European
Commission

JRC TECHNICAL REPORTS

The EU-Wide Individual Farm Model for Common Agricultural Policy Analysis (IFM-CAP v.2)

Manual of the model

Dimitrios Kremmydas
Athanasios Petsakos
Pavel Ciaian
Edoardo Baldoni
Pascal Tillie

2022

Joint Research Centre

EUR 30949 EN

This publication is a technical report by the Joint Research Centre (JRC), the European Commission's science and knowledge service. It aims to provide evidence-based scientific support to the European policymaking process. The scientific output expressed does not imply a policy position of the European Commission. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use that might be made of this publication. For information on the methodology and quality underlying the data used in this publication for which the source is neither Eurostat nor other Commission services, users should contact the referenced source. The designations employed and the presentation of material on the maps do not imply the expression of any opinion whatsoever on the part of the EU concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries.

Contact information
Dimitrios Kremmydas
Address: Edificio Expo. c/Inca Garcilaso, 3. E-41092 Seville (Spain)
E-mail: Dimitrios.Kremmydas@ec.europa.eu
Tel.: +34 954 488 420

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

JRC127014

EUR 30949 EN

PDF ISBN 978-92-76-46336-8 ISSN 1831-9424 doi:10.2760/248136

Luxembourg: Publications Office of the European Union, 2022

© European Union, 2022

The reuse policy of the European Commission is implemented by the Commission Decision 2011/833/EU of 12 December 2011 on the reuse of Commission documents (OJ L 330, 14.12.2011, p. 39). Except where otherwise noted, the reuse of this document is authorised under the Creative Commons Attribution 4.0 International (CC BY 4.0) licence (https://creativecommons.org/licenses/by/4.0/). This means that reuse is allowed provided appropriate credit is given and any changes are indicated. For any use or reproduction of photos or other material that is not owned by the EU, permission must be sought directly from the copyright holders.

All images © European Union 2022, except: cover page: *agricultural images* © https://pixabay.com/ (royalty free images)

How to cite this report:Kremmydas, D., Petsakos, A., Ciaian, P., Baldoni, E., Tillie, P., *The EU-Wide Individual Farm Model for Common Agricultural Policy Analysis (IFM-CAP v.2): Manual of the model*, EUR 30949 EN, Publications Office of the European Commission, Luxembourg, 2022, ISBN 978-92-76-46336-8, doi:10.2760/248136, JRC127014

Contents

Acknowledgements

The development of the Individual Farm Model for Common Agricultural Policy (IFM-CAP) was initiated by the workshop on ‘Development and prospects of farm level modelling for post-2013 CAP 2013 impact analysis’, organised jointly by the Joint Research Centre (JRC) and the Directorate General for Agriculture and Rural Development in Brussels on 6–7 June 2012. Model development started with a simplified prototype that was finalised in December 2015 (Louhichi et al., 2015). The first version of the model (Louhichi et al., 2018b) was used in the impact assessment of the Commission’s legislative proposal for the CAP post 2020. This document describes the second, and updated, version of the model (v.2), whose development has been undertaken by Dimitrios Kremmydas, Athanasios Petsakos, Pavel Ciaian, Edoardo Baldoni and Pascal Tillie, who are based at the Economics of Agriculture Unit at the JRC.

We would like to gratefully acknowledge Kamel Louhichi, Maria Espinosa, Angel Perni, Bouda Vosough Ahmadi, Liesbeth Colen and Sergio Gomez y Paloma, who were previously part of the IFM-CAP team at JRC Seville and who contributed to the development of the original version of the model. We would also like to thank Pierluigi Londero, Sophie Helaine, Koen Mondelaers and Koen Dillen from the Directorate General for Agriculture and Rural Development (Directorate C.2) for their support in the development of the IFM-CAP model, as well as Emmanuel Jacquin, Aurora Ierugan and Ana Torres Fraile from the Directorate General for Agriculture and Rural Development (Directorate C.3) for allowing us access to the Farm Accountancy Data Network (FADN) and for helping us to have a better understanding of the database.

Finally, we would like to thank Filippo Arfini (Department of Economics, University of Parma, Italy), Bruno Henry de Frahan (Catholic University of Louvain, Belgium), Alexander Gocht (Thünen Institute for Rural Studies, Germany), Thomas Heckelei (University of Bonn, Germany) and Alan Matthews (Trinity College Dublin, Ireland) for acting as members of the IFM-CAP advisory group and providing scientific advice regarding the methodological choices made. Their experiences and expertise on modelling agricultural policy impacts were very useful in improving our modelling approach and enriching our thinking and discussions.

Special thanks go to the present head of the Economics of Agriculture Unit, Giampiero Genovese, for supporting the development of the IFM-CAP.

Authors

Dimitrios Kremmydas, Athanasios Petsakos, Pavel Ciaian, Edoardo Baldoni, Pascal Tillie

Abstract

This report presents the second version of the EU-wide individual farm level model for common agricultural policy (IFM-CAP), which aims to assess the impacts of the post-2020 CAP reform on farm economics and the environment. The rationale for such a farm-level model is based on the increasing demand for a microsimulation tool capable of modelling farm-specific policies and capturing farm heterogeneity across the EU in terms of policy representation and impacts. Based on Positive Mathematical Programming, IFM-CAP seeks to improve the quality of policy assessment over existing aggregate and aggregated farm-group models and to provide an assessment of distributional effects on the EU farm population. To guarantee the highest representativeness of the EU agricultural sector, the model is applied to each EU-FADN individual farm (83 292 farms).

This report provides a detailed description of the IFM-CAP model (IFM-CAP v.2) in terms of the design, the mathematical structure, data preparation, modelling of livestock activities, allocation of input costs, and the calibration process. The theoretical background, the technical specification and the outputs that can be generated from this model are also briefly presented and discussed.

Preface

Building on the previous model documentation available in Louhichi et al. (2018b), this manual describes the IFM-CAP model and is written with three use cases in mind (**Figure 1**):

- First, it is addressed at readers who want to **use and interpret the results of the model** for evaluating a given policy. The aim is to obtain a general understanding of how the model functions (policy assumptions, data, calculations) without necessarily getting into too many details. Section 1 provides a brief summary of the model and is the most appropriate for those readers.
- Second, it is addressed at readers who want to **review and/or validate the model**. These readers are mostly concerned with the theoretical underpinnings of the model (i.e. assumptions about farmers' behaviour and general economic reasoning) and the reliability of the data used (validity of parameter estimation procedures, calibration). They should be able to find detailed information regarding these issues but without necessarily getting into the code implementation details. Section 2 is the most appropriate for those readers.
- Finally, this manual is also intended to be useful for those who want to **replicate, alter or contribute** to the codebase of the model. It will enable them to understand how the various components glue together and to learn how to locate the different model components in the code. Section 3 provides all the low-level details related to the design of the code.
-

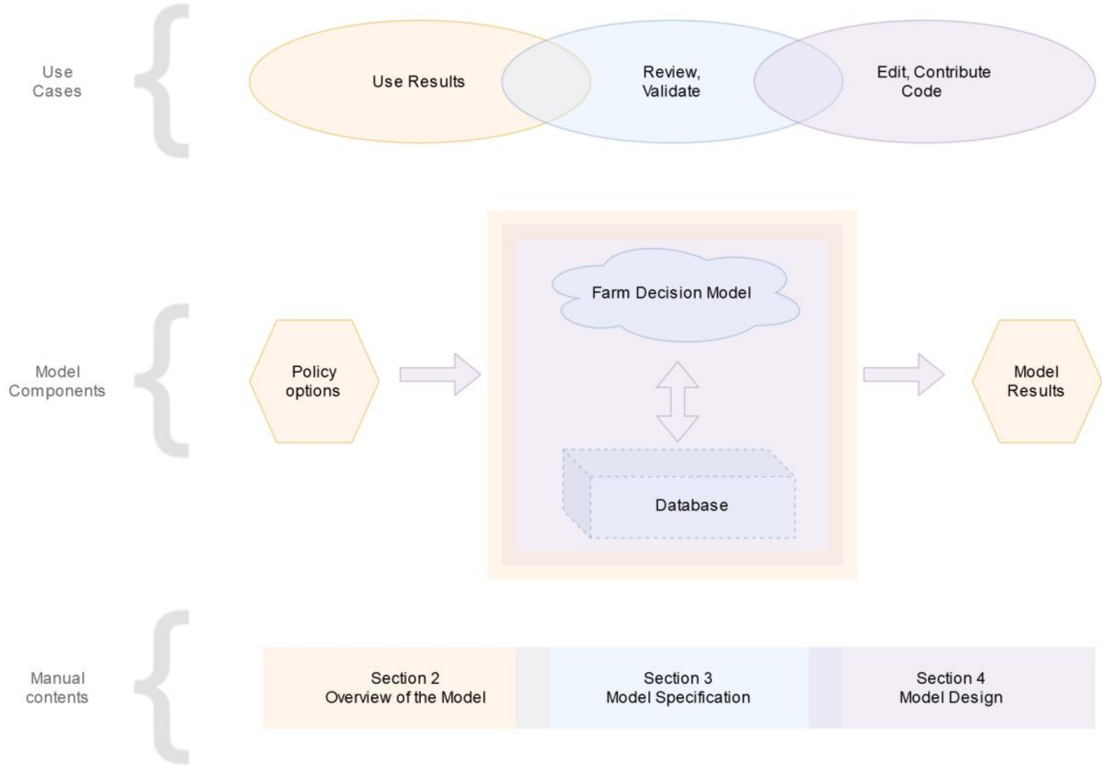


Figure 1. IFM-CAP manual reader use-cases

At first glance, the suggested use-case typology may seem overlapping; someone who is interested in interpreting the results of the model will possibly need more details about its theoretical basis. In addition, reviewing and validating the model cannot be done without examining the code. However, the manual organisation will allow the reader to guide himself from the more general information to the more specific one. Section 1 provides a brief summary of the model and corresponds to the first use-case (‘use the results of the model’). We briefly describe the motivation behind the model, its main assumptions, its core processes, and the necessary data required to run a policy scenario. All these issues are further analysed in Section 2, which provides all the theoretical details behind each model process and covers the requirements of the second reader use-case. Finally, Section 3 explains how the farm decision model and the database are translated into code.

1. A brief overview of the model

Why a farm level model?

In the past few decades, the Common Agricultural Policy (CAP) of the European Union has undergone a major shift from price support to support conditional on respecting specific environmental standards. At the same time, in an attempt to better address the heterogeneity of the agricultural sector within the EU, the approach to policy design has also changed significantly, moving from a 'one-size-fits-all' approach to measures targeting specific regions or farms. These developments have brought about an increased need for tools to model the various objectives of the CAP (income support, environmental sustainability, equity of direct payment distribution) at more disaggregate geographical scales, particularly at farm level.

The first important change to the CAP, which required the application of farm modelling to analyse its impacts, was the introduction of farm-specific decoupled payments as part of the 2003 reform (i.e. the single payment scheme). Subsequent CAP reforms introduced additional changes to decoupled payments, including modulation and capping of payments to large farms, redistributive payments for the first hectares used by farms, the small farmers scheme, and additional payments for young farmers, new entrants and farms located in disadvantaged areas. All these changes further affected payment heterogeneity across farms.

The need for farm-level models became more obvious after the 2013 CAP reform, which introduced greening as an additional conditionality layer of farm-specific obligations for receiving direct payments. Although several studies have attempted to model greening using more aggregated models (e.g. Gocht et al., 2017), the majority of studies about the 2013 CAP reform have been conducted using farm-level models. This reveals the suitability of such tools for *ex ante* policy analyses under the current CAP setting (e.g. Cortignani et al., 2017; Louhichi et al., 2018a; Solazzo and Pierangeli, 2016; Vosough-Ahmadi et al., 2015). Despite their different modelling assumptions, data sets and regional focus, most of these studies utilising farm models were able to produce consistent results that showed a limited impact of greening across EU farms. However, among the important limitations of these models is that most of them covered only selected Member States (MSs) / regions or specific agricultural sectors (e.g. Cimino et al., 2015; Solazzo et al., 2014; Vosough-Ahmadi et al., 2015), and they lacked a modular structure that was easily transferable and generic.

Furthermore, an important challenge for modelling tools in delivering a meaningful *ex ante* impact assessment is provided by the recent European Commission proposal for the CAP post 2020. This proposal aims to provide greater freedom for MSs to decide how best to meet the common objectives while responding to the specific needs of their farmers and rural communities (European Commission, 2018). The CAP post 2020 is the main support instrument envisaged to promote the transition to sustainable and inclusive agricultural production, as outlined in the European Green Deal and reflected in the farm-to-fork and biodiversity strategies (European Commission, 2020a,b). The European Green Deal is a comprehensive policy approach promoting transformation of the EU food system to one that is environmentally friendly, socially responsible, able to preserve ecosystems and biodiversity and able to contribute to a climate-neutral European economy. It takes a holistic approach by targeting the whole EU food system from farmers to consumers by covering food production, transport, distribution, marketing and consumption, and global trade and global food sustainability standards. As the farming sector supplies primary inputs to the food system and uses land in the production process that covers almost 50 % of the EU territory, its adjustment is pivotal for the transition to sustainable food systems, as envisaged within the Green Deal. As a result, the Green Deal could have significant impacts on the farming sector, both directly, because it targets the production process and farming practices, and indirectly, because of the feedback effects channelled from other stages of the food supply chain targeted by the Green Deal. To be able to analyse the economic, social and environmental impacts of the complex nature of the Green Deal's policy objectives requires modelling approaches that capture detailed aspects of farmer decision-making and farm structural change, as well as the interactions between farmers and both biophysical factors and other actors of the food supply chain.

The changes envisaged for the CAP post 2020, relate in particular to the objectives of

1. rebalancing farm income support towards small and medium-sized family farms and reducing it for large farms through capping;
2. stimulating farm structural change with respect to the entry of young farmers in the sector and improvement of intergenerational knowledge transfer;
3. enhancing the provision of public goods, biodiversity, ecosystem services and climate change mitigation and adaptation either through mandatory requirements or through voluntary schemes;
4. promoting greater requirements for EU agriculture to meet societal expectations on animal welfare, food quality, food safety and health issues; and
5. designing more effective rural development measures to promote growth, job creation, social inclusion and business development in rural areas including the development of bioeconomy (European Commission, 2018).

The analysis of these changes requires more detailed and flexible representation of policies in simulation models. The increased focus on agri-environmental interactions also requires better integration of biophysical and economic approaches, while the call for greater use of knowledge and innovation implies that farmers will be encouraged to invest in new production technologies. Assessing the impact of such policies calls for approaches that consider farm-level characteristics and alternative specifications for farmers' behaviour and preferences.

Given the above challenges, and because of the shortcomings of the agricultural policy modelling tools that use aggregated farms, the Joint Research Centre (JRC) developed an individual farm-level simulation model. The tool is called the Individual Farm Model for Common Agricultural Policy (IFM-CAP) and will be used for the *ex ante* assessment of medium-term adaptation of individual farmers to policy and market changes. The main expectations of this microsimulation tool are as follows: (i) it allows a more flexible and comprehensive assessment of a wide range of farm-specific policy measures that cannot be achieved with other models; (ii) it can be applied at EU-wide scale; (iii) it reflects the full heterogeneity of the EU commercial farm in terms of policy representation and impacts; (iv) it covers all the main agricultural production activities in the EU; (v) it permits a detailed analysis of different farming systems; and (vi) it enables the distributional impacts across the farm population to be estimated.

The model in a nutshell

IFM-CAP is a farm-level optimisation model of agricultural supply designed for the economic and environmental analysis of the European agricultural systems. It can be characterised as a template model. It consists of a number of individual farm models – one for each of the 81 107 individual farms in the farm accountancy data network[1] (FADN), covering all EU. All individual farm models have the same structure but use different farm-specific parameters, which in turn determine farm eligibility for specific policy measures (i.e. activation of the respective model constraints). IFM-CAP includes all FADN activities for crops (arable crops, vegetables and permanent crops, fodder and

---

(1) The FADN survey covers only those EU farms that, because of their size, could be considered commercial (the specific size threshold varies in each MS).

grassland, and fallow) and livestock (cattle, pigs, small ruminants and other animals). This means that the model provides EU-wide geographical and production coverage and that it is representative of the effects of CAP policy on the commercial farms of the EU.

IFM-CAP simulates a farm's decision to allocate resources to various crop and livestock activities as an optimisation problem. Each FADN farm selects the level of crop and livestock activities (in hectares and head of livestock, respectively) that maximises its expected utility of income. The expected utility of income for a farm is defined as the expected gross income (2) minus the risk premium, which represents the importance of uncertainty in the farm decision-making process. All CAP decoupled and coupled payments are part of a farm's expected income, which also includes those payments that depend on eligibility rules and on compliance with specific environmental measures.

The optimisation problem also includes technical constraints related to resource endowments, production relationships and policy. For example, the overall activity area of a farm cannot exceed the available land that the farm had in the reference year. Regarding livestock, a balance of the feed requirements with the feed supply (nutrient balances; minimum and maximum percentages of certain types of feeds) is also enforced. In addition, technical constraints are applied for modelling CAP in 2013–2020 (greening) and post 2020 (enhanced conditionality and eco-schemes).

The model uses data derived either directly from the FADN database or through estimation using the FADN and other variables. The observed crop and animal activity levels, subsidies and activity costs refer to the model's **base year** (currently 2017), while time series data (2012–2016) are used to calculate expected yields and prices. Procedures also exist for the identification and correction of out-of-range values and outliers and handling missing values.

Before IFM-CAP is used operationally for policy analysis, it needs to be calibrated. Calibration ensures that the farm optimisation problem, when solved for its base year, gives the same crop and livestock allocation as that observed. The reproduction of the observed activity levels is a critical issue for optimisation models of agricultural supply, and it is probably the most widely used validation criterion. The calibration of IFM-CAP relies on the principles of positive mathematical programming (PMP). The PMP approach used for IFM-CAP meets the calibration objective and it also ensures that the model response to parameter changes is consistent with exogenous information on price elasticities of supply. This last feature of the calibration process increases the probability that the model results will approximate the real-life adjustments that farms will make in response to the policies evaluated.

IFM-CAP is a **comparative static** supply model, that is, it compares the supply side of the EU agricultural sector with and without a policy change, but it does not examine the specific trajectories that led to the policy change simulated. Rather than providing forecasts or projections, the model performs 'what if' analyses; it simulates how a given policy scenario, when compared with a reference situation, can affect a set of performance indicators that are important to both decision-makers and stakeholders (Figure 2). The reference situation is called the **baseline** and it represents a mid-term projection of the European farming sector under the current policy setting, that is, it includes CAP provisions introduced with the 2013 reform (greening measures, capping). The baseline of IFM-CAP is technically a policy scenario itself. It is created by solving the model using adjusted prices and yields that are calculated from trends to 2030, taken from the CAPRI model (Britz & Witzke, 2014). These trends are based on a set of plausible assumptions regarding macroeconomic conditions and other variables of interest and are consistent with the European Commission's annual baseline projections

---

(2) At the beginning of the production year the farm decides which activities to engage in. Because at that time yields and prices are uncertain, the decision on activity levels is based on the income that these activities are expected to generate. This expected income in turn depends on the yields and prices that the farm expects to hold at the end of the production year.

of agricultural commodity markets (3). A more detailed discussion on the generation of the IFM-CAP baseline is given in Section 0.

