros and the highest target value is around 3.3 billion Euros for a Member State. On average European Union Members have a target value of around 408.2 million Euros for chemical plant protection and growth regulatory measures.

The result for (1) Reduction of N surplus is -0.33% indicating that the situation worsened. The deterioration is 33% compared to the distance to go from the reference situation in 2010 towards the policy target, i.e. the distance to go increased by 33%. The outcome for (2) Reduction of N air pollution is positive. The measured value in 2013 is 3264.87 and lower than the policy target value of 3296.14. The result of 1.28 for the indicator means that the distance from reference to target value is covered and that further improvement was made in terms of approaching the long term policy vision of zero emissions. The over-fulfilment accounts for 28% of the distance between reference and 2013 target value. The result for (3) Reduction of N water pollution is -1.46 signifying that the distance to the policy target is 146% higher in 2013, i.e. water pollution from nitrate is 146% worse compared to the reference situation. The (4) Reduction of P surplus is around zero indicating that the situation did not change significantly but that phosphate surplus is still at the level of the reference year. (5) Reduction of toxic substances is -3.23 indicating the situation worsened about 323%, i.e. the use of toxic substances is 323% higher compared to the reference situation.

(1) Reduction of N surplus is not used in the aggregation to the performance metric since it is overlapping with the indicators (2) and (3) and is assigned zero weight for the aggregation. The aggregation of indicators (2) to (5) results in

-0.85 for the performance metric (B) “Clean air and water”. This signifies that overall the situation worsened compared to the reference year. The deterioration is 85% which means that the distance to go to the desired political target is 85% greater compared to the reference year.

The performance metrics (A) Climate stabilization and (C) Biodiversity conservation consist of one aggregate indicator only, namely (6) CO2 equivalent and (7) Shannon land use entropy. Both show negative results at the European level. The emission of climate relevant gases from agriculture in terms of their global warming potential is around 9% higher than the policy target value. The outcome for (C) Biodiversity conservation is -2.27 indicating that the situation worsened around 227% compared to the reference situation.

The interpretation of a deterioration of 227% refers to the distance to go from the reference year towards the political target value of the indicator. The distance to the desired level is also called reference situation in this context. The change might thereby also be directly related to the reference situation but is not identical with the level change of the measured variable. In this example the deterioration of 227% compared to the reference situation corresponds to a decrease of land use diversity of around 1.1% compared to the reference year 2010.

Table 4 presents the results for performance metrics and aggregate indicators for Czech Republic, Denmark, France, and Italy at the Member state level. For all of them the situation worsened concerning the performance domain (C) Biodiversity conservation. The distance to the desired level increased by at least 233% for Italy and by at maximum 285% for Denmark. This corresponds to a decrease in land use entropy of around 1.5% for Italy and 1.9% for Denmark from reference year 2010 to measurement year 2013. The situation for the domain (A) Climate stabilization worsened for all of them except for Italy. The deterioration is at least 36% in case of France and at maximum 206% for Czech Republic corresponding to an increase of CO2 equivalent emissions of around 1.9% for France and of 10.7% for the Czech Republic. The total increase of carbon equivalent emissions amounts to around 1323 tons in France and 666 tons in Czech Republic.

Table 4: Performance metrics and aggregate indicators for Czech Republic, Denmark, France and Italy at the Member State level

		Performance metric			Aggregate indicator				
Region	Category	Climate stabilization	Clean air and water	Biodiversity conservation	N surplus total	N air pollution	N water pollution	P surplus total	Toxic substances
Czech Republic	Metric/Indicator value	-2.06	-17.11	-2.58	-3.04	1.27	-3.04	-56.07	-10.58
	Reference value	6204.20		2.49	269.44	59.62	36.37	3.83	237617.72
	Target value	5880.35		2.99	260.46	57.64	35.16	3.71	229697.13
	Measured value	6870.66		2.42	296.71	57.10	40.06	11.00	303787.08
Denmark	Metric/Indicator value	-0.51	2.54	-2.85	1.27	5.25	0.99	7.92	-4.02
	Reference value	11022.24		2.11	380.76	103.03	16.85	29.93	220791.07
	Target value	10446.88		2.53	368.07	99.60	16.29	28.93	213431.37
	Measured value	11314.69		2.15	364.70	84.99	16.30	22.03	236615.34
France	Metric/Indicator value	-0.36	-2.10	-2.52	-1.26	0.38	-0.72	-5.33	-2.74
	Reference value	70832.15		2.53	2077.51	571.17	136.23	239.81	3378528.60
	Target value	67134.71		3.03	2008.26	552.13	131.69	231.82	3265910.90
	Measured value	72155.68		2.48	2164.77	563.92	139.49	282.40	3485123.40
Italy	Metric/Indicator value	0.22	0.36	-2.33	-0.40	1.19	-0.51	2.70	-1.95
	Reference value	27577.12		2.71	962.38	345.06	65.37	303.67	824045.75
	Target value	26137.60		3.25	930.31	333.56	63.19	293.55	796577.56
	Measured value	27257.18		2.67	975.28	331.37	66.49	276.30	829529.12