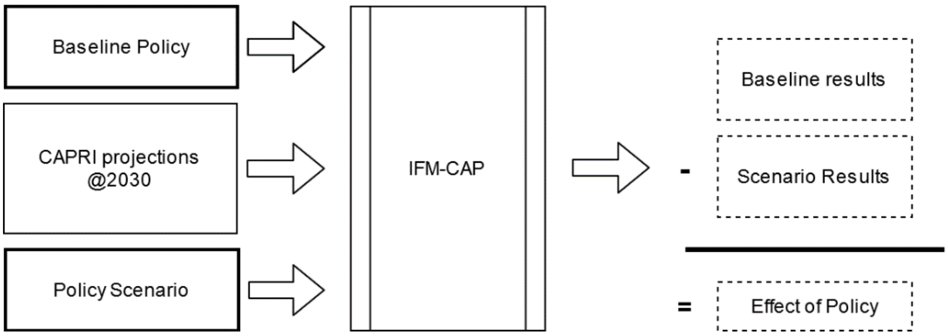


Figure 2. Comparative nature of IFM-CAP

All scenario specificities are modelled as changes over the baseline, which therefore serves as a counterfactual for scenario analysis. This may entail changing the value of some of the model's parameters (e.g. level of individual farm payments, parameters related to thresholds for environmental practices), defining new variables, modifying existing constraints or even introducing new ones. An example of how a policy scenario can be constructed is provided by the implementation of voluntary measures. In this case, additional constraints are introduced, together with the appropriate number of binary variables that control compliance and link the choice of the farm to the payment accompanying the voluntary measure. This particular example is indicative of the model's ability to simulate the self-selection of farms into voluntary measures and to provide estimates of adoption that are not possible with models operating at higher geographical scales.

After reparametrising and adjusting the model according to the desired policy scenario, the optimisation problem is run for each farm in the FADN sample. The main simulation outputs of IFM-CAP are:

1. land allocation to different crop activities (including fallow land);
2. herd size of different livestock activities, volume of feeds used and livestock density;
3. share of arable land in utilised agricultural area (UAA) and share of grassland in UAA;
4. land use change and agricultural production;
5. intermediate input use and intermediate input costs;
6. CAP first and second pillar subsidies and adoption rates of voluntary measures;
7. gross farm income and net farm income;
8. environmental impacts (biodiversity, soil erosion risk, input use).

Based on the resulting farm production choices, a set of production, economic and environmental indicators is calculated at farm level and aggregated using the FADN farm weights. These indicators are available both as averages at various aggregation levels (MS, farm type, economic size and any combination of these) and as distributions over the farm population or selected farm categorisations (e.g. distributions within a given type of farming). The effect of the policy scenario can be evaluated by comparing the

---

(3) For more information on the CAPRI projections, see Blanco-Fonseca (2010), Britz and Witzke (2014) and Himics et al. (2013, 2014).

values of the indicators produced by the policy scenario against their values under the baseline situation.

Some examples of questions that can be answered with IFM-CAP are ‘How is farm income affected by policy reforms?’, ‘Which farms would benefit and which would lose?’, ‘Are the economic impacts equally distributed across all farms?’, ‘Are small farms more affected than large farms?’ and ‘What is the production specialisation of the farms that are most affected?’

Important examples of applications of IFM-CAP cover the analysis of the future pathways for the European agriculture sector (M’barek et al., 2017), the evaluation of the impact of CAP greening (Louhichi et al., 2017; Louhichi et al., 2018a) and the contribution to the impact assessment of the European Commission proposal for the CAP post-2020 (European Commission, 2018b).

In Table 1 and Figure 3 we provide a summary of the main features of the IFM-CAP model.

Table 1. Main features of IFM-CAP

Model name	Individual Farm Model for Common Agricultural Policy Analysis (IFM-CAP).
Institution responsible for development and maintenance	JRC.D.4 Economics of Agriculture Unit (in-house model development and maintenance) and Directorate-General for Agriculture and Rural Development (Directorate C, user feedback)
Type of model	Individual farm model running for the whole FADN sample (and therefore all EU regions and sectors)
Methodology	Comparative static and non-linear programming model
Model calibration	Calibrated to base year using PMP
Objective function	Farm utility maximisation (revenues – accounting costs + subsidies – PMP terms – risk premium)
Revenues	Production value by activity: price × yield × activity level (hectares or head of livestock)
Accounting costs	Operating costs per unit of each production activity
Subsidies	First pillar policies: decoupled payments (single payment scheme, single area payment scheme, basic payment scheme, basic income support for sustainability, greening payments and payments related to voluntary measures) and coupled payments (voluntary coupled support (VCS)) Second pillar policies: payments related to pillar 2 reported in the FADN are assumed to be unchanged and independent of the policy scenario simulated, unless explicitly stated otherwise by the scenario assumptions
Risk premium	Constant absolute risk aversion (CARA) coefficient multiplied by the variance of revenues (and hence income) due to price and yield variations
Constraints	
Land constraint	Sum of area by activity less than or equal to total farm area endowment defined by type of use (arable and grassland)
Labour, capital	Captured by PMP terms
Policy constraints	Farm-specific measures: greening, voluntary agri-environmental pillar 1 measures, capping, modulation, regional ceiling for premiums, etc.
Livestock	Animal demography and livestock constraint, balancing feed demand and feed supply
Other considerations	
Expected prices and yields	Exogenous variables derived at farm level assuming adaptive expectations (based on past 5 years with declining weights)
Subsidies	Exogenous variables derived at farm level from the FADN for the base year
Input costs by activity	Input costs by activity are derived using econometric estimation (highest posterior density (HPD) estimation)
Total farm area endowment	Fixed at base year level
Technological progress	Yes, using an exogenous yield trend
Structural change	No
Changes in management practices	No
Environmental indicators	Crop diversity, soil erosion and input use proxies. Additional indicators, such as nutrient balance (nitrogen and phosphorus) and greenhouse gas emissions, are still under development

Input and output market interactions	No
Time horizon	2030 (extensive use of results from Aglink/CAPRI baseline work)
Potential scenarios	Any CAP scenario related to changes in payments and agri-environmental obligations at farm level, or self-selection of farms into voluntary measures
Model results	
Types of model results	Production, land use, land allocation among activities, farm income, variable costs, subsidies, environmental impacts, distribution of income and CAP benefits among farmers for each scenario (base year, baseline and policy scenarios)
Farm level	Single farm units
Farm group aggregation	By farm typology, farm size or other relevant dimension by using farm weighting factors from the FADN
Regional aggregation	FADN regions, Nomenclature of Territorial Units for Statistics (NUTS), MSs, EU
Data needs and other considerations	
FADN data	2012–2017 individual farm data
Other supporting data	Official statistical sources (e.g. Eurostat (regional statistics, Farm Structure Survey)), scientific literature and other model databases (e.g. CAPRI)
Programming language	General Algebraic Modelling System (GAMS), R language
Visualisation and data analysis	Graphical user interface (GUI) and Qlik (http://www.qlik.com/us/)

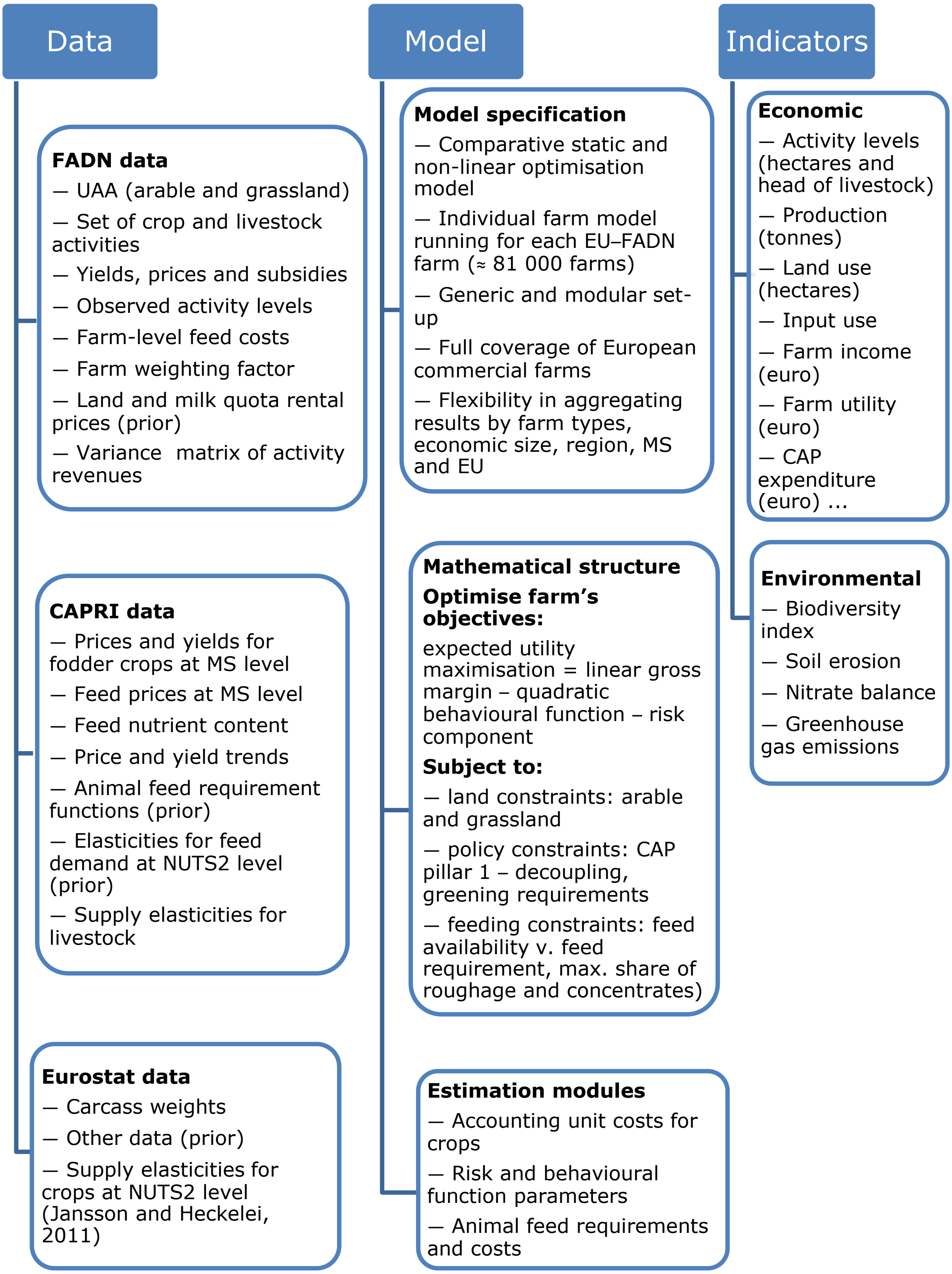


Figure 3. Data, model specification and output indicators for IFM-CAP

2. Model and data specification

IFM-CAP is a constrained optimisation model that maximises an objective function subject to a set of constraints. It assumes that farmers maximise their expected utility of income at given yields, product prices and production subsidies. They are also subject to resource (arable land, grassland, and feed requirements) and policy constraints such as greening and other agri-environmental restrictions. Land constraints are used to match the available land that can be used in a production operation and the possible uses made of it by the different agricultural activities. Constraints involving feed availability and feed requirements are used to ensure that the total energy, protein and fibre requirements of livestock are met by own-produced and/or purchased feed. In addition, for certain animal categories, in order to represent more realistically their diet, additional minimum or maximum requirements by type of feeding are introduced.

Section 2.1 presents in detail the farm optimisation problem, explains the underlying theoretical assumptions and provides a detailed mathematical representation of its objective function and the baseline constraints. An analytical presentation of how the model is parameterised is given in Section 2.2. This includes details of FADN data handling, estimation of costs and feed-related parameters, model calibration and generation of the baseline.

2.1. The farm decision model

Farmers' expected utility of income is defined following the linear mean-variance (E-V) approach (Markowitz, 1952) which assumes an exponential utility function and a normal distribution of income (Freund, 1956; Pratt, 1964; Arribas et al., 2017). According to this approach, which posits constant absolute risk aversion (CARA) preferences, expected utility maximisation is equivalent to the maximisation of the certainty equivalent income (CE), in other words, the lowest income level — known with certainty — for which the farmer would be indifferent between accepting it and engaging in a risky production plan. Denoting farm income by z, the CE is defined as expected income minus the associated risk premium (the cost of risk bearing), which, in turn, is equal to the product of the income variance $\mathrm{V}[z]$ and the CARA coefficient φ:

$$CE = \mathrm{E}[z] - \frac{\varphi}{2}\mathrm{V}[z] \tag{1}$$

The computational advantage of the selected E-V approach with CARA specification is one of the main reasons for its use in the IFM-CAP framework. We note that CARA is a rather restrictive behavioural assumption because it considers preferences of economic agents to be independent of the level of income (wealth). Nevertheless, it is widely employed in empirical agricultural research because of its mathematical simplicity, that is, the certainty equivalent income is defined as a linear function of the mean and variance of income. The resulting optimisation problem is a quadratic programming problem for which the literature provides several solution methods. More sophisticated, non-linear E-V specifications may consider the more plausible assumption of decreasing absolute risk aversion (DARA). (4). However, non-linear E-V models can become too difficult to evaluate numerically and may lead to non-convex programming problems (Hazell and Norton, 1986).

Farmers' expected income $\mathrm{E}[z]$ is defined as the sum of expected activity gross margins and decoupled farm payments minus a non-linear implicit cost function. The activity gross margin is equal to the total expected activity revenue, including sales from agricultural products and coupled compensation payments, minus the accounting

---

(4) See, for example, Coyle (1999), Petsakos and Rozakis (2015) and Sckokai and Moro (2006).

variable costs of production. The accounting variable costs include the costs of seeds, fertilisers, crop protection and feeding and other specific costs. The implicit cost function is a quadratic behavioural function introduced to calibrate the farm model to an observed base year with respect to activity levels and animal feeding practices, as is usually carried out for PMP models. It is intended to capture the effects of factors that are not explicitly included elsewhere in the model, for instance labour requirements, capital constraints, managerial capacity and other factors that may lead to increasing implicit marginal costs (5).

The FADN database provides only total accounting costs per variable input category (e.g. seeds, fertilisers, pesticides, feed), without indicating the unit input costs for each activity (crop and animal), which is needed to capture policy impacts and to represent technologies in an explicit way. For crop activities, we overcome this lack of information by using a Bayesian econometric estimation of unit input costs based on the farm-level input costs per category reported in the FADN database, assuming a Leontief technology, as explained in detail in Section 2.2.1.9. Unit input costs are estimated for the whole 2011–2016 period using cross-sectional data. For livestock activities, we use the farm-level feeding costs reported in the FADN database and various external sources to estimate animal feed (nutrient) requirements and to balance feed requirements and feed availability at farm level, as described in Section 2.2.2.

The separation of the Leontief production function (i.e. accounting variable costs) from the quadratic behavioural function was motivated by the fact that the primal technology representation through the Leontief production function has the following advantages:

1. provides an explicit link between production activities and the total physical input use
2. eases the link to calculating environmental indicators calculation and
3. allows the simulation of policy measures linked to specific farm management

According to Heckelei and Wolff (2003), the main disadvantage of this approach is the lack of rationalisation, as intermediate input uses are assumed to be independent of the unknown marginal costs captured by the quadratic behavioural function.

Regarding the income variance, $\mathrm{V}[z]$, most of the literature incorporates uncertainty in the gross margin per unit of activity (e.g. Cortignani and Severini, 2012; Jansson et al., 2014) or in the revenues per unit of activity (Arata et al., 2017; Coyle, 1999; Paris and Arfini, 2000; Petsakos and Rozakis, 2015; Sckokai and Moro, 2006). In the former case, the authors assume that prices, yields and costs are stochastic. In the latter case, the authors argue that costs are non-random variables because in static decision models all costs are known when decisions are made (Antle, 1983; Petsakos and Rozakis, 2015) or because costs are less stochastic than revenues from the farmer's perspective, so that the variance in the gross margin can be approximated by the variance in revenues (Jansson et al., 2014). In the IFM-CAP framework, we opted for the second approach by considering that uncertainty applies only to prices and yields (i.e. revenues) but without differentiating between sources of uncertainty. For each farm and activity, the expected revenue per hectare or per head given is calculated as the product of the expected yield and price assuming adaptive expectations (based on the past five observations with declining weights, that is, for 2011–2016). The same data are used to calculate the covariance matrix of activity revenues for each farm. Section 2.2.1.7 presents in detail the process for calculating these parameters.

An identical model structure is applied for all modelled FADN farms to ensure uniform handling of all the individual farm models and their results. In other words, individual FADN farms are represented by individual farm models that have identical objective

---

(5) In principle, any non-linear function that represents increasing marginal costs (or, equivalently, decreasing marginal productivity) can reproduce the base year solution. A quadratic (convex) cost function is the most popular non-linear behavioural function used in PMP models.

functions, constraints and variables to those in the generic format of IFM-CAP, although the values of their model parameters are farm specific. No cross-farm constraints or relationships are assumed in the current version of the model.

2.1.1. Generic mathematical formulation

The generic mathematical formulation of the expected utility maximisation problem for an individual farm is as follows:

$$\begin{aligned}\max_{x_i, \zeta_{i,m} \geq 0} \mathrm{E}[U] = &\sum_i E[gm_i] x_i + e - \sum_i x_i \left(d_i + 0.5 \sum_j Q_{i,j} x_j \right) \\ &- \sum_{i \in animals} x_i \zeta_{i,m} \left(d^F_{i,m} - 0.5 \sum_m Q^F_{i,m} \zeta_{i,m} \right) \\ &-0.5\varphi \sum_{i,j} x_i \Omega_{i,j} x_j\end{aligned} \tag{2}$$

subject to:

$$\sum_m \mathrm{A}^F_{n,m,v} \zeta_{i,m} \lesseqgtr b^F_{i,n,v} \; [\theta^F_{i,n,v}] \qquad \text{where } i \in \text{ set of animal activities} \tag{3}$$

$$\sum_i A_{t,i} x_i \leq b_t \; [\theta_t] \tag{4}$$

where indices i, j denote the agricultural (crop and livestock) activities, m denotes marketable commodities (i.e., feed purchased, and farm output sold in the market or used as animal feed),(6) t the resource and policy constraints related to activities (e.g., agricultural land, greening obligations), while v denotes animal feeding constraints and n the different types of nutrient or energy requirements. x_i is the level of activity i (hectares and head), $\zeta_{i,m}$ is the amount of feed m given to animal activity i (tons per head), $\mathrm{E}[gm_i]$ is the expected gross margin for activity i (EUR/ha or EUR/head), e denotes decoupled payments (EUR), d_i is the intercept of the activity-specific behavioural (implicit cost) function (the linear PMP terms), $Q_{i,j}$ is its slope (the nonlinear PMP terms —a diagonal positive semi-definite matrix), $d^F_{i,m}$ is the linear term of the behavioural function related to animal feeding, $Q^F_{i,m}$ is the nonlinear part of the same function (a diagonal positive semi-definite matrix), φ is the farmer's CARA coefficient and Ω_{ij} is the covariance matrix of activity revenues per hectare or per head. Inequality (3) represents the general structure of the animal feeding constraints, where $\mathrm{A}^F_{n,m,v}$ is a matrix of coefficients representing the amount of nutrient n in feed m, while $b^F_{i,n,v}$ is the quantity limit of nutrient n given to animal i (lower or upper, or satisfied as equality), and $\theta^F_{i,n,v}$ is the shadow price of the v-th feeding constraint. $A_{t,i}$ are coefficients for resource and policy constraints, b_t are available resource levels and upper bounds for policy constraints, while θ_t are their corresponding shadow prices.

---

(6) Mathematically this means that the set of feeds in IFM-CAP, and the set of farm outputs, some of which can be used as feeds themselves, are subsets of the set of all marketable commodities included in the model.

The expected activity gross margin is defined as:

$$\mathrm{E}[gm_i] = \sum_m p_m(1-\xi_m)y_{i,m} + v_i - C_i \tag{5}$$

where $y_{i,m}$ is the expected yield of output m from activity i, p_m denotes the expected price for commodity m (including for feed and young animals), ξ_m are estimated production losses, v_i are coupled payments linked to activity i, and C_i are the accounting variable costs. The calculation of variable costs differs between crop and animal activities. Specifically, for crops, $C_i = \sum_k c_{i,k}$, where k are intermediate inputs (i.e. fertiliser, seeds, crop protection, etc.) and $c_{i,k}$ are the per hectare costs of each input type. For animals, $C_{f,i} = \sum_{m \in Feed} p_m \zeta_{i,m}$, where feed m given to animal activity i is evaluated at price p_m.

Note that the model assumes that all variance on income stems from prices p_m and yields $y_{i,m}$, and that all other model parameters, including coupled payments, activity costs and all PMP terms (d_i, $Q_{i,j}$, $d^F_{i,m}$ and $Q^F_{i,m}$), are known with certainty.

Expected prices, yields, accounting unit costs, subsidies, matrix of coefficients and land availability are derived from the FADN database or calculated in the data preparation step (see Section 2.2). The unknown parameters d_i, $Q_{i,j}$, $d^F_{i,m}$, $Q^F_{i,m}$ and φ are recovered simultaneously in each NUTS 2 during calibration (see Section 2.2.2 for a detailed presentation of the calibration approach).

Below we provide the full specification of the farm decision problem in algebraic notation.

2.1.2. Detailed mathematical formulation

The objective function and the constraints of each farm model with all variables and parameters can be presented analytically as follows.

Objective function

$\sum_{m \in M}\left[p_m\left(\sum_{i \in C} y_{i,m}x_i\right)\right]$	Revenue of crop activities
$-\sum_k \sum_{i \in C} c_{i,k}x_i$	Variable cost of crop activities
$+\sum_{m \in M}\left[p_m\left(\sum_{i \in A} y_{i,m}x_i\right)\right]$	Revenue of livestock activities
$-\sum_{m \in F}\left[p_m\left(\sum_{i \in A} \zeta_{i,m}x_i\right)\right]$	Cost of feed
$-\sum_{m \in F}\left[p_m\xi_m\left(\sum_{i \in I} y_{i,m}x_i\right)\right]$	Cost of losses
$+(1-\beta^{DP})\sum_{i \in V}(v_i x_i^V)$	Voluntary coupled support, for that part of activity i that is entitled coupled payments (x_i^V). Binary variable β^{DP} takes a value of 1 when the farm does not comply with the minimum requirements for receiving direct payments (see equation (13))

$+(1-\beta^{DP})\left(e^{other}+h^{ent}\sum_{s} e_s\right)$	Decoupled payments related to mandatory farm measures. Parameter e_s denotes the payment per hectare of eligible land (h^{ent}) related to mandatory measure s. Binary variable β^{DP} takes a value of 1 when the farm does not comply with the minimum requirements for receiving direct payments (see constraint (13))
$+(1-\beta^{DP})\beta^{ELS}\left(h^{ent}\sum_{s} e_s^{ELS}\right)$	Decoupled payments related to voluntary farm measures. Parameter e_s^{ELS} denotes the payment per hectare of eligible land (h^{ent}) related to voluntary measure s Binary variable β^{DP} takes a value of 1 when the farm does not comply with the minimum requirements for receiving direct payments [see constraint (13)]. Binary variable β^{ELS} takes a value of 1 if the farm adopts the voluntary measures (e.g., eco-schemes) and therefore it is entitled to unitary payments e_s^{ELS} (see constraint (43))
$-\sum_{i\in I} x_i d_i$	Linear PMP terms for production activities
$-\frac{1}{2}\sum_{i\in I}\sum_{j\in I} x_j Q_{i,j} x_i$	Quadratic PMP terms for production activities
$-\sum_{i\in A}\sum_{m\in F} x_i d_{i,m}^{F}\zeta_{i,m}$	Linear PMP terms for feed to calibrate the feed input coefficient $\zeta_{i,m}$
$-\frac{1}{2}\sum_{i\in A}\sum_{m\in F} x_i \zeta_{i,m} Q_{i,m}^{F}\zeta_{i,m}$	Quadratic PMP terms for feed
$-\frac{\varphi}{2}\sum_{i\in I}\sum_{j\in I} x_j \Omega_{i,j} x_i$	Risk premium
$-\sum_{i\,\in\,CTC} \eta_i^{CTC} c^{CTC}$	Costs for introducing additional catch crop cover area, beyond what farms do by default in the base year. This additional cover may be needed to comply with greening/environmental requirements. For greening (2013 CAP), the additional catch crop areas are explained in constraints (28)–(31). For eco-schemes (post-2020 CAP2020), they relate to the winter cover obligation and are explained in equations (39)–(42)
$-\sum_{i\,\in\,PERM} \eta_i^{PERM} c^{PERM}$	Costs for introducing additional cover area between tree rows, beyond what farms do by default in the base year. The additional cover area may be needed to comply with mandatory or voluntary obligations (eco-schemes) for CAP2020+. They are explained in equations (36)–(38)

Land constraint

$$\sum_{i \in C} x_i = h^{tot} \tag{6}$$

The land constraint is specified as an equality so that non-productive activities such as leaving land fallow are now part of the constraint. This specification will always return a non-zero shadow price for land.

Constraints related to animal feeding

$\forall i \in A,$ $\forall n \in N^+$

$$\sum_{m \in F} \zeta_{i,m} \cdot \delta_{n,m} \geq b^F_{i,n} \tag{7}$$

This constraint ensures that the sum of different types of feed m given to animal i (sum of variable $\zeta_{i,m}$ over m) covers the minimum requirement for nutrient n. This minimum requirement is given by parameter $b^F_{i,n}$.

$\forall i \in A,$ $\forall n \in N^o$

$$\sum_{m \in F} \zeta_{i,m} \cdot \delta_{n,m} = b^F_{i,n} \tag{8}$$

This constraint ensures that the sum of different types of feed m given to animal i (sum of variable $\zeta_{i,m}$ over m) exactly satisfies the requirement for nutrient n. This requirement is given by parameter $b^F_{i,n}$.

$\forall i \in A$

$$\sum_{m \in FG^+} \zeta_{i,m} \cdot \delta^{DM}_m \geq b^{DM}_{i,FG^+} \tag{9}$$

This constraint ensures that the contribution to dry matter of feeds belonging to the specific feed type FG^+ must be greater than the minimum allowed, b^{DM}_{i,FG^+}.

$\forall i \in A$

$$\sum_{m \in FG^-} \zeta_{i,m} \cdot \delta^{DM}_{m^-} \leq b^{DM}_{i,FG^-} \tag{10}$$

This constraint ensures that the contribution to dry matter of feeds belonging to the specific feed type FG^- must be less than the maximum allowed, b^{DM}_{i,FG^-}.

Constraints related to voluntary coupled support (VCS):

$\forall i \in V$

$$x^V_i \leq x_i \tag{11}$$

The part of a production activity receiving coupled support (x^V_i) cannot exceed the overall level of that production activity in the farm (x_i).

$$\sum_{i \in V} x^V_i \leq b^V_{max} \tag{12}$$

This constraint ensures that the sum of activities receiving a specific type of coupled support cannot exceed a predefined maximum level (b^V_{max}).

Constraints related to direct payments

The minimum requirements for receiving direct payments differ among MSs as follows.

1. In some MSs, a farm is eligible for direct payments only if payments exceed a certain value threshold (in euro).
2. In other MSs, the threshold refers to a minimum eligible area (in hectares).
    (a) A farm receives direct payments if it has more eligible hectares than the threshold.
    (b) When a farm does not have the required level of eligible hectares, it can still receive payments if the level of CAP direct payments is above a given threshold and the farm receives at least EUR 1 in VCS payments related to animals.

Note: Cases (1) and (2) are mutually exclusive.

Source: European Commission (2019b).

$$\left(e^{other} + h^{ent} \cdot \sum_{s} e^{S}\right) + Z^{+} \cdot \beta^{DP} \geq b^{DP}_{min} \tag{13}$$

This constraint corresponds to case (1) described above. If the farm's total value of direct payments is lower than b^{DP}_{min}, then the binary variable β^{DP} takes a value of 1 (Z^{+} is a random big number), indicating that the farm is not eligible to receive direct payments. When the direct payments that the farm can receive are greater than b^{DP}_{min}, then β^{DP} can take a value of either 0 or 1, as both values will satisfy the constraint. However, as $\beta^{DP} = 0$ increases the value of the objective function, it will always be selected as the optimal solution.

$$h^{ent} + Z^{+} \cdot \beta^{DP} \geq b^{ent}_{min} \tag{14}$$

This constraint corresponds to case (2a) described above. If the farm's total eligible area is smaller than b^{ent}_{min}, then the binary variable β^{DP} takes a value of 1 (Z^{+} is a random big number), indicating that the farm is not eligible to receive direct payments. When the eligible area is greater than b^{ent}_{min}, then β^{DP} can take a value of either 0 or 1, as both values will satisfy the constraint. However, as $\beta^{DP} = 0$ increases the value of the objective function, it will always be selected as the optimal solution.

$$\sum_{i \in (V \cap A)} \left(v_i \cdot x_i^{V}\right) + Z^{+} \cdot \beta^{DP} \geq b^{DP}_{min} \tag{15}$$

This constraint corresponds to case (2b) described above and applies to farms whose total eligible area is smaller than b^{ent}_{min}, but which engage in animal activities that can potentially receive some VCS payments. The constraint works exactly like constraint (13). If the VCS payments are lower than the minimum value defined by b^{DP}_{min}, then the binary variable β^{DP} takes a value of 1 (Z^{+} is a random big number), indicating that the farm is not eligible to receive direct payments. When the VCS payments for the farm's animal activities are greater than b^{DP}_{min}, then β^{DP} can take a value of either 0 or 1, as both values will satisfy the constraint. However, as $\beta^{DP} = 0$ increases the value of the objective function, it will always be selected as the optimal solution.

Constraints related to crop diversification (CAP 2013)

$$\sum_{i \in ARAB} x_i \leq 10 + Z^{+} \cdot \beta^{10} \tag{16}$$

Binary variable β^{10} controls whether the total area of arable crops is greater than 10 ha. If this condition is not satisfied, β^{10} becomes 0 (Z^{+} is a random big number) and the farm is exempted from the 75 % diversification constraint in inequality (18)).

$$\sum_{i\,\in\,ARAB} x_i \leq 30 + Z^{+} \cdot \beta^{30} \tag{17}$$

Binary variable β^{30} controls whether the total area of arable crops is above 30 ha. If this condition is not satisfied, β^{30} becomes 0 (Z^{+} is a random big number) and the farm is exempted from the 95 % diversification constraint in inequality (19)).

$$\forall i \in R \qquad x_i - 0.75 \sum_{i\,\in\,ARAB} x_i \leq (1-\beta^{10}) \cdot Z^{+} + (\beta^{A} + \beta^{G}) \cdot Z^{+} \tag{18}$$

Article 44(1), paragraph 1, of Regulation (EU) No 1307/2013 (European Parliament and Council, 2013) states that:

> *Where the arable land of the farmer covers between 10 and 30 hectares and is not entirely cultivated with crops under water for a significant part of the year or for a significant part of the crop cycle, there shall be at least two different crops on that arable land. The main crop shall not cover more than 75 % of that arable land.*

If a farm has less than 10 ha of arable land (exempted from diversification), then $\beta^{10} = 0$ and the inequality always holds because the value of the left-hand side will always be lower than Z^{+} (a random big number). If a farm has more than 10 ha of arable land (not exempted from diversification due to land size), then $\beta^{10} = 1$ and no single crop i can occupy more than 75 % of the total area under diversification.

Binary variables β^{A} and β^{G} control exemptions to diversification stemming from the selected crop mix. They are defined in constraints (20)–(22) and (23)–(25), respectively. If any of these two binary variables is equal to 1, then the inequality is satisfied regardless of the value of the left-hand side.

$$\forall i, j \in R \qquad x_i + x_j - 0.95 \sum_{i\,\in\,ARAB} x_i \leq (1-\beta^{30}) \cdot Z^{+} + (\beta^{A} + \beta^{G}) \cdot Z^{+} \tag{19}$$

Article 44(1), paragraph 2, of Regulation (EU) No 1307/2013 (European Parliament and Council, 2013) states that:

> *Where the arable land of the farmer covers more than 30 hectares and is not entirely cultivated with crops under water for a significant part of the year or for a significant part of the crop cycle, there shall be at least three different crops on that arable land. The main crop shall not cover more than 75 % of that arable land and the two main crops together shall not cover more than 95 % of that arable land.*

If a farm has less than 30 ha of arable land, then $\beta^{30} = 0$ and the inequality always holds because the value of the left-hand side will always be lower than Z^{+} (a random big number). If a farm has more than 30 ha of arable land, then $\beta^{30} = 1$ and two crops i *and* j can never occupy more than 95 % of the total area under diversification. In this case, the 75 % diversification constraint (18)) is also satisfied. Note that a farm with $\beta^{30} = 0$ may still be subject to the 75 % diversification constraint.

Binary variables β^{A} *and* β^{G} control exemptions to diversification stemming from the selected crop mix. They are defined in constraints (20)–(22) and (23)–(25), respectively. If any of these two binary values is equal to 1, then the inequality is satisfied regardless of the value of the left-hand side.

$$x_{FALL} + x_{MAIF} + x_{OFAR} - 0.75 \sum_{i\,\in\,ARAB} x_i \geq (\beta^{A} - 1) \cdot Z^{+} \tag{20}$$

Article 44(3)(a) of Regulation (EU) No 1307/2013 (European Parliament and Council, 2013) states that:

> *[Crop diversification constraints] shall not apply to holdings:*
>
> *(a) where more than 75 % of the arable land is used for the production of grasses or other herbaceous forage, is land lying fallow, or is subject to a combination of these uses, provided that the arable area not covered by these uses does not exceed 30 hectares;*

This condition is modelled as follows. If the total area of fallow land (*FALL*), fodder maize (*MAIF*) and other fodder on arable land (*OFAR*) is greater than 75 % of the total arable area, and the arable area occupied by any remaining activities is less than 30 ha, then the farm is exempted from the diversification requirements ($\beta^{A} = 1$). This last condition is controlled by binary variable β^{A30} in constraints (21) and (22).

$$\sum_{i \in ARAB} x_i - (x_{FALL} + x_{MAIF} + x_{OFAR}) \leq Z^{+} \cdot \beta^{A30} + 30 \tag{21}$$

The binary variable β^{A30} controls the 30 ha condition for the exemption defined in constraint (20). This condition is modelled as follows. If the difference between total arable area and total area of *FALL*, *MAIF* and *OFAR* is greater than 30 ha, then $\beta^{A30} = 1$ and the farm is not exempted. In this case β^{A} becomes 0. If the difference between total arable area and total area of *FALL*, *MAIF* and *OFAR* is less than 30 ha, then β^{A30} can take a value of either 0 (exemption) or 1 (non-exemption), as both values will satisfy the constraint. However, if the binary variable β^{A} also takes a value of 1 because of equation (20), then β^{A30} will be forced to take a value of 0.

Binary variable β^{A30} is linked to binary value β^{A} by equation (22).

$$\beta^{A} \leq (1 - \beta^{A30}) \tag{22}$$

Binary variables β^{A} *and* β^{A30} cannot take a value of 1 at the same time.

$$x_{PGRA} + x_{RGRA} + x_{OFAR} + x_{MAIF} + x_{PARI} - 0.75 \cdot h^{ent} \geq (\beta^{G} - 1) \cdot Z^{+} \tag{23}$$

Article 44(3)(b) of Regulation (EU) No 1307/2013 (European Parliament and Council, 2013) states that:

> *[Crop diversification constraints] shall not apply to holdings:*
>
> *(b) where more than 75 % of the eligible agricultural area is permanent grassland, is used for the production of grasses or other herbaceous forage or for the cultivation of crops under water for a significant part of the year or for a significant part of the crop cycle, or is subject to a combination of these uses, provided that the arable area not covered by these uses does not exceed 30 hectares;*

This condition is modelled as follows. If the total area of permanent grassland (*PGRA*), rough grazing (*RGRA*), *OFAR*, *MAIF* and paddy rice (*PARI*) is greater than 75 % of the UAA (h^{ent}), and if the arable area occupied by any remaining activities is less than 30 ha, then the farm is exempted from the diversification requirements ($\beta^{G} = 1$). This last condition is controlled by binary variable β^{G30} in constraints (24) and (25).

$$\sum_{i \in ARAB} x_i - (x_{OFAR} + x_{MAIF} + x_{PARI}) \leq Z^{+} \cdot \beta^{G30} + 30 \tag{24}$$

The binary variable β^{G30} controls the 30 ha condition for the exemption defined in constraint (23). This condition is modelled as follows. If the difference between total arable area and total area of *OFAR*, *MAIF* and *PARI* is greater than 30 ha, then $\beta^{G30} = 1$ and the farm is not exempted. In this case β^{G} becomes 0. If the difference between total arable area and total area of *OFAR*, *MAIF* and

PARI is less than 30 ha, then β^{G30} can take a value of either 0 (exemption) or 1 (non-exemption), as both values will satisfy the constraint. However, if binary variable β^{G} is also 1 because of equation (23), then β^{G30} will be forced to take a value of 0.

Binary variable β^{G30} is linked to binary β^{G} by equation (25).

$$\beta^{G} \leq (1 - \beta^{G30}) \tag{25}$$

Binary variables β^{G} and β^{G30} cannot take a value of 1 at the same time.

Constraints related to permanent grassland preservation

$$x_{PGRA} + x_{RGRA} \leq (1 - w^{GRAS})(\bar{x}_{PGRA} + \bar{x}_{RGRA}) \tag{26}$$

Grassland areas (*PGRA* and *RGRA*) cannot decrease by more than (w^{GRAS}) % compared with the reference land allocation. The greening requirements for the 2013–2020 CAP require that $w^{GRAS} = 0.05$. As the measure applies at NUTS 2 level, or for a group of NUTS 2 regions, this constraint is activated in a second model run, after having calculated the unconstrained change in grassland areas. Specifically, if grassland decreases by more than ($\mathrm{w}^{\mathrm{GRAS}}$) % at regional level, then the farms that exceed this threshold solve again but with the additional grassland constraint.

$$x_{PGRA} + x_{RGRA} \leq \bar{x}_{PGRA} + \bar{x}_{RGRA} \tag{27}$$

For farms in Natura 2000 regions, the model does not allow any decrease in grassland areas (*PGRA* and *RGRA*).