In the domain (B) Clean air and water Czech Republic and France show a worsening of the situation of around 1711% and 210% respectively whereas the situation in Denmark and Italy improved of 254% and 36% respectively. The results for the underlying aggregate indicators are quite heterogeneous. The indicator (5) Reduction of toxic substances use shows a negative and the indicator (2) Reduction of N air pollution shows a positive outcome for all four Member States. The indicators (2) Reduction of N water pollution and (3) Reduction of total P surplus both have a negative result for Czech Republic and France and are both positive for Denmark.

The resulting performance metric (B) Clean air and water based on uniform weights for the four aggregate indicators is negative for Czech Republic and France and positive for Denmark and Italy. This demonstrates that the higher complexity of the performance domain in terms of different areas of pollution taken into account is reflected in the results at the aggregate indicator level. The overall outcome for the performance metric might conceal important developments at the more differentiated underlying indicators that might be contrary. Italy shows an improvement of the situation for the domain (B) Clean air and water where at the same time the situation in the underlying differentiated areas (3) Reduction of N water pollution and (5) Reduction of toxic substances use actually worsened. This may point at using the performance metrics carefully while taking the complementary information from underlying differentiated indicators into account.

Furthermore, the outcome for metrics and indicators does not allow for judging the status quo across countries. For example the result for the indicator (4) Reduction of total P surplus shows a strong deterioration of around 5607% in Czech Republic whereas it indicates an improvement of 792% in Denmark. However, both the level of total P surplus as well as its surplus per hectare is much lower in Czech Republic. The usable agricultural area of Denmark is just around 68% the value in Czech Republic. However, its total P surplus is 7.8 times greater in 2010. Total P surplus per hectare in Denmark is around 10.7 kg in the reference year whereas this level is at only around 0.93 kg in Czech Republic. A comparison between Italy and France reveals an analogous scenario. Whereas Italy shows an improvement of 270% in the (4) Reduction of total P surplus the result indicates a deterioration of the situation in France of around 533%. However, with Italy's much smaller usable agricultural area – around 48% the value of France – its total P surplus is about 2.6 times larger.

Certainly, a precise investigation of the situation in order to derive causes and dependencies and to allow for an assessment would call for more targeted reference categories than the usable agricultural area. Better judgments may be possible based on comparisons of the major crop types, the shares of high and low intensity cropping, grazing and fodder activities, and essentially by incorporating the geographic conditions. However, the example reveals that regions' geographic conditions and their differences in the status quo may have a tremendous impact on the meaning of the indicators. A precise judgment requires complementary information and comparisons across regions need to incorporate information on the local conditions as well.

Table 5 and 6 show the performance metrics, underlying aggregate indicators, reference, target and measured values for Czech Republic differentiated over its NUTS2 regions. The performance metrics turned out to be negative for all three domains. For the domain (C) Biodiversity conservation this is reflected in the outcome in each NUTS2 region. For the domains (A) Climate stabilization and (B) Clean air and water the result is negative for six out of eight NUTS2 regions.

Of the indicators that the metric (B) Clean air and water is derived from (5) Reduction of N air pollution is positive for most regions. Only Severovýchod showed a worsening of the situation. (3) Reduction of N water pollution turned out to be negative for all NUTS2 regions except for Severozápad. For (4) Reduction of P surplus the outcome is heterogeneous whereas for (5) Reduction of toxic substances use the result indicates a worsening of the situation in all NUTS2 regions.

The example demonstrates that using the information from the underlying indicators as well allows for more differentiated assessments on the whole performance domain than sticking to the metric alone. Improvements in the area of N emissions to the air seem to be more feasible or less compromised in most NUTS2 regions of the Czech Republic than in the areas of N water pollution and toxic substances use.