Constraints related to ecological focus areas (EFA CAP 2013 – greening)

$$Z^{+} \cdot \beta^{15} \geq \left(\sum_{i \in ARAB} x_i - 15 \right) \cdot (1 - \psi) \tag{28}$$

This constraint controls whether the arable area of a farm is of adequate size to comply with the ecological focus area (EFA) requirements, while also considering possible exemptions because of large forest areas ($\psi = 1$ means exemption). If $\psi = 0$ and the farm has more than 15 ha of arable land, the constraint is satisfied when $\beta^{15} = 1$, which means that the farm needs to comply with the EFA requirements.

$$0.05 \sum_{i \in ARAB} x_i \leq \left[\sum_{i \in EFA} x_i \cdot w_i^{EFA} + \sum_{i \in CTC} g_i^{CTC} \cdot w_i^{EFA} \right] + (1 - \beta^{15}) \cdot Z^{+} + (\beta^{A} + \beta^{G}) \cdot Z^{+} \tag{29}$$

If the farm has more than 15 ha of arable land ($\beta^{15} = 1$) and is not exempted from greening obligations for other reasons related to types of crop activities ($\beta^{A}, \beta^{G} = 0$ means non-exemption), then the sum of crop activities that count towards EFA requirements and the area under cover need to be greater than 5 % of the farm's arable land.

The EFA contribution of crop areas (x_i) and of the areas covered by catch crops (g_i^{CTC}) is weighted by w_i^{EFA}. These weights are specified in Commission Delegated Regulation (EU) 639/2014 (European Parliament and Council, 2014a). In the case of IFM-CAP, they correspond to 0.3 for all green cover and nitrogen-fixing crops, and to 1 for fallow land.

$\forall i \in CTC$

$$g_i^{CTC} \leq x_i \tag{30}$$

The total area of crop i that is under green cover from catch crops (g_i^{CTC}) cannot exceed the total area allocated to crop i (x_i).

$\forall i \in CTC$

$$g_i^{CTC} = x_i \cdot \rho_i^{CTC} + \eta_i^{CTC} \tag{31}$$

The total area of crop i that is under green cover from catch crops (g_i^{CTC}) is defined as the sum of the default cover share applied by the farmer (ρ_i^{CTC}), plus any additional cover (η_i^{CTC}), which is accompanied by costs, as shown in the last term of the objective function.

Constraints related to crop rotation (post-2020 CAP)

$\forall i \in ARAB$

$$x_i \leq w_0^{ROT} \sum_{j \in ARAB} x_j \tag{32}$$

This is the mandatory crop rotation measure. Crop rotation of X years is modelled as a requirement to grow a crop i in an area that is no more than N/X of the total arable area, where N is the number of times a crop can appear over a sequence of X years. The N/X fraction is given by parameter w_0^{ROT}.

$\forall i \in ARAB$

$$x_i - \epsilon_i^{ROT} \leq w_1^{ROT} \sum_{j \in ARAB} x_j \tag{33}$$

This is the voluntary crop rotation measure. Crop rotation of X years is modelled as a requirement to grow a given crop in an area that is no more than N/X of the total arable area, where N is the number of times a crop can appear over a sequence of X years. The N/X fraction is given by parameter w_1^{ROT}.

Variable ϵ_i^{ROT} represents the area by which the farm overshoots the rotation requirement at the optimum ($\epsilon_i^{ROT} = 0$ suggests that the farm applies the rotation correctly; in other words, it adopts the rotation measure). Variable ϵ_i^{ROT} is linked to the binary variable β^{ELS} in the objective function through constraint (43), which controls the adoption of voluntary measures.

As $w_1^{ROT} < w_0^{ROT}$ (the voluntary constraint is more demanding than the mandatory constraint), equation (32) is also automatically satisfied when the farm adopts the rotation eco-scheme.

Constraints related to landscape elements (post-2020 CAP)

$$x_{FALL} \geq w_0^{FALL} \left(\sum_{j \in \{ARAB \cup PERM\}} x_j \right) \tag{34}$$

The mandatory area allocated to landscape elements is modelled as a fallow land share constraint. The constraint calls for the area of fallow land to be greater than, or equal to, a share (parameter w_0^{FALL}) of the UAA (the sum of arable and permanent crops).

$$x_{FALL} \geq w_1^{FALL} \left(\sum_{j \in \{ARAB \cup PERM\}} x_j \right) - \epsilon^{FALL} \tag{35}$$

The voluntary additional area allocated to landscape elements (eco-scheme) is also modelled as a fallow land share constraint. The constraint calls for the area of fallow land to be greater than, or equal to, a share (parameter w_1^{FALL}) of the UAA (the sum of arable and permanent crops).

If the farm does not adopt the eco-scheme, x_{FALL} is lower than is needed and the control variable ϵ^{FALL} becomes positive to satisfy the constraint. Variable ϵ^{FALL} is linked to the binary variable β^{ELS} in the objective function through constraint (43), which controls the adoption of voluntary measures.

As $w_1^{FALL} > w_0^{FALL}$ (the voluntary constraint is more demanding than the mandatory constraint), equation (34) is also automatically satisfied when the farm adopts the eco-scheme.

Constraints related to cover crops between tree rows (post-2020 CAP)

$$\sum_{i \in PERM} g_i^{PERM} \geq w_0^{PERM} \left(\sum_{i \in PERM} x_i \right) \qquad (36)$$

The total area of permanent crops that is under green cover according to the measure 'cover crops between tree rows' (sum of g_i^{PERM}) must be greater than, or equal to, a share of the total area allocated to permanent crops, given by parameter w_0^{PERM}.

$$\forall i \in PERM \qquad g_i^{PERM} \leq x_i \qquad (37)$$

This equation ensures consistency for permanent crop cover. The green area of permanent crop i, which is under green cover according to the measure 'cover crops between tree rows', cannot be greater than the area of that permanent crop.

$$\forall i \in PERM \qquad g_i^{PERM} = x_i \cdot \rho_i^{PERM} + \eta_i^{PERM} \qquad (38)$$

This is a balance equation for permanent crop cover. The green area cover related to permanent crop i (g_i^{PERM}) is defined as the sum of the default cover share applied by the farmer (ρ_i^{PERM}) and any additional cover (η_i^{PERM}), which is accompanied by costs, as shown in the last term of the objective function.

Constraints related to winter cover crops (post-2020 CAP)

$$\sum_{i \in WINT} g_i^{WINT} \geq w_0^{WINT} \sum_{i \in WINT} x_i \qquad (39)$$

The total area covered by winter crops that counts towards compliance with the mandatory obligation for winter crop cover (sum of g_i^{WINT}) must be greater than a percentage of the areas of those crops for which the measure may apply (e.g. spring wheat is in set $WINT$, whereas winter wheat is not). This percentage is given by parameter w_0^{WINT}.

$$\sum_{i \in WINT} g_i^{WINT} \geq w_1^{WINT} \sum_{i \in WINT} x_i - \epsilon^{WINT} \qquad (40)$$

This equation defines a voluntary, more demanding obligation for winter cover (eco-scheme). As in equation (39), the total area covered by winter crops that counts towards compliance with the winter crop cover eco-scheme must be greater than a percentage of the areas of those crops for which the measure may apply (e.g. spring wheat is in set $WINT$, whereas winter wheat is not). For the eco-scheme case, this percentage is given by parameter w_1^{WINT}, for which $w_1^{WINT} > w_0^{WINT}$.

If the farm does not respect the constraint, the total value of the green area is lower than that required for compliance and the control variable ϵ^{WINT} becomes positive to satisfy the constraint. Variable ϵ^{WINT} is linked to the binary variable β^{ELS} in the objective function through constraint (43), which controls the adoption of voluntary measures.

As $w_1^{WINT} > w_0^{WINT}$, equation (39) is also automatically satisfied when the farm adopts the eco-scheme.

$\forall i \in WINT$	$g_i^{WINT} \le x_i$	(41)

This equation ensures consistency for permanent crop cover. The area of crop i that is under green cover during winter (g_i^{WINT}) cannot be greater than the area of that crop.

$\forall i \in WINT$	$g_i^{WINT} = \eta_i^{CTC} + x_i \cdot \rho_i^{CTC} + +x_i \cdot \rho_i^{MULCH} + +x_i \cdot \rho_i^{WINT}$	(42)

This is a balance equation for winter cover. The total green area under winter cover is equal to the default areas with (i) catch crops, (ii) mulching and (iii) winter cover crops, plus any additional green area (η_i^{CTC}) that is needed to satisfy the farm obligation, which is accompanied by costs, as shown in the last term of the objective function. This additional area under winter cover is assumed to be covered by catch crops.

The default (base year) areas with catch crops, mulching and winter cover are defined as a share of the area allocated to the appropriate crops, where the shares are given by parameters ρ_i^{CTC}, ρ_i^{MULCH} and ρ_i^{WINT}, respectively.

Constraint controlling adoption of eco-schemes (post-2020 CAP)

	$\left(\sum_{i \in ARAB} \epsilon_i^{ROT}\right) + \epsilon^{FALL} + \epsilon^{WINT} \le (1 - \beta^{ELS}) \cdot Z^{+}$	(43)

Controlling the adoption of eco-schemes is based on the requirement that the farm adopts all the previously defined eco-schemes (rotation, landscape elements and winter soil cover). If any of the control variables do not take a value of 0 (i.e. the left-hand side of the equation), then the binary variable β^{ELS} becomes 0 to satisfy the constraint and the farm cannot receive the corresponding eco-scheme payment. Otherwise, if the left-hand side takes a value of 0, then β^{ELS} can become either 0 or 1, as both values will satisfy the constraint. However, as $\beta^{ELS} = 1$ increases the value of the objective function, it will always be selected as the optimal solution.

Sets

I	Set of all production activities
A	Set of animal activities ($A \subset I$)
C	Set of crop activities ($C \subset I$)
EFA	Set of EFA crops ($EFA \subset C$)
$ARAB$	Set of arable crops ($ARAB \subset C$)
$PERM$	Set of permanent crops ($PERM \subset C$)

$WINT$	Set of crops that can be covered during winter ($WINT \subset C$)
CTC	Set of crops that can receive green area cover by catch crops ($CTC \subset C$)
V	Set of activities receiving VCS ($V \subset P$)
M	Set of marketable outputs/commodities
F	Set of outputs/commodities (produced on farm or purchased) that can be used as animal feed ($F \subset M$)
FG^{+}	Set of feed groups (e.g. concentrated and rough types of feed) that must provide at least a minimum quantity of dry matter ($FG^{+} \subset F$)
FG^{-}	Set of feed groups (e.g. concentrated and rough types of feed) that cannot provide more than a maximum quantity of dry matter ($FG^{-} \subset F$)
N^{o}	Set of nutrients corresponding to feeding requirements that must be satisfied as equality
N^{+}	Set of nutrients corresponding to feeding requirements that must be satisfied as inequality

Continuous variables

x_i	Level of activity i (ha or head of livestock)
x_i^V	Level of activity i receiving VCS (ha or head of livestock)
q_m	Amount of output m under quota (tonnes)
$\zeta_{i,m}$	Amount of feed m given to animal activity i (kg fresh weight)
g_i^{CTC}	Total area of crop i that is under cover by catch crops (ha)
η_i^{CTC}	Additional area of crop i that is under cover by catch crops – beyond base year default cover (ha)
g_i^{PERM}	Total area of permanent crop i that is under green cover (ha)
η_i^{PERM}	Additional area of permanent crop i that is under green cover (the 'cover crop between tree rows' measure) – beyond base year default cover (ha)
g_i^{WINT}	Total area of crop i that is under green winter cover (ha)
ϵ_i^{ROT}	Area by which the farm overshoots the rotation obligation for crop i (ha)
ϵ^{FALL}	Area by which the farm overshoots the eco-scheme (voluntary) obligation of providing additional areas for fallow land and preserving landscape elements (ha)
ϵ^{WINT}	Area by which the farm overshoots the eco-scheme (voluntary) obligation of providing winter cover area (ha)

Binary variables

β^{DP}	Takes a value of 1 if a farm does not satisfy the minimum requirements for receiving direct payments
β^{ELS}	Takes a value of 1 if a farm adopts voluntary farm measures (e.g. eco-schemes)
β^{G}	Takes a value of 1 if a farm is exempted from greening measures because more than 75 % of the eligible agricultural area is permanent grassland, is being used for the production of grasses or other herbaceous forage or is being used for the cultivation of crops under water for a significant part of the year or for a significant part of the crop cycle, or is subject to a combination of these uses, provided that the arable area not covered by these uses does not exceed 30 ha
β^{A}	Takes a value of 1 if a farm is exempted from greening measures because more than 75 % of the arable land is used for the production of grasses or other herbaceous forage, is lying fallow or is subject to a combination of these uses, provided that the arable area not covered by these uses does not exceed 30 ha
β^{A30}	Takes a value of 1 if the arable land of the farm covers more than 30 ha and more than 75 % of the arable land is covered by grasses or other herbaceous forage or is lying fallow (farm not exempted from diversification)
β^{G30}	Takes a value of 1 if the arable land of the farm covers more than 30 ha and is not entirely cultivated with crops under water for a significant part of the year or for a significant part of the crop cycle (farm not exempted from diversification)
β^{30}	Takes a value of 1 if the arable land of the farm covers more than 30 ha
β^{10}	Takes a value of 1 if the arable land of the farm covers more than 10 ha
β^{15}	Takes a value of 1 if the arable land of the farm covers more than 15 ha

Parameters

$y_{i,m}$	Yield of activity i in output m (tonnes/ha or head of livestock)
p_m	Market price for output m (euro/tonne)
p_m^Q	Quota price for output m (euro/tonne)
$c_{i,k}$	Expenses for input k in activity i (euro/ha)
ξ_m	Percentage of losses of output m (%)
c^{CTC}	Cost related to introducing additional catch crop green cover area (euro/ha)
c^{PERM}	Cost related to introducing additional green cover area between tree rows (euro/ha)

d_i	Linear PMP terms for production activities
$Q_{i,j}$	Non-linear PMP terms for production activities. It is a diagonal $I \times I$ matrix
$d_{i,m}^F$	Linear PMP terms for feed
$Q_{i,m}^F$	Non-linear PMP terms for feed. $Q_{i,m}^F$ represents a diagonal $M \times M$ matrix for each individual i animal activity
φ	CARA coefficient
$\Omega_{i,j}$	Covariance matrix of activity profits
v_i	Unitary value of VCS for activity i (euro/ha or head of livestock)
e_s	Unitary value for each decoupled payment type s, linked to mandatory farm measures (euro/ha)
e_s^{ELS}	Unitary value for each decoupled payment type s, linked to voluntary farm measures (euro/ha)
e^{other}	Sum of any other payments received by the farm, e.g. pillar 2 (euro)
h^{ent}	Total area eligible to receive decoupled payments (ha). The eligible area in MSs implementing the single payment scheme is equal to the amount of the farm's entitlements, whereas in MSs that implement the single area payment scheme it is equal to the total agricultural area
h^{tot}	Total UAA (ha)
$\delta_{n,m}$	Content of nutrient n in feed m (kg of nutrient/kg of feed)
δ_m^{DM}	Dry matter content in feed m (kg of dry matter/kg of feed)
$b_{i,n}^F$	Requirement of nutrient n for animal activity i (kg/head of livestock)
$b_{i,FG}^{DM}$	Dry matter requirements for animal activity i from feed group FG (kg/head of livestock)
b_{max}^V	Maximum activity levels to receive VCS (ha or head of livestock)
b_{min}^{DP}	Minimum payment value for becoming eligible to receive direct payments (euro)
b_{min}^{ent}	Minimum entitlements for becoming eligible to receive direct payments (hectares)
Z^+	A large number, used in some equations that contain binary variables
ψ	Binary parameter for exemption from EFA requirements because of large forest area (0 = not exempted, 1 = exempted). This parameter is relevant for farms in Scandinavian and Baltic MSs
w_i^{EFA}	Weight of the green share of crop i that counts towards EFA requirements (%)
ρ_i^{CTC}	Default share of the area allocated to crop i that is under cover by catch crops

	in the base year (%)
ρ_i^{PERM}	Default share of the area of permanent crop i that is under green cover in the base year (%)
ρ_i^{MULCH}	Default share of the area of arable crop i that is under mulching in the base year (%)
ρ_i^{WINT}	Default share of the area of arable crop i that is under green winter cover in the base year (%)
w^{GRAS}	Permitted reduction in percentage of grassland (%)
w_0^{ROT}	Required share of each crop under a mandatory rotation scheme (%)
w_1^{ROT}	Required share of each crop under a voluntary rotation scheme (%). As voluntary farm obligations (eco-schemes) are defined as being stricter than mandatory obligations, $\rho_0^{ROT} < \rho_1^{ROT}$
w_0^{FALL}	Required share of mandatory fallow land for the 'fallow land and landscape elements' measure (%)
w_1^{FALL}	Required share of additional voluntary fallow land for the 'fallow land and landscape elements' measure (%). As voluntary farm obligations are defined as being stricter than mandatory obligations, $w_0^{FALL} < w_1^{FALL}$
w_0^{PERM}	Required share of mandatory permanent crop area under green cover for the 'cover crops between tree rows' measure (%)
w_0^{WINT}	Required share of mandatory winter cover area (%)
w_1^{WINT}	Required share of additional voluntary winter cover area (%). As voluntary farm obligations (eco-schemes) are defined as being stricter than mandatory obligations, $w_0^{WINT} < w_1^{WINT}$

2.2. The model database

Most of the model’s data are derived either directly or by data operations on the FADN database. More specifically:

1. **Farm and crop activity data.** The model’s data is based on the 2017 FADN dataset. All farms represented in the FADN sample of that year (81 107 farms), are included in the model. However, to parameterise some aspects of the model, past observations (2012–2016) are also required. Consequently, we load and process data for FADN 2012 to 2017. FADN contains mostly farm specific data while the model requires activity level data. For this, several estimations are performed. First, yields, product prices, production subsidies and accounting unit costs for all crop and animal activities in each farm need to be derived. Out-of-range (i.e., negative, outliers) or zero values for prices and yields are identified during the data screening process (see Section 2.2.1.5) and they are corrected using simple rules of thumb. Data on accounting unit costs for crops (i.e., specific costs related to seeds, fertilisers, crop protection and other crop-specific costs) are estimated using a Bayesian approach with prior information on input–output coefficients from the DG AGRI input allocation module (see Section 2.2.1.9). Costs for labour, energy, water, and capital resources are not explicitly included in the current version of the model. Instead, we assume that they are captured by the behavioural function (i.e., PMP terms). Based on the time-series data for prices and yields (assuming costs to be fixed), we obtain the expected gross margins for activities and the covariance matrix of activity revenues. In certain cases, for example, when individual farm-level data are missing from the FADN database, or when outliers are detected, we use the aggregated FADN data at the level of the farm type to replace missing values or outliers.
2. **Livestock activity data.** Livestock activities are inherently more complex than crop activities. For this, we have developed a dedicated livestock module. Sections 2.2.2.1 and 2.2.2.2 explains in detail the rationale behind the livestock modelling and the data pre-processing in IFM-CAP. Section 2.2.2.3 describes the mapping of the original FADN data on the IFM-CAP livestock activities. Section 2.2.2.4 describes the feeding costs estimation. They are estimated using a Bayesian approach with prior information on animal feed requirements from CAPRI and data on farm-level feed costs, feed prices, feed nutrient contents and fodder yields from FADN, CAPRI and Eurostat.
3. **Calibration data.** After estimating and imputing the necessary activity-level data, certain parameters that are included in the objective function allow the model to calibrate (i.e. to replicate the base year situation). We recover these parameters during the model’s calibration process, as described in detail in Section 2.2.3.
4. **Baseline data.** IFM-CAP is a static comparative model and this requires the estimation of a baseline. Section 2.2.4 describes in detail the assumptions and the construction of the baseline model.

2.2.1. Farm and crop activity data

2.2.1.1. *Farm-level data*

The data that are readily available in the FADN database are the available farmland (i.e. total UAA, arable land and grassland), structural characteristics (e.g. production specialisation, organic farming and belonging to a Natura 2000 area) and entitlements for receiving direct payments. We use these data for setting lower and upper bounds for resource and policy constraints and/or for defining exemptions from specific constraints. They are directly available from the FADN database and no transformation is required. They mainly include data from Tables A, B and E in the FADN database (Table 2).