Table 5: Performance metrics and aggregate indicators for Czech Republic differentiated over NUTS2 regions

		Performance metric			Aggregate indicator				
Region	Category	Climate stabilization	Clean air and water	Biodiversity conservation	N surplus total	N air pollution	N water pollution	P surplus total	Toxic substances
Czech Republic	Metric/Indicator value	-2.06	-17.11	-2.58	-3.04	1.27	-3.04	-56.07	-10.58
	Reference value	6204.20		2.49	269.44	59.62	36.37	3.83	237617.72
	Target value	5880.35		2.99	260.46	57.64	35.16	3.71	229697.13
	Measured value	6870.66		2.42	296.71	57.10	40.06	11.00	303787.08
Praha	Metric/Indicator value	0.80	1.05	-3.01	-3.46	11.45	-3.44	39.51	-43.33
	Reference value	24.64		2.12	0.83	0.45	0.16	0.16	676.71
	Target value	23.35		2.54	0.81	0.44	0.16	0.15	654.15
	Measured value	23.61		1.99	0.93	0.28	0.18	-0.05	1563.30
Strední Cechy	Metric/Indicator value	-1.33	-17.42	-2.65	-0.45	2.21	-1.39	-57.64	-12.87
	Reference value	964.94		2.42	50.07	10.15	7.21	-1.79	39622.76
	Target value	914.57		2.91	48.40	9.81	6.97	-1.73	38302.00
	Measured value	1032.15		2.35	50.82	9.41	7.55	-5.23	53514.95
Jihozápad	Metric/Indicator value	-1.94	-20.72	-2.76	-1.65	1.40	-1.57	-75.00	-7.70
	Reference value	1330.18		2.32	52.69	12.41	6.44	2.58	44385.22
	Target value	1260.74		2.79	50.93	11.99	6.23	2.50	42905.72
	Measured value	1464.66		2.24	55.59	11.83	6.78	9.05	52714.03
Severozápad	Metric/Indicator value	0.93	7.86	-2.72	8.42	5.11	5.00	31.76	-10.41
	Reference value	494.57		2.29	29.66	4.46	2.54	1.38	19391.62
	Target value	468.75		2.75	28.67	4.31	2.46	1.34	18745.23
	Measured value	470.56		2.28	21.33	3.70	2.12	-0.08	24688.53
Severovýchod	Metric/Indicator value	-2.95	-3.82	-2.64	-7.40	-0.22	-5.48	1.12	-10.70
	Reference value	1040.26		2.49	40.70	9.64	4.97	1.65	36982.07
	Target value	985.96		2.98	39.35	9.31	4.81	1.59	35749.33
	Measured value	1200.40		2.36	50.74	9.70	5.88	1.59	47421.74

Table 6: Performance metrics and aggregate indicators for Czech Republic - continuation

		Performance metric			Aggregate indicator				
Region	Category	Climate stabilization	Clean air and water	Biodiversity conservation	N surplus total	N air pollution	N water pollution	P surplus total	Toxic substances
Jihovýchod	Metric/Indicator value	-3.11	-2.13	-2.48	-9.66	0.60	-6.50	10.01	-12.61
	Reference value	1255.54		2.57	49.54	13.00	8.23	-2.51	55131.36
	Target value	1190.01		3.08	47.88	12.57	7.95	-2.43	53293.65
	Measured value	1459.08		2.52	65.48	12.74	10.01	-1.68	74003.07
Strední Morava	Metric/Indicator value	-2.12	-0.26	-2.67	-2.93	0.10	-2.86	9.91	-8.22
	Reference value	707.68		2.41	28.85	6.29	4.51	0.69	27076.93
	Target value	670.74		2.89	27.89	6.08	4.36	0.66	26174.37
	Measured value	786.17		2.33	31.66	6.27	4.94	0.46	32597.96
Moravskoslezko	Metric/Indicator value	-2.36	-26.49	-2.81	-5.35	0.51	-3.95	-94.29	-8.23
	Reference value	386.40		2.30	17.11	3.23	2.29	1.68	14351.04
	Target value	366.23		2.76	16.54	3.12	2.22	1.62	13872.67
	Measured value	434.05		2.19	20.16	3.17	2.59	6.94	17283.49

When comparing the absolute values of the indicators it turns out that they differ tremendously in their range. The indicator (4) Reduction of P surplus shows the largest range of 133.8 between minimum of -94.29 and maximum 39.51. It is thereby around 11.5 times larger than the range of indicator (2) Reduction of N air pollution between its minimum of -0.22 and maximum 11.45. For the example of Czech Republic it is remarkable that large absolute values of indicator results tend to be connected to low levels of the underlying variable. The levels of P surplus are usually much lower in the NUTS2 regions of Czech Republic than the levels of N air and water pollution. Slight absolute changes in the measured value of 2013 thereby result in larger variations of the indicator result.