Table 2. FADN tables

Table	Related data
A	General information on the holding
B	Type of occupation
C	Labour
D	Assets
E	Quotas and other rights
F	Debts
G	Value added tax
H	Inputs
I	Crops
J	Livestock production
K	Animal products and services
L	Other gainful activities directly related to the farm

2.2.1.2. *Times series data*

One related issue when constructing aggregate time series from FADN data is the continuity of Eurostat's NUTS 2 nomenclature from 2010 to 2017. Although the FADN database provides a relatively continuous FADN region nomenclature, the NUTS nomenclature is used when we need to combine data with data from other Eurostat databases. The NUTS nomenclature is revised regularly (7) and this may pose challenges when constructing time series at a NUTS level spanning several years. In addition, the FADN database may adopt a different Eurostat classification of NUTS in each of the survey years. Both issues are shown in Table 3, in which the percentage of unique NUTS 2 codes in the FADN database that match with the different NUTS 2 classifications is presented for each survey year between 2010 and 2017. From these percentages, it is clear that, in 2017, the FADN database adopted the Eurostat NUTS 2 classification for 2016. In fact, 100 % of the unique NUTS 2 codes used in 2017 in the FADN database could be matched to the 2016 NUTS 2 codes. In 2014–2016, 100 % of the NUTS 2 codes used in the FADN database matched those of the NUTS 2 classification of 2013 instead. From 2010 to 2013, given the relatively high percentages of matching codes (99.6 %, 99.6 %, 100 % and 100 %, respectively), it appears that the FADN database adopted the 2010 classification. These percentages reveal both that FADN adopts different Eurostat classifications in different survey years, and that Eurostat classifications change over time.

---

(7) https://ec.europa.eu/eurostat/web/nuts/history

Table 3. Percentage of NUTS 2 codes in the FADN database that match Eurostat classifications by survey year

FADN survey year	Eurostat 2016	Eurostat 2013	Eurostat 2010	Eurostat 2006
2017	**100.0**	83.8	79.0	70.3
2016	85.0	**100.0**	93.8	85.0
2015	85.0	**100.0**	93.8	85.0
2014	85.0	**100.0**	93.8	85.0
2013	80.2	93.8	**100.0**	87.2
2012	80.0	93.8	**100.0**	87.6
2011	80.7	94.6	**99.6**	87.0
2010	80.7	94.6	**99.6**	87.0

In order to construct time series data at NUTS level for IFM-CAP that address both issues at once, we built an artificial NUTS nomenclature that preserves continuity for the required time range, similarly to what was carried out by Baldoni and Ciaian (2021). Specifically, for regions that have changed NUTS name/code over time, we maintained a common name/code throughout the period. For regions that changed shape, reshaping was necessary. This involved either spatial aggregation or disaggregation of NUTS regions at different hierarchical levels.

2.2.1.3. Crop activities

The extraction of data from FADN regarding crop activities is relatively straightforward. The details that we present in this and the subsequent sections are referring to crop activities. For livestock activities, section 2.2.2 provides the relevant details.

In Table 4, we provide in detail the crop activities modelled in IFM-CAP. In total, we model 42 crop activities, which cover the vast majority of the FADN original activities (8). Although the model uses more aggregated activities (see the 'IFM-CAP2 code' column) than the FADN ones, the detection and correction of outliers and the imputation of missing values were performed on the more detailed FADN activity nomenclature, (see 'FADN code' column). This provides better estimates, as the original FADN data set contains more homogeneous activities. For example, in IFM-CAP the PULS activity contains three original FADN activities, that is, peas, field beans and sweet lupins (10210), lentils, chickpeas and vetches (10220) and other protein crops (10290). Checking for yield outliers in these original activities is more convenient, as the yield of each category is expected to have a lower dispersion.

The crop activity-related information that is readily available includes the total area and the irrigated area, the opening and closing valuation of any stocks, the quantities and the value of sales, the value of the farm household consumption, and the value of production that was used as an input for the holding, for example animal feed. Prices and yields are not reported directly and need to be calculated (see Section 2.2.1.4).

For activities that report by-products (wine, grapes and olives), the values of the by-products are allocated to the value of the main product. For some activities that are composed of very heterogeneous products (flowers and ornamental plants; nurseries; other permanent crops; aromatic and medicinal plants; mushrooms), we express the production as the output value and set the price to 1.

---

(8) We do not model the following activities in Table I of the FADN database: land ready for sowing leased to others (11300), kitchen gardens (20000), growth of young plantations (40800), unutilised agricultural land where no agricultural use is intended (50100) and wooded area (50200).

Table 4. Land use activities (crops)

Category	FADN code		IFM-CAP2 code	Crop name	Comments
	Up to 2013	2014 and after			
(1) Cereals	120	10110	SWHE	Common wheat and spelt	
	121	10120	DWHE	Durum wheat	
	122	10130	RYEM	Rye	Includes mixtures of rye and other cereals sown in the autumn (meslin).
	123	10140	BARL	Barley	
	124	10150	OATS	Oats	Includes mixtures of oats and other cereals sown in the spring.
	126	10160	MAIZ	Grain maize	Excludes sweet maize cobs for human consumption.
	127	10170	PARI	Rice	Excludes cereals and maize harvested green for animal feed (including silage).
	128 + 125	10190	OCER	Other cereals for the production of grain	Includes cereals harvested dry for grain and not recorded under previous items (other cereal mixes, millet, triticale, buckwheat, sorghum and canary seed).
	NB: Excludes cereals and maize harvested green for animal feed (including silage)				
(2) Pulses	360	10210	PULS	Peas, field beans and sweet lupins	
	361	10220		Lentils, chickpeas and vetches	Also includes chickling vetch.
	330	10290		Other protein crops	
	NB: Crops sown and harvested mainly for their protein content and for the production of grain. Excludes leguminous crops harvested green, those grown as vegetables and oil protein crops.				
(3) Oilseeds	331	10604	RAPE	Rape and turnip rape	Crops grown for the production of oil, harvested as dry grains.
	332	10605	SUNF	Sunflower	Crops harvested as dry grains.
	333	10606	SOYA	Soya	Crops harvested as dry grains.
	364	10607	OOIL	Linseed (oil flax)	Varieties grown for producing oil, harvested as dry grains.
	334	10608		Other oilseed crops	Other crops grown for their oil content, harvested as dry grain. Includes mustard, poppy, safflower (*Carthamus*), sesame seed, earth almond, peanuts, pumpkins for oil, flax other than fibre flax
(4) Other arable	130	10300	POTA	Potatoes (including early and seed)	Includes early and seed potatoes (i.e. propagation material).

	131	10400	SUGB	Sugar beet (excluding seeds)	Excludes seeds and seedlings
	144	10500	OCRO	Fodder roots and brassicas (excluding seeds)	Includes mangolds, swedes, fodder carrots, fodder turnips, forage kale, half-sugar mangolds, fodder parsnips, other fodder roots and brassicas. Includes swede, Jerusalem artichoke, yam and manioc when used for fodder. Excludes seed crops.
	134	10601	TOBA	Tobacco	
	347	10603	TEXT	Cotton	
	133	10602	OIND	Hops	
	373	10609		Flax	Varieties grown for producing fibre.
	374	10610		Hemp	Other plants grown for their fibre content, not mentioned elsewhere. Includes jute, abaca (Manila), sisal and kenaf
	Part of 348	10611		Other fibre plants	
	345	10612		Aromatic plants, medical and culinary plants	Plants or parts of plants used for pharmaceutical purposes, perfume manufacture or human consumption. Excludes chicory and tea and coffee
	346	10613		Sugar cane	
	Part of 348	10690		Other industrial crops not mentioned elsewhere	Includes chicory and miscanthus.
(5) Vegetables and flowers	136	10711	VGOF	Open field fresh vegetables, melons and strawberries	Crops grown in rotation with field-scale crops. The harvested production is generally used for industrial processing rather than directly for fresh consumption.
	137	10712	VGMG	Market gardening fresh vegetables, melons and strawberries	Crops grown under short rotation with other horticultural crops, with almost continuous occupation of the land and several harvests per year. The harvested production is generally used for fresh consumption rather than industrial processing.
	138	10720	VGUG	Under glass fresh vegetables, melons and strawberries	Crops grown under shelter (greenhouses, permanent frames, accessible plastic tunnels) during the whole or for the predominant part of the growing season
	140	10810	FLOW	Flowers and ornamental plants (outdoors)	
	141	10820		Flowers and ornamental plants (under glass)	
	NB: This category includes cauliflower and broccoli, lettuce, tomatoes, sweet maize for human consumption, onions, garlic, carrots, strawberries and melons. It also includes pineapple, sweet maize and leguminous crops grown as vegetables (e.g. green beans and peas). It includes vegetables grown for roots, bulbs or tubers (e.g. Jerusalem artichokes, sweet potatoes, yam, manioc, turnips and swedes for human consumption). It excludes potatoes and mushrooms and nurseries.				
(6) Fodder activities	147	10910	OFAR	Temporary grass (harvested green)	Grass plants for grazing, hay or silage included as a part of a normal crop rotation, lasting at least 1 crop year and less than 5 years. Includes mixtures of predominantly grass plants and other forage crops, grazed, harvested green or as dried hay
	Part of (327 + 328)	10922		Leguminous plants	Leguminous plants grown and harvested green as the whole plant mainly for forage. Includes various species of clover (annual or perennial, e.g. crimson, red, white, Egyptian, Persian, different types of lucerne/alfalfa) and other leguminous plants grown for fodder (e.g. sainfoin, sweet clover, vetches, trefoil, melilot, sweet lupins, serradella, fenugreek and

					sulla).
	Part of (327 + 328)	10923		Other plants harvested green not mentioned elsewhere	Other arable crops intended mainly for animal fodder, harvested green and not mentioned elsewhere. Includes annual crops – cereals, ray grasses, sorghum, certain graminaceous plants (e.g. meadow grass), plants belonging to other families (cruciferous), plants not mentioned elsewhere (e.g. rape, California bluebell) – if harvested green.
	142 + 143	11000		Arable land seed and seedlings	Includes seeds and seedlings of vegetables, flowers, horticultural plants and arable crops other than cereals, dry pulses, potatoes and oilseed crops.
	148	11100		Other arable land crops	Includes arable crops not mentioned elsewhere and typically of low economic importance.
	326	10921	MAIF	Green maize	All forms of maize not harvested for grain (whole cob, parts of or whole plant).
	150	30100	PGRA	Pasture and meadow, excluding rough grazing	Includes grassland grown for 5 years or more on cultivated land. Excludes pastures and meadow not in use
	151	30200	RGRA	Rough grazing	Includes low yielding permanent grassland (generally uncultivated and unfertilised land, including scrub, used as poor-quality pasture).
	314	30300		Permanent grassland not used for production and eligible for subsidies	Areas of permanent grassland and meadows no longer used for production purposes that, in line with Council Regulation (EC) No 73/2009 or, where applicable, the most recent legislation are maintained in good agricultural and environmental condition and are eligible for financial support.
	NB: Includes all ‘green’ arable crops intended for animal feed and/or renewable energy production, grown in rotation with other crops and occupying the same parcel for less than 5 years (annual or multiannual fodder crops). It includes cereals, industrial plants and other arable land crops harvested and/or used green (including dried hay). It excludes fodder roots and brassicas.				
(7) Permanent	Part of 349	40111	APPL	Apples	
	Part of 349	40112		Pears	
	Part of 350	40113	PEAC	Peaches and nectarines	
	Part of 350	40114	OFRU	Other fruit of temperate zones	Includes fruit tree plantations that are traditionally cropped in temperate climates for producing fruits, such as quinces, medlars, apricots, cherries (including sour cherries), plums (including mirabelle plums, greengages and damsons) and other stone fruit not specified elsewhere (e.g. sloes and loquats)
	353 + part of 182	40115		Fruit of subtropical or tropical zones	Includes fruit tree plantations that are traditionally cropped in subtropical or tropical climates for producing fruits such as annona, pineapples, avocados, bananas, lychees, papaya, mangos, guava, passion fruit, figs, other fruits of woody plants (e.g. dates, persimmons and pomegranate), prickly pear and kiwi.
	Part of 352	40120	BERR	Berry species	Berry plantations that are traditionally cropped both in temperate and in subtropical climates for producing berries. Includes blackcurrants, redcurrants and white currants, raspberries, gooseberries, blackberries, blueberries and cranberries. Mulberry trees, elderberries and sea buckthorn are also included.
	351	40130	NUTS	Nuts	Nut tree plantations that are traditionally cropped in temperate and subtropical climates.

					Includes walnuts, hazelnuts, almonds, dulcis, chestnuts and other nuts not otherwise specified (e.g. pine seeds and pistachio nuts).
	354	40210	CITR	Oranges	
	355	40220		Tangerines, mandarins, clementines and similar	Includes tangerines, mandarins, clementines, satsumas, mandarins’ oranges, kings and hybrids (e.g. fortune, ortanique, Clemenvilla/Nova and Nadorcott/Afourer).
	356	40230		Lemons	
	357	40290		Other citrus fruit	Includes other citrus fruit not mentioned elsewhere (e.g. bitter orange, bergamot, fingered citron, acid limes and fortunella).
	281	40310	TABO	Table olives	Plantations of varieties grown for producing table olives.
	282	40320	OLIV	Olives for oil production (sold as fruit)	Olive plantations grown for oil production but sold as fruit.
	283	40330		Olive oil	
	284	40340		Olive by-products	
	289	40411	TWIN	Quality wine with a protected designation of origin (PDO)	Wine from grape varieties normally grown for the production of wines with a PDO.
	294	40412		Quality wine with a protected geographical indication (PGI)	Wine from grape varieties normally grown for the production of wines with a PGI.
	295	40420		Other wines	Wine without a geographical indication. Grape varieties normally grown for the production of wines other than PDO and PGI wines, including varietal wines.
	286	40451		Grapes for quality wine with a PDO	Grape varieties grown for the production of grapes for PDO wines
	292	40452		Grapes for quality wine with a PGI	Grape varieties grown for the production of grapes for PGI wines.
	293	40460		Grapes for other wines	Grapes varieties grown for the production of wine without a geographical indication.
	285	40430	TAGR	Table grapes	Grape varieties grown for the production of fresh grapes.
	291	40440		Raisins	Grape varieties grown for the production of raisins.
	157	40500	NURS	Nurseries	Includes plants grown in the open air for subsequent transplantation: (a) vine and rootstock nurseries; (b) fruit tree and berry nurseries; (c) ornamental nurseries; (d) commercial nurseries of forest trees (excluding those for the holding’s own requirements grown within woodland); (e) trees and bushes for planting in gardens, parks, at the roadside and on embankments (e.g. hedgerow plants, rose trees and other ornamental bushes, and ornamental conifers), including in all cases their stocks and young seedlings.
	158	40600	OCRO	Other permanent crops	Includes osier willow, bamboo, rush, rattan, carob trees, tea, coffee and truffles.
	156	40700		Permanent crops under glass	

	139	60000		Mushrooms	
(8) Fallow land	315	11210	FALL	Fallow land without any subsidies	Fallow land for which no financial aid or subsidy is paid. Excludes areas of arable land taken out of production for more than 5 years, or under 5 years when the farmer clearly states that it is taken out of production (not only for resting)
	316	11220		Fallow land subject to the payment of subsidies, no economic use	Fallow land for which the holding is entitled to financial aid or subsidies

NB: FADN codes are given both for the data set up to 2013 and for the data set from 2014 onwards.

Source: European Commission document RI/CC 1680 v.6
accessible from https://circabc.europa.eu/sd/a/56ed82b0-9e19-4f92-893d-238919294204/RICC_1680_v2.0_accounting_year_2015.pdf

2.2.1.4.Prices and yields

Yields are relatively straightforward to calculate from the production quantity:

$$Y = PRQ/A$$

where Y is yield, PRQ is production quantity (tonnes) and A is the production area (⁹) (ha).

Calculating prices is less straightforward. In the previous version of IFM-CAP, prices were calculated as the quotients of the total output value to the production quantity:

$$p_1 = TO/PRQ$$

$$TO = SAV + FCV + FUV - (OV - CV)$$

where p is price, TO is the total output value, PRQ is the production quantity, SAV is the sales value, FCV is the farm consumption value, FUV is the farm use value, OV is the opening stock value and CV is the closing stock value. The opening and the closing stocks were valued at the beginning and at the end of the production year, respectively; these values may be different from prices at the time of sale.

In the FADN data set from 2014 onwards, the sales quantities are explicitly reported and thus the calculation of prices becomes more straightforward:

$$p_2 = \frac{SAV}{SAQ}$$

where SAQ is the sales quantity.

The two price definitions are equal when the ratio of the sales value to the total output equals the ratio of the sales quantity to the production quantity:

$$p_1 = p_2 \Rightarrow \frac{SAV}{SAQ} = \frac{TO}{PRQ} \Rightarrow \frac{SAV}{TO} = \frac{SAQ}{PRQ}$$

Still, often the sales quantities are not present or there are data inconsistencies, especially for the years close to 2014. For those reasons, we applied a hybrid approach (between the total output and the sales quantity) for calculating the prices. We use the following algorithm:

1. Calculate the price using Sales Quantity:
   IF { [Sales Quantity] > 0.01 AND [Sales Value] >0 }
   THEN [Price1]= [Sales Value] / [Sales Quantity]
2. Calculate the price using the Total Output Value:
   IF { [Production Quantity] >0 AND [Total Output] >0 }
   THEN [Price2]= [Total Output] / [Production Quantity]
3. If only [Price1] or only [Price2] is available (e.g. Production quantity is higher than zero and Sales quantity is zero), then use the one that is available
4. If both [Price1] and [Price2] are available, then take the minimum[10].

(⁹) The exact definition of this variable is 'main crop area, excluding follow-up crops'. Note that the production variable contains the production of both the main crop and the follow-up crops. This implies that, for a specific crop, we assign the production of both the main crop and the follow-up crops to the main crop area. We do this because, first, it was impossible to distinguish between the production quantity of the main crop and that of the follow-up crops and, second, there were many inconsistencies between the area of the 'main crop' and the areas of the 'follow-up' crops.

[10] This makes sense, since the [TO]/[PRQ] is an overestimation of the actual price. In reality
PRICE x (OQ+PRQ)=SV+FCV+FUV+CV
=> PRICE = (SV+FCV+FUV+CV)/(OQ+PRQ) ≥ (SV+FCV+FUV+CV)/(PRQ)

2.2.1.5. Outliers

The purpose of FADN data screening was to remove aberrations and to check the extent to which the data need to be adjusted to meet the IFM-CAP modelling requirements. The key data that were screened were yields and product prices. The outlier procedure was applied to the 2010–2017 time series.

Outliers are observations that are numerically distant from the assumed distribution of the data. In this case, outliers concern prices and yields and may arise for the following reasons.

- Prices and yields are derived from other FADN data (based on the total production value, production quantity and areas), so their values in some farms may deviate significantly from the rest of the sample if the underlying data do not contain sufficient information to identify their true values (e.g. because of high carry-over stock combined with high prices).
- Yields of, and prices achieved for, specific activities included in a given aggregated activity group (e.g. flowers, other cereals, other vegetables), as well as for crops whose yields are strongly dependent on climatic conditions or the variety cultivated (e.g. tobacco, potatoes, olive trees), are highly heterogeneous.
- Yields and prices may have been recorded under exceptional circumstances that cannot be considered 'normal', for example during adverse weather conditions or outbreaks of pests. In this case production is minimal or zero.
- It is possible that a farmer may have entered incorrect information in the FADN farm returns, in particular for output quantity and/or output value.

For prices and yields, we carried out normality tests and, for consistency reasons, we used the interquartile range (IQR), a non-parametric method, to determine the outliers. The IQR is a measure of statistical dispersion, being equal to the difference between the upper quartile and the lower quartile:

$$IQR = Q3 - Q1$$

More precisely, it is a trimmed estimator, defined as the 25 % trimmed mid range, and is the most significant basic robust measure of scale. It is the third quartile of a box and whisker plot minus the first quartile. An outlier is defined as any value that lies more than 1.5 times the length of the IQR from the first quartile (lower outlier) or the third quartile (upper outlier). Therefore:

$$If\ Xi < (Q1 - 1.5 \times IQR) > lower\ outlier$$

$$If\ Xi > (Q3 + 1.5 \times IQR) > upper\ outlier$$

Detection of outliers was conducted at NUTS 0 level for all FADN crops except fallow land and grasslands, as information on prices and yields is not reported for these crops. To calculate IQRs, prices and yields that were either zero or missing were not considered.

The values that were flagged as outliers were replaced with NUTS 2 imputed values (see Section 2.2.1.6). In Table 5, we provide an example of IQR outlier detection for cereal crops; the yield ranges before IQR outlier detection and after IQR outlier detection and the subsequent removal of outlier yields are provided.

Table 5. Yield ranges before and after IQR outlier detection for cereal crops

	Common wheat		Durum wheat		Barley		Oats		Rye		Rice		Grain maize		Other cereals	
	Before	After	Original	After	Original	After	Original	After	Original	After	Original	After	Original	After	Original	After
Belgium	1–13.5	6.6–13.5			0.6–60	6.1–13.2	1–10.5	3.8–10.5	1–8.1	2.9–8.1			0.7–84.1	10.1–20.9	0.8–11.4	4.1–11.4
Bulgaria	0.1–44	2.7–9	0.7–8.7	2.4–7	0.1–21.6	2.4–8.8	0–14.7	1.2–5.7	0.2–10.1	1–4.3	2.1–9.2	3–9.2	0–43.3	3.4–15	0.1–14.8	1.6–7.8
Czechia	0.2–10.7	4.2–10.7	0.6–8.7	3.3–8.7	0.2–10.6	3.8–9.2	0.1–9	2.8–7.5	0.5–9.2	3.6–9.2			0.3–38.8	5.8–16.3	0.3–10.8	3.3–10.8
Denmark	0–11.9	5.7–11.9			0–11.7	4.6–10.3	0–10.4	3.9–10.4	0–11.6	4.5–11.6			0–14.9	4.3–14.9	0.7–10	3.2–10
Germany	0.1–36.1	5.6–13.8	0.1–19	3.5–10.7	0.1–14.8	5-13	0.1–22.6	3.5–10.4	0.1–26	3.8–13.7			0–24.1	7.3–17.6	0.2–22.4	4.6–12.6
Estonia	0–9.5	1.9–8.4			0–8.3	1.8–8.3	0–9.1	1.6–6.6	0.2–9	1.5–7.9					0.1–8	1.1–8
Ireland	0–13.5	6.8–13.5			0–15.9	5.7–11.7	0–10.7	5.2–10.7	0.8–3	1.1–3					1–17.3	1–7.5
Greece	0.1–20	2.2–6.3	0–40	1.9–7.8	0.1–61.7	1.8–7	0.1–27.1	1.3–6.7	0.2–19.4	1.8–4.2	0.5–13.2	6.3–13.2	0.2–110.2	9.5–20.8	0–12	1.5–5
Spain	0.1–50.3	2.1–10.7	0.1–45.8	1.7–6.1	0–20.1	1.6–8.2	0–384	1.1–7.6	0.1–14 444	1.2–5.8	0.2–18.4	6–12	1–1 365.8	9.6–21.8	0–34.4	1.1–6.7
France	0.1–17.8	4.8–14.8	0–11.9	3.4–11.9	0.1–12.8	4.4–12.8	0.1–19.2	2.7–10.1	0.1–11.2	2.9–10.2	1.4–8.4	3.7–8.4	0–34.4	6.3–19.1	0.1–27.1	3.2–11.3
Croatia	0.4–10	3.7–10	6.7–6.7	6.7–6.7	0.3–9.2	2.9–9.2	0.3–10	2.4–6.3	1.8–6.4	3.1–6.4			0.1–50	5.5–16.8	0–32.1	3–9.5
Italy	0.2–82	4.1–11.3	0.2–202.4	2.7–8.6	0–25.3	2.7–10	0.1–10.1	2.1–5.9	0.2–21.8	2.3–9.1	1.4–12	5.8–9.6	0–225	7.8–20.8	0.1–56.6	2–13.5
Cyprus	0.3–55.6	1.1–5.8	0.1–48.6	1.4–7.7	0–15.2	1.3–6.7	0.1–6.7	1.4–6.7					3.3–20	5.5–20	0.5–15.6	3–3.8
Latvia	0.1–10.3	2.2–9.1			0.1–9.6	1.8–7.1	0.1–10	1.8–5.7	0.2–8.7	1.8–8					0–8.9	1.3–7
Lithuania	0.3–12.5	2.5–11.8			0.2–10.6	2.1–8.7	0.2–7.8	1.6–6	0.1–9.4	1.6–6			0.5–11	2.2–11	0.1–12.6	1.6–9.2
Luxembourg	0.9–40	4.6–10	4.4–4.7	4.7–4.7	0.5–9.8	3.9–9.8	0.3–9.4	3.2–9.4	0.3–12.5	3.5–9.8			0.8–18.6	4.5–18.6	0.8–9.6	4.2–9.6
Hungary	0.3–10	3.4–10	1.1–8	3.4–8	0–10.7	2.9–9.7	0.1–8	1.8–7	0.1–12.5	1.7–7.2	1–6.4	2.3–6.4	0.1–15.1	4.2–15.1	0.3–10	2.8–8.2
Netherlands	0.8–15.1	6.9–15.1			0.1–11.6	5.2–11.6	1.7–12.5	3.1–12.5	0.6–7.6	1.4–7.6			3.4–99.9	9.4–18.2	0.8–8.3	2.1–8.3
Austria	0.4–82.6	3.9–12.5	0.6–8.9	3.3–8.9	0.2–23.2	3.5–11.7	0.4–20.7	2.7–10	0.2–10.5	2.3–10.5			0–84.8	7.3–22.5	0.1–74	2.9–13.1
Poland	0–94.2	3.7–11	4.9–4.9	4.9–4.9	0.1–104.1	3.3–8.8	0.1–203.8	2.5–7.2	0.1–300.3	2.3–7.7			0.2–208.5	6.1–18.4	0.1–391	3–9.2
Portugal	0.1–6.6	0.9–4.4	1.1–6	2-6	0–6	0.9–6	0.1–30	0.7–3.8	0.2–6	0.8–2.2	0.3–9	3.7–9	0.4–60	1.7–21.2	0.5–9.6	0.6–1
Romania	0.3–1 920	2.8–8.2	0.8–6.6	2.8–6.6	0.3–11.5	2.6–7.8	0.5–15	1.6–7	1–7.4	2.2–7.4	1.2–5.8	3.5–5.8	0.1–2 773.3	3.5–11.4	0.2–10.9	2.8–8
Slovenia	0.3–20.7	3.4–12.1			0.3–48.4	3.2–11.7	0.3–40	2.2–10	0.3–8.1	2.2–8.1			0.1–870	5.6–25	0.2–15.6	2–10.2
Slovakia	0.1–12.3	2.6–11.2	0.7–9.5	2.7–9.5	0–10.5	2.2–10.5	0–8	1.3–6.9	0.3–11.8	1.7–8			0–19	3.6–16.6	0.4–17	2–7.4
Finland	0.1–8.2	2.9–8.2			0–8.1	2.6–7.8	0.2–10.5	2.3–7.8	0.3–8.5	1.7–8.5			4.3–4.3	4.3–4.3	0.1–11.5	0.3–3.8
Sweden	0.1–13.7	3.9–13.7			0.1–11.4	3–9.9	0.1–9.7	2.8–9.7	0.4–11.2	3.6–11.2			1-20	5–9.5	0.5–10	2.8–8
EU	0–1920	0.9–15.1	0–202.4	1.4–11.9	0–104.1	0.9–13.2	0–384	0.7–12.5	0–14 444.4	0.8–13.7	0.2–18.4	2.3–13.2	0–2 773.3	1.7–25	0–391	0.3–13.5

2.2.1.6.Imputing missing or incomplete data

Imputing missing or outlier yields and prices

The presence of at least one missing yield or price in a farm's data would require removing the farm completely from the model. Thus, in order not to exclude a significant number of farms, we replace outlier and missing prices and yields (e.g. when production quantity and sales quantity were unavailable) with existing non-missing and non-outlier data.

For this, we follow a cascading procedure in which we replace a missing data point with the median value for the year at the farm's NUTS 3, NUTS 2, NUTS 1, NUTS 0, neighbouring countries or EU level, depending on whether the number of observations is more than a certain threshold ([11]). In case the cascading procedure cannot find a valid data point, we replace the missing data point with the median value of the NUTS 0 and EU data points for the crop across all years.

For example, if the yield for barley is missing for a farm that belongs to EL432 for 2017, if there are more than three non-missing and no outlier observations in EL432 for that crop in 2017, we assign the median yield of those observations. If not, if there are more than four non-missing observations at the NUTS 2 level of the farm for 2017, we use the median of those observations, and so on. If we cannot find a median value from the neighbouring countries for that year, we replace the missing data point with the median value at the NUTS 0 level (EL) across all years and, if this is not available, we replace it with the EU median value for that crop across all years. At the end of this procedure there are no missing values.

In Table 6, we show the number of values that were either missing or flagged as outliers and had to be imputed. Around 21% of the yield data points and 23% of the price data points were imputed.

Table 6. Numbers of yield and price observations with imputed values for 2010–2017

Yields	Missing/outlier data points		526 964 (21 %)	Prices	Missing/outlier data points		574 537 (23 %)
	Number of replacements from	NUTS 3 (yearly)	320 848		Number of replacements from	NUTS 3 (yearly)	335 143
		NUTS 2 (yearly)	52 578			NUTS 2 (yearly)	71 750
		NUTS 1 (yearly)	22 617			NUTS 1 (yearly)	26 325
		NUTS 0 (yearly)	28 760			NUTS 0 (yearly)	37 011
		Neighbouring NUTS0(yearly)	38 317			Neighbouring NUTS0(yearly)	37 933
		NUTS 0 (all years)	29 500			NUTS 0 (all years)	64 660
		EU (all years)	34 342			EU (all years)	1 715

---

[11] This threshold was three observations at NUTS 3 level, four at NUTS 2 level, seven at NUTS 1 and NUTS 0 levels, and nine at the neighbouring NUTS 0 level.

Grassland production

For grassland-related activities[12], production values are missing in the vast majority of cases and need to be imputed. For this, we use the CAPRI database.

The CAPRI database contains two grassland activities: extensive (*GRAE*) and intensive (*GRAI*). We use the prices directly; *GRAE* corresponds to *CRG* activities and *GRAI* to *CGRSXRG* activities in the FADN database.

For yields, we observed that CAPRI values overestimate the observed grassland yields (13) and so we applied a correction procedure at the NUTS 2 level. First, we retrieved the mean grassland yield from which the CAPRI *GRAE* and *GRAI* yields were computed:

$$\left.\begin{aligned} yield_{GRAE} &= 0.6 \cdot yield_{mean} \\ yield_{GRAI} &= 1.4 \cdot yield_{mean} \end{aligned}\right\} yield_{mean} = \frac{yield_{GRAE} + yield_{GRAI}}{2}$$

$$\frac{yield_{GRAE}}{yield_{GRAI}} = \frac{0.6 \cdot yield_{mean}}{1.4 \cdot yield_{mean}} \Rightarrow yield_{GRAE} = 0.428 \cdot yield_{GRAI}$$

Then, given this mean yield, for the yields of the FADN *CRG* and *CGRSXRG* activities, it holds that

$$yield_{mean} = yield_{CRG} \cdot share_{CRG} + yield_{CGRSXRG} \cdot share_{CGRSXRG}$$

$$yield_{mean} = 0.428 \cdot yield_{CGRSXRG} \cdot share_{CRG} + yield_{CGRSXRG} \cdot share_{CRG}$$

$$yield_{CGRSXRG} = \frac{yield_{mean}}{0.428 \cdot share_{CRG} + share_{CGRSXRG}}.$$

where $share_{CRG}$ and $share_{CGRSXRG}$ are the shares of *CRG* and *CGRSXRG* in the total grassland area in a NUTS 2 region.

Straw production

We calculate straw production using the residue-to-crop ratio (RCR) as a function of crop yield. The RCR indicates how much residue is produced as a function of the main agricultural crop produced measured on a total dry matter basis. The RCR can vary widely, depending, for example, on the type of crop, crop productivity, crop mix, crop variety, climate conditions and agricultural practices. Based on a review of the literature, Scarlat et al. (2010) reported an RCR of between 0.6 and 2.8, depending on the crop type and the study reviewed. Edwards et al. (2005) estimated a cereal RCR function for the EU based on a wide range of studies. Their estimated ratio ranged between 0.62 and 0.94 and was negatively correlated with the cereal yield. Koopmans and Koppejan (1997) reported RCRs for 13 crops of between 0.2 and 4, depending on the crop and the study reviewed. Furthermore, this literature implies that the amount of residue produced can be linked to crop productivity and can be approximated by a functional form (negatively) depending on the crop yield (Edwards et al., 2005; Scarlat et al., 2010).

Following Scarlat et al. (2010), we assume the following relationships between RCR and yield:

1. $RCR_{wheat} = 1.6057 - 0.3629 \ln(Yield_{wheat})$

---

(12) CRG: Rough grazing; CGRSXRG: Pasture and meadow.

(13) Although the majority of the production values for *CGRSXRG* are missing, there are a number of data points that can be used to estimate yields at country level.

2. $RCR_{rye} = 1.5142 - 0.3007 \ln(Yield_{rye})$
3. $RCR_{oats} = 1.3002 - 0.1874 \ln(Yield_{oats})$
4. $RCR_{Barley} = 1.3796 - 0.2751 \ln(Yield_{Barley})$
5. $RCR_{maize} = 1.3373 - 0.1807 \ln(Yield_{maize})$
6. $RCR_{rice} = 3.845 - 1.2256 \ln(Yield_{rice})$
7. $RCR_{sunflower} = 3.2189 - 1.1097 \ln(Yield_{sunflower})$
8. $RCR_{rapeseed} = 2.0475 - 0.452 \ln(Yield_{rapeseed})$
9. $RCR_{other\ crops} = 2.0311 - 0.5118 \ln\left(Yield_{other\ crops}\right)$.

where $Yield_i$ is yield (tonnes/ha) for crop *i*. Note that the coefficient corresponding to the RCR for other crops is calculated as the average coefficient value over all crops.

The straw yield, *StrawYield$_i$*, for crop *i* is obtained by multiplying the RCR calculated in no 1-9 by crop yield (in fresh weight per year) and the collection rate, *CollRate$_i$*:

10. $StrawYield_i = CollRate_i \, RCR_i \, Yield_i$

The actual residue collection rate varies depending on a number of factors, such as collection equipment used, crop variety, harvest height, yield and environmental requirements. Studies provide estimates of crop collection rates of between 30 % and 75 % (Bakker, 2013; Scarlat et al., 2010). Following Scarlat et al. (2010), we assume collection rates as reported in Table 7.

Note that in official statistical sources (e.g. Eurostat and the Food and Agriculture Organization), crop yields are usually not recorded on a dry matter basis. Instead, they are recorded in the form in which the crops are harvested (fresh or wet weight). As a result, the straw yield calculated in equation 10 is not measured on a dry matter basis but contains the moisture level of the grain crop (i.e. between 15 % and 20 % depending on the crop).

Table 7. Crop residue collection rates

	Collection rate, *Collrate*(%)
Wheat	40
Rye	40
Oats	40
Barley	40
Maize	50
Rice	50
Sunflower	50
Rapeseed	50

Source: Scarlat et al. (2010)

2.2.1.7. Revenue expectation and variance

As defined in the model's objective function, farmers are assumed to maximise their **expected** utility of income, that is, farmers' decision-making is based on **expected** prices and yields for given costs (i.e. only revenue is assumed to be stochastic). Although the E-V framework dictates that expected values for these parameters are the mean of the respective distribution for each farm, many authors consider this assumption to be unrealistic and suggest alternative methods of calculating farmer expectations (e.g. Brink and McCarl, 1978; McCarl and Spreen, 2003). The formulation of expectations for the calibration of the IFM-CAP model is based on the theory of adaptive expectations, which results in expectations being a weighted average of past observations. Expected prices and yields are constructed at the level of the farm type or NUTS 2 region. Subsequently, the individual farm-level expectation is constructed to account for farm-specific factors.

The adaptive expectations theory assumes that expectations are revised based on past forecasting error. This corresponds to expected prices being equal to the weighted average of past prices with geometrically declining weights (Nerlove, 1958). Alternative formulations, such as naive expectations, in which expected values are set as equal to the last observed prices, or future prices, were considered. However, based on the available empirical literature, a statistical test based on FADN data and data limitations (14), a simplified version of the adaptive expectation approach was considered. This approach (i.e. covering only five periods in the past) can be applied to all products and is believed to provide the best solution for the IFM-CAP model.

Ideally, the model would generate expected prices at the individual farm level to account for farm-specific transaction costs and quality differences across farms. However, as not all activities and products are observed at all farms or in each of the past 5 years, an approach for generating farm-specific expected prices for every product is proposed. In the first step, average expected prices are generated for each farm type in each NUTS 2 region, consisting of the weighted average over the past 5 years of prices at farm type level. The weights approximate geometrically declining weights, that is, recent observations get a higher weight than observations made in the past (15).

In the second step, the average farm-specific deviation from the average price for that farm type is calculated for each product based on the actual farm-specific prices observed in the past (i.e. over the period 2012–2016). This deviation is then added or subtracted from the average for the farm type to obtain an individual farm-specific expected price. Note that this farm-specific deviation is assumed to remain constant over time. As such, it will not influence the expected prices in the baseline or other scenarios (i.e. each farm within the same farm level and region will experience the same absolute price change in the scenarios). Below, the different steps in the construction of the expected prices and, accordingly, yields and unit costs are described in detail.

---

(14) The empirical literature on different models of price expectations is inconclusive, without a clear preference for backward-looking prices (naive or adaptive expectations), (quasi-)rational expectations, future prices or monthly prices (Chavas, 2000; Haile et al., 2016; Kenyon, 2001; Nerlove and Bessler, 2001; Shideed and White, 1989). Chavas et al. (1983) found that future prices may correspond better to the price formation process for some crops, but future price information is not available for all activities and farmers in all countries (Chavas, 2000), which complicates implementation of future prices in the IFM-CAP model. In contrast, backward-looking expectations can be homogeneously constructed for all IFM-CAP farms using FADN data. We performed a simple econometric test (comparing R^2 and the root mean square error) to compare the use of different types of backward-looking expectations with regard to the supply response of wheat and maize in the Netherlands and France. There were no significant differences in the results obtained using expectations based on past prices at the individual, farm-type or regional level, naive expectations or a (weighted) combination of past prices (adaptive expectations). Therefore, our choice of the use of a simplified formula of adaptive expectations is driven mainly by pragmatic arguments related to data error and availability; using a weighted combination of farm-type or regional-level prices for the last 3 years prior to the base year of 2012 allows smoothing of some potential errors in the data compared with the use of naive prices. In the current version of the model, policy-related price changes (e.g. sugar reform) that could have influenced price expectations for specific crops over the period examined are not considered.