This clearly has an impact on the outcome for the performance metric. Indicators with a larger variation may turn the result from negative to positive and vice versa. Putting it differently, the uniform weighing might be skewed by large variations in the range of aggregate variables. This points again at the crucial role that the weighing of indicators may play. Table 7 presents metrics and indicators for Italy differentiated over NUTS2 regions. At the Member State level the results are positive for the performance domains (A) Climate stabilization and (B) Clean air and water and negative for the domain (C) Biodiversity conservation. Looking at the indicators underlying the domain (B) Clean air and water the results are heterogeneous at the Member State level with positive results for (2) Reduction of N air pollution and (4) Reduction of total P surplus and a negative outcome for (3) Reduction of N water pollution and (4) Reduction of toxic substances use.

The negative development in the biodiversity domain at the Member State level is reflected in the outcome at the NUTS2 level. Land use diversity decreased in all NUTS2 regions of Italy. However, for the other two performance domains the results at the NUTS2 level is heterogeneous. Both for (A) Climate stabilization and (B) Clean air and water there are around 40% of NUTS2 regions where the situation worsened compared to the reference year and around 60% of NUTS2 regions where the situation improved.

Results are similarly heterogeneous when looking at the aggregate indicators for the domain (B) Clean air and water. A difference to notice compared to the outcome for the NUTS2 regions of Czech Republic is that positive and negative results for the aggregate indicators are more evenly distributed over them. Whereas the outcome for the indicator (2) Reduction of N air pollution in the NUTS2 regions of Czech Republic turned out to be positive for almost all of them the results for Italy’s NUTS2 regions is much more diverse for the respective indicator. A pattern that seems to exist for Italy’s NUTS2 regions is that the results for the reduction of P surplus are mostly positive except for the region Lombardia. From this it can be followed that in Italy improvements in the area of P surplus seems to be more feasible and less compromised compared to the other areas of the performance domain (B) Clean air and water.

Table 7: Performance metrics and aggregate indicators for Italy differentiated over NUTS2 regions

	Performance metric			Aggregate indicator				
Region	Climate stabilization	Clean air and water	Biodiversity conservation	N surplus total	N air pollution	N water pollution	P surplus total	Toxic substances
Italy	0.22	0.36	-2.33	-0.40	1.19	-0.51	2.70	-1.95
Lombardia	-0.29	-0.67	-2.81	-1.36	0.42	-1.71	-0.76	-0.63
Emilia-romagna	0.37	0.83	-2.58	0.13	1.96	-0.51	4.92	-3.05
Lazio	0.60	0.78	-2.57	4.88	2.05	2.95	1.78	-3.67
Campania	1.47	2.66	-2.29	6.35	3.83	4.23	3.45	-0.89
Sicilia	1.31	0.50	-2.55	3.92	1.55	2.19	2.31	-4.05
Sardegna	0.61	2.36	-3.16	-1.11	1.32	0.95	2.81	4.38
Piemonte	-0.70	0.36	-2.66	-7.43	-0.12	-3.36	3.34	1.57
Valle d'Aosta	-0.78	-1.90	-4.07	-7.70	-1.05	-4.40	3.63	-5.76
Liguria	-1.46	-0.37	-3.32	-3.82	-0.12	-1.43	6.62	-6.54
Trentino-Alto Adige	-0.42	-0.98	-4.08	-2.34	-0.55	-0.53	0.45	-3.30
Veneto	1.23	1.85	-2.94	0.86	3.45	0.47	4.30	-0.82
Friuli-Venezia Giulia	-1.08	-2.62	-2.76	-3.39	1.35	-4.08	1.43	-9.20
Toscana	1.04	2.36	-2.43	1.54	1.15	1.07	6.99	0.23
Umbria	-0.01	-0.08	-2.37	-0.25	0.94	-2.48	5.99	-4.76
Marche	0.15	1.05	-2.63	-0.02	0.96	-1.43	6.41	-1.76
Abruzzo	0.23	0.85	-2.64	-0.31	0.64	-0.64	5.88	-2.47
Molise	0.03	2.83	-2.61	0.79	2.47	-0.56	5.28	4.15
Puglia	-0.45	-0.75	-2.57	-2.69	-1.62	-1.21	2.79	-2.96
Basilicata	0.22	-0.51	-2.55	-2.46	-0.69	-1.40	1.94	-1.87
Calabria	0.63	-0.51	-2.65	0.42	0.43	-0.01	3.49	-5.94

5 CONCLUSIONS

The procedure of selecting individual variables and the decisions on how to aggregate them in order to calculate the performance indicators has a critical impact on the quantitative outcome. By those decisions all assessments based on the indicators are centrally influenced.