(15) The weights used correspond to those used in the CAPRI expectations module. They correspond to an adaptive expectations model with a correction factor of 0.55. To make sure that the sum of the weights equals 1, the weights on observations in (t – 2) and (t – 3) are slightly higher than would be the case if prices further back in time were included as well. This results in weights of 0.55, 0.30 and 0.15 for observations 1, 2 or 3 years ago, respectively. If only two observations in the past 3 years are available, these weights are adjusted in an ad hoc way, that is, to 0.67 for the more recent observation and 0.33 for the later observation.

Generation of adaptive expected prices at farm type level

After the exclusion of outliers in the price data, for each product the expected price at farm type level is constructed as the weighted average of prices in the past 3 years prior to the base year of 2017 (i.e. 2016, 2015 and 2014). If expected prices for a farm type cannot be constructed, the expected price is calculated at NUTS 2 level instead. If regional-level prices are also missing, expected prices are generated at MS level, or at EU level if needed.

For fodder crops, if information is missing from the FADN database or if the difference between FADN and CAPRI values is greater than ± 25 % at MS level, we use annual prices and yields at national level from the CAPRI database.

More specifically, the following forms of adaptive expectations are constructed at farm type level for each MS, going back three years prior to the base year:

- $p^e_{FTit} = \sum_{n=1}^{5} w_{t-n}\ p_{FT,i,t-n}$ with $w_{t-1} = 0.55, w_{t-2} = 0.30, w_{t-3} = 0.15, w_{t-4} = 0.05, w_{t-5} = 0.01$,

  where *FT* is farm type, *i* is product *t* is year and *p* is the average price for the farm type if data exist in the three successive years;

- $p^e_{FTit} = \sum_{n=1}^{2} w_{t-n}\ p_{FT,i,t-n}$ with $w_{t-1} = 0.67, w_{t-2} = 0.33$,

  if data exist in only two of the three successive years, in which case all the combinations are implemented;

- $p^e_{FTit} = \sum_{t=2014}^{2016} p_{FT,i,t}\,/N$,

  if data exist in only one of the three years previous years (which occurs in only a few cases).

In the final case we also include prices from 2013 and 2012 if available. *N* is the number of years (between 2012 and 2016) with available data.

The regional-level expected prices for each MS are calculated following a similar formula, with the index *FT* (standing for farm type) being replaced by the index *r* (standing for NUTS 2 region).

Generation of the farm-specific deviation from the farm-type or regional average

For each individual farm and for each product, we then calculate the farm-specific deviation from the weighted average price by year for that farm type and take the average over all years between 2012 and 2016 for which price data are available (dev_FT_{fi}). If the farm was observed only in the 2017 sample, the deviation is based on 2017 only.

$$dev\_FT_{fi} = \frac{1}{N} \sum_{t=2012 \text{ to } 2016} (p_{fit} - p_{FTit})$$

In the same way, for each product, the average farm-specific deviation from the regional (NUTS 2) average price is calculated (dev_R_{fi}). For products not produced at the farm in 2012–2016, the farm-specific deviation is set to zero.

Generation of farm-specific expectations

Finally, for each product, the farm-specific expected price is constructed as follows:

- $p^e_{fit} = p^e_{FTit} + dev\_FT_{fi}$, if farm-type expectations are available;
- $p^e_{fit} = p^e_{rit} + dev\_R_{fi}$, if farm-type expectations are missing but expectations are available at the NUTS 2 level;
- $p^e_{fit} = p^e_{MSit}$, if both farm-type and NUTS 2 expectations are missing;
- $p^e_{fit} = p^e_{EUit}$, if, at MS level, expectations are also missing.

For the baseline and simulation scenarios, the formulation of expectations is based on projected prices and yields, to which the individual farm-specific deviation is applied according to the formulas above.

2.2.1.8.Subsidies and payments

We rely on subsidy data available in the FADN database for 2017, which correspond to the 2014–2020 CAP. The data cover both decoupled and coupled CAP payments.

For the case of pillar 1 and pillar 2 payments, data extraction is relatively straightforward (Table 8).

Table 8. Extraction of pillar 1 decoupled and pillar 2 coupled payments from the FADN database

FADN variable	Payment
Pillar 1	
2007–2013 CAP	
SSPSN	Single payment scheme (normal)
SSPSPERMGRS	Single payment scheme (grassland)
SSPSS	Single payment scheme (special entitlements)
2014–2020 CAP	
SBPS	Basic payment scheme
SSAPS	Single payment scheme
SPRCTCLIMENV	Payment for agricultural practices beneficial for the climate and the environment
SANC	Payment for areas with natural constraints
SPS1300	Redistributive payments
SYF	Payment for young farmers
SSFS	Small farmers scheme
Pillar 2	
SINVSUB	Rural development investment subsidies
SA1OTHSUB	Other axis 1 payments
SAEAWSUB	Agri-environment and animal welfare payments
SORGSUB	Organic farming subsidy
SN2000SUB	Natura 2000 payments, excluding forestry
SNHNDMNTSUB	Natural handicap payments to farmers in mountain and other area

FADN variable	Payment
SFRSUB	Payments for forestry, including Natura 2000 payments
SA2OTHSUB	Other axis 2 payments
SRDOTHSUB	Other payments for rural development

However, a problem arises with extracting the share of payment that comes from the EU budget. More specifically, a payment may belong to one or more of the three categories: (i) EU-financed payments, (ii) EU and MS co-financed payments and (iii) MS financing. For pillar 1 decoupled payments, the data clearly pertains to the first category (i.e. decoupled payments are 100% EU financed payments). However, for pillar 2 payments, some payments belong to the co-financed and some others to the MS-financed categories. Since it is impossible to distinguish the share of EU financing and as pillar 2 payments are not endogenously modelled, we sum both financing categories and assume that they are related to the pillar 2 CAP policy.

For coupled payments, the data extraction and the corresponding modelling are much more complex. The implementation of coupled support is MS specific, varies from one production year to another and has payment requirements (e.g. related to minimum and maximum numbers of animals, intensity of production and crop variety). In addition, the FADN coupled support payment categories are too generic to be attributed to the MS-specific coupled support measures (Table 9).

See, for example, excerpts from the 2019 coupled payments for Bulgaria and Czechia.

Bulgaria	1	Milk and milk products	Measure for Coupled Support for Milk Cows	194 923	15 386 818	15 386 818
Bulgaria	11	Milk and milk products	Measure for Coupled Support for Milk Cows under selection control	83 185	18 819 427	18 819 427
Bulgaria	12	Milk and milk products	Measure for Coupled Support for Milk Cows in Mountain Areas (5-9 animals)	7 895	613 550	613 550
Czech Republic	3	Fruit and vegetables	Fruit species with very high labour intensity	5 957	2 647 562	2 647 552
Czech Republic	4	Fruit and vegetables	Fruit species with high labour intensity	4 460	1 056 150	1 056 146

Source: MS communication of coupled support to the Directorate-General for Agriculture and Rural Development for claim year 2019 (European Commission, 2019c).

In Bulgaria, support is given for a small herd of milk cows in a mountain area in one case and for milk cows under a specific animal management mode in another case. In Czechia, support is provided for fruit species with very high labour intensity in one case and for fruit species with high labour intensity in another case.

Consequently, there are two potential approaches for modelling coupled payments:

1. disregard the FADN coupled payment data and instead try to use the payment unit values from the external data source for MS implementation and link the IFM-CAP activities to those;
2. use the FADN coupled payment data for each individual farm, linking IFM-CAP activities to FADN coupled payment categories.

We used the second approach for two reasons. First, IFM-CAP model is not detailed enough to capture the different aspects of MS implementation. For example, there is no information on the plant or animal variety, no spatial information on the fam location, etc.. Second, the FADN-reported coupled payments received by each farm represent implicitly the various implementation details of the MS that the farm belongs to. For example, if a farm has fruit species that indeed use high labour intensity methods and is eligible to receive a payment, this will be reflected in the amount received in one of the

FADN coupled payment categories. If the farm has eligible fruit species but is not using high labour intensity methods, the FADN coupled payment category will not have any amount registered for the farm.

Thus, we map the various categories of FADN coupled payments to IFM-CAP activities, using as auxiliary information the MS implementation details. An example of the mapping for cereals is provided in Table 10. For example in the first line, the 'coupled support for cereals' (SCOPSUBCER) is mapped to durum and soft wheat for Greece and Lituania, to durum wheat for France and Italy and to barley and rye for Latvia.

The steps for mapping FADN coupled payment categories and IFM-CAP activities at MS level are as follows.

1. For each MS, we compute the number of farms that have received each type of FADN coupled payment ($NU_{NUTS0,c}$)
2. For MSs where $NU_{NUTS0,c}$ = 0, the unit value of the coupled payment is zero for all activities ($UV_{NUTS0,c,a} = 0$).
3. For MSs where $NU_{NUTS0,c} > 0$, we try to link the specific payment to one or more IFM-CAP activity. We do this based on the three criteria below:
   (i) the MS implementation details,
   (ii) index A1, for each MS, coupled payment *c* and activity *a*:
   $$A1_{NUTS0,c,a} = \frac{\text{Number of farms with payment } c \text{ and activity } a}{\text{Number of farms with payment } c},$$
   (iii) index A2, for each MS, coupled payment *c* and activity *a*:
   $$A2_{NUTS0,c,a} = \frac{\text{Number of farms with payment } c \text{ and activity } a}{\text{Number of farms with activity } a}.$$

   Considering the values of A1 and A2 (higher values mean a higher probability that an activity is linked to the coupled payment) and considering the implementaion details, we manually assign IFM-CAP activities to coupled payments at MS level.
4. After the mapping of activities and coupled payments in each MS is completed ($MAP_{NUTS0,c,a}$), we estimate the coupled payment unit value at country level (16). For each coupled payment, the estimated unit value multiplied by the activity level should return the payment value reported in the FADN database:
   $$\sum_{f,a}\left(UV_{NUTS0,c} \cdot x_{f,a}\right) = BUDGET_c \ \forall MAP_{NUTS0,c,a}$$
   where *f* are the farms of a NUTS 0, *a* represents the activities connected to payment *c*, and $UV_{NUTS0,c}$ is the unit value of coupled payment *c*. Note that in this way we asume that, for coupled payments that may be granted for more than one activity, the unit payment is equal for all connected activities.

In order to include the specifications of the maximum areas eligible for coupled payments that are often found in MS implementations, we set the observed activity level of a farm as the maximum activity for which the coupled payment can be received. The farm can attain a higher activity level than this but without receiving a coupled payment.

---

(16) We assume that the coupled payment unit value at country level is the same for all farms that receive the payment; this assumption is in line with the way that coupled payments are implemented in CAP.

Table 9. FADN coupled payment variables

FADN code	FADN coupled payment category
SANIMNODEFSUB	Animals, not defined
SANIMSUBDPOTH	Animal subsidy: other coupled payments not mentioned elsewhere
SARACRPNODEFSUB	Arable crops, not defined
SCATLSUB	Cattle, not defined
SCITRPLTSSUB	Citrus plantations
SCOPSUBOIL	COP subsidy: oilseeds
SCOPSUBPROT	COP subsidy: protein crops
SBEFSUB	Coupled support, animals: beef and veal
SDAIRSUB	Coupled support, animals: dairy
SPIGPLTRSUB	Coupled support, animals: pigs and poultry
SSHEPGTSUB	Coupled support, animals: sheep and goats
SCOPSUBCER	Coupled support, cereals, EU financed
SCRPPERMNDEFSUB	Coupled support: permanent crops not mentioned elsewhere
SSCOTNSUB	Crop-specific payment for cotton subsidy
SDRYFODSUB	Dried fodder subsidy
SFLWLNDSUB	Fallow land
SCRPINDSUBFLAX	Industrial crops subsidy: flax
SCRPINDSUBLEG	Industrial crops subsidy: grain legumes
SCRPINDSUBHEMP	Industrial crops subsidy: hemp
SCRPINDSUBHOP	Industrial crops subsidy: hops
SCRPINDSUBRICE	Industrial crops subsidy: rice
SCRPINDSUBSUGCN	Industrial crops subsidy: sugar cane
SNRPCOTNSUB	National restructuring programme for the cotton sector subsidy
SPLTSOILSUB	Olive plantations
SCRPINDOTHSUB	Other industrial crops subsidy
SCRPPERMSUBBERRY	Permanent crops subsidy: berries
SCRPPERMSUBNUT	Permanent crops subsidy: nuts
SPERMGRSSUB	Permanent grassland
SPMSFSUB	Pome and stone fruit
SPOTSUB	Potatoes
SSSEDSUB	Seed production subsidy
SSUGBTSUB	Sugar beet
SVEGSUB	Vegetables
SPLTVINSUB	Vineyards

NB: COP refer to ‘cereals, oilseeds and protein crops’.

Table 10. Excerpt from the correspondence between FADN coupled payment codes and IFM-CAP activities

Coupled payment	Financed	Bulgaria	Czechia	Ireland	Greece	Spain	France	Croatia	Italy	Latvia	Lithuania	Luxembourg	Hungary	Poland	Portugal
SCOPSUBCER	EU				DWHE, SWHE		DWHE		DWHE	BARL, RYEM	SWHE, DWHE				
SCOPSUBCER	Co														
SCOPSUBCER	MS														
SCOPSUBOIL	EU					SUNF, RAPE, OOIL	SOYA		SOYA, SUNF, RAPE	RAPE					
SCOPSUBOIL	Co														
SCOPSUBOIL	MS														
SCOPSUBPROT	EU	PULS, OFAR, SOYA	OFAR		PULS, OFAR	PULS	PULS		OFAR, PULS	PULS, OFAR	PULS	PULS	OFAR	PULS, OFAR	
SCOPSUBPROT	Co							OFAR							
SCOPSUBPROT	MS			PULS											
SARACRPNODEFSUB	EU										OFAR				PGRA
SNRPCOTNSUB	EU					TEXT									
SSCOTNSUB	EU	TEXT			TEXT	TEXT									
SSCOTNSUB	MS					TEXT									

NB: Co, co-financed.

2.2.1.9.Input unit costs

Activity-specific unit input costs enter the expected utility maximization of farmers, expressed by equation (2), through gross margins. However, FADN collects the monetary value of inputs at the farm level without distributing them over specific farm activities. Therefore, to parametrize the farmers' utility function, activity-specific unit input costs need to be estimated from available FADN information through a process of statistical allocation.

Four variable input categories are considered in the model: seeds, fertilisers, plant protection and other specific inputs. The definition of these input categories is in Table 11.

Table 11. Key for allocating input costs developed by DG AGRI

Cost item	IFM-CAP code	FADN code
Seeds and seedlings	SEED	SE290 (home-gown) + F72 (purchased)
Fertiliser	NITF	SE295
Crop protection	PLAP	SE300
Other crop specific costs	CSPE	SE305

We use FADN farm-level data of the year 2017, i.e. the base year, to perform this statistical allocation and obtain activity-specific, farm-level unit input costs for the four input categories considered. Input-output coefficients are estimated separately for all farming types[17] included in the FADN at NUTS2 level using the High Posterior Density (HPD) estimation proposed by Heckelei et al. (2005, 2008). This approach requires assumptions about the production technology, the specification of the HPD objective function, and the definition of prior information. Post-estimation corrections are then applied to the estimated input-output coefficients to ensure that the sum of estimated costs equals the reported FADN costs.

Literature Review

There is considerable literature on the allocation of farm-level input costs across farm activities. This literature have evolved over time and used a variety of statistical approaches, from linear regression models to more sophisticated Bayesian estimation methods. Early studies in the EU have either use linear programming or regression approaches (Ray, 1985; Errington, 1989). However, results were often considered unacceptable due to corner solutions or zero values, or due to the non-negativity of the estimated coefficients (Louhichi et al., 2012; Moxey and Tiffin, 1994; Mindmore, 1990).

Thus, new approaches started to emerge based on statistical methods capable of easily incorporating prior knowledge on the parameters into the estimation framework and effectively constrain coefficients' estimates (Moxey and Tiffin, 1994). Lance and Miller (1998a,b) and Leon et al. (1999) proposed the use of Generalized Maximum Entropy (GME) (Jaynes, 1957a,b; Golan et al., 1996; Paris and Howitt, 1998) to the estimation of activity specific input costs incorporating external prior knowledge on the parameters. In the presence of informative priors (Lance and Miller, 1998a; Paris and Caputo, 2001), these entropy-based methods were both able to perform well in recovering activity-

---

(17) We use the TF14 FADN classification to define farming types.

specific costs (Louhichi et al., 2012), and suitable for dealing with the issues of singularities, non-negativity constraints, and zero-observations (Leon et al., 1999).

Still, to tackle some of the limitations of entropy-based methods, Heckelei et al. (2005, 2008) proposed the High Posterior Density (HPD) estimation, a fully Bayesian approach to the estimation of underdetermined systems of equations. HPD was able to incorporate prior information more transparently than entropy-based approaches and was computationally simpler in terms of number of equations and parameters to be estimated (Louhichi et al., 2012; Heckelei et al., 2008). Louhichi et al. (2012), using informative priors for activity-specific input costs on a sample of French farms, compared the performance of GME and HPD methods and concluded that the two methods were able to produce equivalent results.

Leontief technology specification for intermediate inputs

For the estimation of unit input costs, we assume a linear Leontief technology for intermediate inputs (i.e. different inputs increase proportionally to each other and increase linearly with production activity levels). The linear technology in input costs and output values is defined as follows:

$$\mathbf{z} = \mathbf{A\theta} + \mathbf{u} \tag{44}$$

where $\mathbf{z}$ is the (FK x 1) vector of input costs, $\mathbf{A}$ is the (FK x KN) block-diagonal matrix of output values, $\boldsymbol{\theta}$ is the (KN x 1) vector of input-output coefficients, and $\mathbf{u}$ is the (FK x 1) error vector. This relationship can be expressed by farm and input category as follows:

$$z_{f,k} = \sum_i a_{f,i}\theta_{k,i} + u_{f,k} \tag{45}$$

where $z_{f,k}$ is the cost of input category *k* (k=1,..,K) for farm *f* (f=1,…,F), $a_{f,i}$ is the value of activity *i* (i=1,…,N) for farm *f*, $\theta_{k,i}$ is the input-output coefficients, and $u_{f,k}$ is the error term. It is assumed that farms within the same NUTS2 region and the same farming type have a common technology and therefore the same input–output coefficients $\theta_{k,i}$.

This assumption on the functional form of the technology may be considered restrictive. However, this form of input demand equation has been assumed widely in the literature (e.g. Léon et al., 1999; Kleinhanss, 2011) and, it is a convenient way to both include behavioural assumptions in the estimation framework and to obtain unit input costs per hectare using available data.

This linear technology embeds the behavioural assumption that total revenues equal total costs at the farm level. Following Léon et al. (1999), this is achieved by introducing a residual input category 'value added' defined as the difference between the total revenues and the sum of all variable input costs considered[18], and by imposing the following restriction on the input-output coefficients for each activity *i*:

$$\sum_i \theta_{k,i} = 1 \tag{46}$$

---

(18) Similarly to other input categories, value added is restricted to be positive, assuming that, for each type of output *i* averaged (across all farms), total cost cannot exceed total revenue.