For the metrics in the assessment domain (4) Reduced environmental impacts of the food system the way of calculating the aggregate indicators based on the procedure as suggested in chapter 3.3 is straightforward to implement and easy to interpret. The interpretation is linked to a reference situation in terms of a distance to go towards the policy goal. The definition allows for an interpretation in terms of "situation improved/worsened by x percent compared to the reference situation". This makes it possible to compare developments in different domains of performance metrics or aggregate indicators and over Member States or groups of countries. Using the metrics to assess the case studies in WP5, therefore, enables to identify whether or not implementing an innovation results in an improvement in terms of decreased distance towards the policy goal. By assessing different innovations it is possible to identify which innovations are needed to reach a certain policy goal. If for example policy makers have a certain set of targets they are aiming at, those innovations that contribute most to reaching the targets can be selected forming together an innovation pathway. The results of this deliverable, therefore, show that the metrics can be used to identify innovation pathways within WP5. However, there are several crucial issues that need to be taken into account.

The straightforward interpretation as presented in chapter 4 on the experimental results works fine if the increase or reduction of a variable is monotonically related to an improvement or worsening of the judged situation. This is straightforward for the variables of the domain (B) Clean air and water, where it is easy to follow the conclusion that, for example, less nitrate emissions are better. In case of variables where an overshooting might be a potential scenario connected to negative consequences it is important to ensure that this negative progress is reflected by the indicator as well. An example for such a case might be indicators that are related to peoples' nutritional situation. It would require that the distance to optimum in terms of over- or underachievement is penalized in both directions.

Elaborated and agreed-on methods for deriving the policy target values of the indicators is the essential task in making the performance metrics usable and meaningful. Policy targets might be derived from official documents or they might be scientifically justified or be based on consensus in the community of the users of the metrics. If policy targets can be derived from official documents or declarations we need a translation of this political intent into the desired levels of the variables. For most indicators it will not be possible to directly translate related policy targets into desired levels as it is the case, for example, with the European Union's global warming target. This requires additional assumptions, for example on the overall emission level in the timeframe that is considered to be consistent with the global warming target. Nevertheless, a political vision may be defined independently of the targets. For the example of global warming the long term vision might be an emission level that is close to the physical minimum.

Political targets might also vary over countries or be conflicting. The procedure of deriving the target needs to be standardized to ensure consistency of the indicators and comparability across countries. The process may start with a case-to-case consideration. However, with gaining experience it might be possible to see patterns that allow for a standardization of the procedure.

The selection of aggregate indicators and their weighing for calculating the performance metrics has an essential impact on the outcome for the metrics' value. If the weights are changed this may have a significant impact on the overall assessment. The metric "Clean air and water" at the European level turned out to be negative. However, if air pollution was given a slightly higher and water pollution and use of toxic substances a slightly lower weight, the performance metric would turn out positive. However, the approach as described in chapter 3.3 makes this transparent. The default uniform weighting may be modified if users have a different view or based on scientific rationale.

If metrics comprise different underlying indicators exploiting the information from them as well allows for more differentiated assessments on the whole performance domain than sticking to the metric alone. The overall outcome for the performance metric might conceal important developments at the more differentiated underlying indicators' level that might be contrary. If for example results show that over the different regions of a country certain environmental aspects tend to be more compromised concerning their improvement than other aspects this may point at important geographic differences that need to be taken into account as well.

Regions' geographic conditions and their differences in the starting point concerning the improvement area might influence the significance of indicators.

In a region where for example pollution from nitrate emissions is extraordinarily high due to weak farming practices the potential for reduction and thereby improvement is much higher compared to regions where more efficient farming practices are prevalent. Even though the indicator shows a great improvement the actual situation might still be much worse compared to the other regions with lesser improvements. Certainly, this also depends on the way the policy targets are defined. For the case of long term policy visions of, for example zero emissions in the far future, a linear trajectory path would entail greater absolute cuts for less efficient regions where all of them have to meet the visionary target in the same year. However, the example demonstrates that detailed evaluations demand for complementary information and comparisons across regions need to incorporate information on the local conditions as well.

6 REFERENCES

Armington, P.S. (1969). A theory of demand for products distinguished by place of production, Staff Papers, International Monetary Fund, Vol. 16, No. 1, pp. 159-178.

Benton T.G., Bryant D.M., Cole L. and H.Q.P. Crick (2002). Linking agricultural practice to insect and bird populations: a historical study over three decades, J. Appl. Ecol., 39 (2002), pp. 673–687.

Britz, W. and Witzke, P. (2014). CAPRI model documentation 2014. Institute for Food and Resource Economics, University of Bonn, Bonn.

Christensen, V., Guenette, S., Heymans, J. J., Walters, C. J., Watson, R., Zeller, D., and D. Pauly (2003). Hundred-year decline of North Atlantic predatory fishes. Fish and fisheries, 4(1), 1-24.

Coll, M., Libralato, S., Tudela, S., Palomera, I. and F. Pranovi (2008). Ecosystem overfishing in the ocean. PLoS one, 3(12), e3881.

Diana, J. S. (2009). Aquaculture production and biodiversity conservation. Bioscience, 59(1), 27-38.

Donald P.F., Green R.E. and M.F. Heath (2001). Agricultural intensification and the collapse of Europe's farmland bird populations, Proc. Roy. Soc. Lond. Ser. B: Biol. Sci., 268 (2001), pp. 25–29.

Dulvy, N. K., Sadovy, Y. and Reynolds, J. D. (2003) Extinction vulnerability in marine populations. Fish and fisheries, 4(1), 25-64.

EC (1991). Council Directive of 12 December 1991 concerning the protection of waters against pollution caused by nitrates from agricultural sources (91/676/EEC).

EU (2000). Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy. Off. J. European Union L327.

EC (2003). Directive 2003/87/EC of the European Parliament and of the Council of 13 October 2003 establishing a scheme for greenhouse gas emission allowance trading within the Community and amending Council Directive 96/61/EC. OJ L 275, p. 32–46.

EU (2006). Directive 2006/118/EC of the European Parliament and of the Council of 12 December 2006 on the protection of groundwater against pollution and deterioration, OJ L 372, p. 19–31.

Gocht, A. and Britz, W. (2011). EU-wide farm type supply models in CAPRI—How to consistently disaggregate sector models into farm type models. J. Policy Model. 33, 146–167. doi:10.1016/j.jpolmod.2010.10.006

Howarth, L. M., Roberts, C. M., Thurstan, R. H. and Stewart, B. D. (2014). The unintended consequences of simplifying the sea: making the case for complexity. Fish and Fisheries, 15(4), 690-711.

Jansson, T. and Heckelei, T., 2011. Estimating a primal model of regional crop supply in the European Union. J. Agric. Econ. 62, 137–152. doi:10.1111/j.1477-9552.2010.00270.x

Lewison, R. L., Crowder, L. B., Read, A. J. and S.A. Freeman (2004). Understanding impacts of fisheries bycatch on marine megafauna. Trends in Ecology & Evolution, 19(11), 598-604.

Reidsma P., Tekelenburg T., van den Berg and R. Alkemade (2006). Impacts of land-use change on biodiversity: An assessment of agricultural biodiversity in the European Union, Agriculture, Ecosystems & Environment, Vol. 114, No. 1, pp. 86–102, http://dx.doi.org/10.1016/j.agee.2005.11.026.

Petit S., Firbank R., Wyatt R. and D. Howard (2001). MIRABEL: models for integrated review and assessment of biodiversity in European landscapes, Ambio, 30 (2001), pp. 81–88.

Pinsky, M. L., Jensen, O. P., Ricard, D. and S.R. Palumbi (2011) Unexpected patterns of fisheries collapse in the world's oceans. Proceedings of the National Academy of Sciences, 108(20), 8317-8322.

Tillin, H. M., Hiddink, J. G., Jennings, S., and M.J. Kaiser (2006) Chronic bottom trawling alters the functional composition of benthic invertebrate communities on a sea-basin scale. Marine Ecology Progress Series, 318, 31-45.

Watson, R. A., et al. (2015). Marine foods sourced from farther as their use of global ocean primary production increases. Nature communications, 6.