Moreover, this linear technology allows to derive a farm-specific, activity-specific input cost per activity level using the following formula:

$$c^*{}_{f,i,k} = \theta_{k,i} \frac{a_{f,i}}{x_{f,i}} \tag{47}$$

where $c^*{}_{f,i,k}$ is the estimated cost per activity level of input *k* used by farm *f* to produce output *i*, $\theta_{k,i}$ is the estimated input-output coefficient, and $a_{f,i}/x_{f,i}$ is the unit value per activity level of activity *i* (Kleinhanss, 2011).

High posterior density estimation

The HPD approach minimises the normalised least square deviation between the estimated input–output coefficients and the prior information subject to technology constraints. Model parameters are treated as stochastic outcomes. In this context, the method distinguishes between the prior density $\mathbf{p}(\boldsymbol{\theta})$, which summarises a priori information on parameters, and the likelihood function $\mathbf{L}(\boldsymbol{\theta}|\mathbf{A})$, which represents information obtained from the data in conjunction with the assumed model. The combination of the prior density and the likelihood function results in a posterior density (e.g. Zellner, 1971, p. 14), which can be expressed as:

$$\mathbf{z}(\boldsymbol{\theta}|\mathbf{A}) \propto \mathbf{p}(\boldsymbol{\theta})\, \mathbf{L}(\boldsymbol{\theta}|\mathbf{A}) \tag{48}$$

where $\mathbf{z}$ denotes posterior density, ∝ is the proportionality, $\boldsymbol{\theta}$ are the parameters to be estimated and $\mathbf{A}$ is the matrix of output values. This approach is extensively discussed in Heckelei et al. (2008). This leads to the following estimation problem:

$$\min \mathrm{HPD} = [\boldsymbol{\theta} - \overline{\boldsymbol{\theta}}]'\Sigma^{-1}[\boldsymbol{\theta} - \overline{\boldsymbol{\theta}}]$$

Subject to:

$$\mathbf{z} = \mathbf{A}\boldsymbol{\theta} + \mathbf{u} \tag{49}$$

$$\mathbf{I}'\boldsymbol{\theta} = \mathbf{1}$$

where $\overline{\boldsymbol{\theta}}$ is the (KN x 1) vector of prior values and HPD is the prior density function of the form $\boldsymbol{\theta} \sim \mathrm{N}(\overline{\boldsymbol{\theta}},\Sigma)$. The prior values $\bar{\theta}_{k,i}$ are the mean input–output coefficients by NUTS2 region and farm type. The covariance matrix Σ is set equal to a diagonal matrix with, as elements, twice the standard deviation of the prior input–output coefficients squared, $(2\sigma^h)^2$.

Prior input-output coefficients

The priors of the input-output coefficient are obtained as weighted averages of the farm-level ratios between total input cost for input category *K* and total output value of the relevant productions. For farm *f*, this ratio for input category *k* is represented by:

$$r_{f,k} = \frac{C_{f,k}}{\sum_i a_{f,i}} \tag{50}$$

where $C_{f,k}$ represents the input cost of input category *k* for farm *f*, and $\sum_i a_{f,i}$ represents the sum of the relevant output values of farm *f*. In the case of fertilizers, pesticides, and other specific costs, the relevant productions includes all crop activities. For seeds costs, relevant productions excludes permanent crops whose seeds costs are assumed to equal to zero. This farm-level ratio applies equally to each of the activity of the farm.

The activity-specific prior for each of the input categories at NUTS2-TF14 level are then obtained as weighted averages of these farm-level ratios[19]. The averages are calculated only using farms that belong to the relevant NUTS2-TF14 and using FADN sampling weights for aggregation. The activity-specific prior for a generic NUTS2-TF14 combination is obtained as follows:

$$\bar{\theta}_{k,i} = \frac{\sum_f w_f \ * r_{f,k}}{\sum_f w_f} \tag{51}$$

where index *i* identifies the farm activity and the index *k* identifies the input category, and w_f is the FADN sampling weight of farm *f*[20].

Unit input cost correction

The linear technology imposes equality of total revenues and total costs at the farm level. However, this does not imply that estimated farm-level input costs exactly equal FADN accounting costs for each input category k. To ensure that total estimated input costs equal total accounting costs at the farm-level reported in FADN, we apply a post-estimation correction to the input-output coefficient. For each farm, we allocate the remaining non-distributed costs proportionally across the different activities, leading to a farm-specific corrected input–output coefficient $\tilde{\theta}_{f,i,k}$:

$$\tilde{\theta}_{f,i,k} = \theta_{k,i} \frac{z_{f,k}}{\sum_i a_{f,i} \theta_{k,i}} \tag{52}$$

Based on these corrected coefficients, $\tilde{\theta}_{f,i,k}$, and the value of production per observed activity level, $a_{f,i}/x_{f,i}$, the corrected input costs per activity level of activity *i*, $c_{f,i,k}$, can be computed as follows:

$$c_{f,i,k} = \tilde{\theta}_{f,i,k} \ \frac{a_{f,i}}{x_{f,i}} \tag{53}$$

---

(19) Ratios that exceeds the value of one, i.e., they are not consistent with the behavioral assumptions, are excluded from the calculations.

(20) Because the farms included in the relevant NUTS2-TF14 combination will each have a different mix of activities, and because each farm-level ratio is applied to all the activities of the farm, the prior obtained is also indexed by activity *i*.

2.2.2. Livestock activities

2.2.2.1. Literature review on modelling livestock activities at farm level

Livestock production systems are complex systems composed of biological, economic, environmental, social and behavioural elements. The main components of livestock production systems can be grouped under four main categories: (i) biological processes; (ii) herd demography/dynamic; (iii) livestock–crop interactions that consists of (a) feeding, grazing and nutritional demand and supply, and (b) manure production and application; and (iv) economic behaviour of the farmers or farm managers. Incorporating these four elements in models along with their interactions is crucial in accurately capturing the behaviour of the whole livestock system.

Among these elements, inclusion of herd demography in models, particularly in static optimisation models, is challenging. An important reason for this is that characteristics of the livestock life cycles make the production activities highly interlinked and dynamic processes; changes in one component can affect the other components of the livestock production systems. Moreover, explicit modelling of herd demography and its dynamic requires detailed data and information on various technical and biological parameters of livestock systems that are often not accessible for a broad range of farming systems. Two main types of models that are used in mathematical programming literature and have attempted to incorporate livestock activities and their demography/dynamic are biophysical models and economic models. These model types are briefly introduced and discussed below (Louhichi et al., 2018b).

Biophysical models

Biophysical models usually attempt to identify optimal farm practices by endogenously defining biological parameters such as animal replacement rate, lactation length, slaughter weight, milk yield, etc. Herd dynamic is featured in many biophysical models (e.g. Gartner, 1982; Kristensen, 1992; Koots and Gibson, 1998; Nielsen et al., 2005; Cabrera, 2012; Kalantari et al., 2014). At the core of biophysical models are specific livestock categories for which herd dynamic and optimal management choices are analysed. For example, key parameters determining herd dynamic in dairy cow models are replacement decisions and reproductive performance of the herd. Both parameters are key drivers of how the herd evolves over time and have a significant impact on the productivity and profitability of dairy farming.

The typical and most widely used biophysical models are single-component models that consider only one animal category, such as dairy cows, suckler cows, breeding sows or breeding sheep, whereas other on-farm livestock categories are treated in a simplified way, for example by assuming unlimited supply of replacement heifers and sale of calves after calving in dairy cow models (Nielsen et al., 2005; Cabrera, 2012; Kalantari et al., 2014). Extensions of the single-component models consider multiple livestock categories that are regarded as multiple-component models (Gartner, 1982; Kristensen, 1992; Koots and Gibson 1998). The structure of the herd is endogenous in biophysical models and the model parameters determine the optimal herd demography. Livestock categories are defined by a set of characteristics (e.g. lactation period, milk production level, calving period, weight, etc.) and often define management practices used to identify optimal choices in a particular production system.

Economic models

In contrast to biophysical models, in economic models all or the majority of the parameters (e.g. lactation period, milk production level, calving period, weight, etc.) are exogenously determined and therefore productivities of different animal categories are exogenously defined. Explicit modelling of herd dynamic of individual farms in economic models also poses challenges, as it requires detailed information on various technical parameters of livestock systems. The central element in many economic models that incorporate livestock dynamic is animals' reproductive characteristics. For example, replacement rate is the key parameter that determines the herd dynamic of dairy and breeding suckler cow systems. A common assumption in many economic models is that cows are replaced by heifers raised on the farm (e.g. Lelyon et al., 2010). Some models also allow for purchase of replacement heifers alongside their own raised heifers (Veysset et al., 2005). Other livestock categories are derived from the number of cows based on reproductive performances of the herd. This is defined either exogenously by parameters such as calves per cow (i.e. shares) or cow replacement rate or endogenously by management practices. Another important management decision determining the herd dynamic is the choice of sales and purchases of different livestock categories.

This modelling approach implies that the demography of reproductive animals (e.g. dairy or breeding suckler cows) is fully endogenous, whereas the rest of the livestock activities can vary from being fully exogenous to fully endogenous. The herd composition and size are co-determined by the herd reproductive performance (e.g. cow replacement rate, calf per cow, etc.) and animal sale and purchase decisions. One possible extreme situation is when sale and purchase activities are not modelled, implying that all non-cow livestock categories (i.e. demography) are exogenous and are determined exclusively by the number of adult cows observed by the herd reproductive performance (Thorne et al., 2009). The other extreme situation is when sale and purchase activities are allowed for all livestock categories. In this case, the herd composition is fully endogenous and depends on the relative return of various livestock activities. Between these two extreme situations, there are many possibilities for partially endogenous herd dynamics. In fact, most of the applied economic models consider partially endogenous herd dynamics. The type of livestock farm modelled largely defines the behaviour of herd dynamic and possible livestock activities. The static characteristics of many economic models reflect the steady-state equilibrium of the modelled farms. The equilibrium solutions reflect the full adjustments of herd demography to the simulated economic and policy shocks.

Economic models can be categorised under normative and positive approaches. *Normative models* usually refer to linear programming models typically result in a wide divergence between the simulated results of considered activities, including livestock numbers and the on-farm observed values. The inclusion of a risk term may improve the model performance, but still may not fully reproduce the actually observed activity level. Despite this weakness, there are many applications of normative models, including livestock modelling. The main focus of these models is on analysing the difference between the simulated scenarios rather than on the accuracy of reproducing the observed livestock activities in the baseline simulations.

From the methodological point of view of incorporating livestock herd dynamic in mathematical programming models, normative models tend to explicitly represent the herd dynamic. They often explicitly represent intergenerational dependences or links and the flows between different livestock categories as well as herd reproduction parameters, such as cow replacement rate, that are key drivers of livestock herd dynamism (e.g. Nicholson et al., 1994; Ramsden et al., 1999; Visagie and Ghebretsadik, 2005; Ducros et al., 2005; Veysset et al., 2005; Havlík et al., 2006; Crosson et al., 2006; Acs et al., 2010; Lelyon et al., 2011; Jones and Salter, 2013). The main characteristics of the normative models used in the cited studies are summarised In Annex A.

Regarding *positive models*, PMP has been the preferred method of many scientists and policy makers in calibrating models that generate the actually observed activities and outcomes for farmers. In other words, PMP assumes that farmers’ choice of combination of activities is optimal. This provides a reliable tool to simulate policy scenarios and predict future changes. Although the PMP approach does not require an explicit representation of the herd dynamic or management practices regarding the sale and purchase activities for different animal categories, the use PMP for livestock activities may not be as straightforward as for other activities, such as crops. An important reason for this is the intergenerational dependences or the linkages between adult animals and their youngstock, which may be retained to replace the breeding animals or may be sold in the market. As stated earlier, most of the livestock systems and related herd demography follow a cyclical pattern that implies the importance of dynamism in these systems. Incorporating this dynamism into static PMP models such as IFM-CAP, therefore, requires certain considerations and assumptions. The main characteristics of the positive models reviewed for the purpose of further development of IFM-CAP model are summarised Annex A.

2.2.2.2.Modelled livestock activities in IFM-CAP

In the current version of the IFM-CAP, as in the approach used in modelling crop activities, PMP terms have been estimated for each livestock category, that is, adult animals and their youngstock separately, without explicitly modelling intergenerational dependences. All livestock activities, therefore, are endogenously determined by the model. The advantage of this approach, compared with the earlier version of the IFM-CAP (where the numbers of young animals were determined by shares of adult animals), is that the number of youngstock is not fully dependent on the number adult animals; this can, therefore, represent real farm management practices. A potential disadvantage, however, may be that the simulated effects for the livestock sector will depend mainly on the sale and purchase prices of animal outputs and this may not reflect the livestock management systems actually practised by farmers. As a result, it may be possible for the model to react differently from the patterns observed in reality, for example drastically reducing the number of young animals. We envisage that this is not likely to be the case, but, if this behaviour were observed, then an additional constraint linking adult and young animals will be added and their levels (i.e. shares of youngstock) will be introduced exogenously to adults. In the current version of the model, sale and purchase activities of various livestock categories are not explicitly modelled and therefore these are implicitly captured by PMP terms for each activity (i.e. only animal products can be sold).

2.2.2.3.Definition of livestock activities and outputs

Four categories of livestock activities are modelled in IFM-CAP: cattle (dairy and beef), pigs, small ruminants (sheep and goats) and other animals. For certain categories (e.g. cattle and small ruminants) two different systems can be considered: raising and fattening systems.

FADN data are used to identify the predominant livestock activities across regions of the EU. Table 12 describes the set of livestock activities included in IFM-CAP and the rules used for extracting their numbers (i.e. activity level) by animal category from FADN (Table D) for the base year period. The set of livestock outputs modelled in IFM-CAP are the following: beef, cow milk (for feeding and sales), milk from sheep and goats (for feeding and sales), meat from sheep and goats, poultry meat, pork and young animals (male and female calves and piglets). Table 13 presents the list of livestock outputs and the rules used to define their values. Figure 4 and Figure 5 present graphically the relation between each animal activity and outputs. Table 14 compares the total EU livestock production between IFM-CAP and Eurostat and shows that the IFM-CAP production quantities are close to those reported in Eurostat.

Table 12. Livestock activities

Livestock activity name	**IFM-CAP activity**	**FADN code**	
		2013 and before	**2014 and after**
Cattle			
Dairy cows	DCOW	30AV	LCOWBUFDAIR
Heifers breeding	HEIR	28AV + MIN(26AV,28AV)	LHEIFFAT+MAX(0,LBOV1_2F-LHEIFBRE)
Raising male calves	CAMR	MAX(0,(24AV-28AV))	MAX(0,LBOV0-LHEIFBRE)
Raising female calves	CAFR	MIN(28AV,24AV)	MIN(LHEIFBRE,LBOV0)

Other cows	SCOW	32AV	LCOWOTH
Heifers fattening	HEIF	29AV + MAX (0,26AV-28AV)	LHEIFFAT+MAX(0,LBOV1_2F-LHEIFBRE)
Male adult cattle	BULF	25AV + 27AV	LBOV1_2M+LBOV2
Fattening male calves	CAMF	0.5*23AV	0.5*LBOVFAT
Fattening female calves	CAFF	0.5*23AV	0.5*LBOVFAT
Pigs			
Pig fattening	PIGF	45AV + 46AV	LPIGFAT + LPIGOTH
Pig breeding	SOWS	44AV	LSOWBRE
Goats and sheep			
Milk ewes and goats	SHGM	38AV + 40AV	LEWEBRE+ LGOATBRE
Sheep and goat fattening	SHGF	39AV + 41AV	LSHEPOTH + LGOATOTH
Other animals			
Laying hens	HENS	48AV/1 000	LHENSLAY
Poultry fattening	POUF	(47AV + 49AV)/1 000	LPLTRBROYL + LPLTROTH
Other animals	OANI	50AV+ 22AV	LRABBRE+LEQD

Note: For 2014 and after, LBOV0 and LBOVFAT are computed from LBOV1, as they are discontinued

Table 13. Definition of outputs for livestock activities

Output / activity	Activity	Output	Extraction rule
Female/Male calves produced/dairy cow	DCOW	YCAF,YCAM	0.5*LBOV1_SN*(DCOW_AN/(DCOW_AN+SCOW_AN))/DCOW_AN
Female/Male calves produced/suckler cow	SCOW	YCAF,YCAM	0.5*LBOV1_SN*(SCOW_AN/(DCOW_AN+SCOW_AN))/SCOW_AN
Beef produced/dairy cow	DCOW	BEEF	DCOW_SN*CW/DCOW_AN
Beef produced/suckler cow	SCOW	BEEF	SCOW_SN*CW/SCOW_AN
Beef produced/bull	BULF	BEEF	BULF_SN*CW/BULF_AN
Beef produced/heifer fattening	HEIF	BEEF	(HEIF_SN*CW/HEIF_AN) + (max(0,LBOV1_2F_SN-LHEIFBRE_SN)*CW/max(0,LBOV1_2F_AN-LHEIFBRE_AN))
Beef produced/calf fattening	CAMF/CAFF	BEEF	CAMF_SN*CW/CAMF_AN
Milk for sale produced/dairy cow	DCOW	COMI	(PMLKCOW_PRQ+PMLKBUF_PRQ)/DCOW_AN
Milk for feeding produced/dairy cow	DCOW	COMF	MC*(PMLKCOW_PRQ+PMLKBUF_PRQ)/DCOW_AN
Milk for feeding produced/suckler cow	SCOW	COMF	Θ*(PMLKCOW_PRQ+PMLKBUF_PRQ)/SCOW_AN
Piglets produced/sow	SOWS	YPIG	LPIGLET_SN/SOWS_AN
Pork produced/sow	SOWS	PORK	SOWS_SN*CW/SOWS_AN
Pork produced/pig fattening	PIGF	PORK	PIGF_SN*CW/PIGF_AN
Meat produced/sheep and goats for milk production	SHGM	SGMT	SHGM_SN*CW/SHGM_AN
Meat produced/sheep and goats for fattening	SHGF	SGMT	SHGF_SN*CW/SHGF_AN
Milk for sale produced/sheep and goats for milk production	SHGM	SGMI	(PMLKSHEP_PRQ+PMLKGOAT_PRQ)/SHGF_AN
Milk for feeding produced/sheep and goats	SHGM	SGMF	MC*(PMLKSHEP_PRQ+PMLKGOAT_PRQ)/SHGF_AN
Poultry meat produced/poultry fattening	POUF	POUM	POUF_SN*CW/POUF_AN
Eggs / Laying hens	HENS	EGGS	0.001*PEGGC_PRQ*57/HENS_AN

Notes: -Letter codes (e.g. DCOW, BULF, LPIGLET, etc.) denote FADN and IFM-CAP animal categories as provided in Table 10.

- 'SN' is the sales number, as reported in FADN, of the corresponding animal category. E.g. SCOW_SN is the sales number of Suckler cows
- 'PMLKCOW', 'PMLKBUF','PMLKSHEP' and 'PMLKGOAT' refers to milk produced by cows, buffalos, sheep and goats respectively. 'PRQ' is the production quantity, as reported in FADN for the corresponding animal product. For example, 'PMLKSHEP_PRQ' is the milk quantity produced from sheep.
- CW: carcass weight at MS level derived from Eurostat;
- MC: share COMF/COMI and SGMF/SGMI at NUTS2 level from CAPRI.
- Θ is an estimated coefficient of the production of a SCOW in relation to that of DCOW. We estimate it as the ratio of the median production of SCOW to DCOW in NUTS0 level.

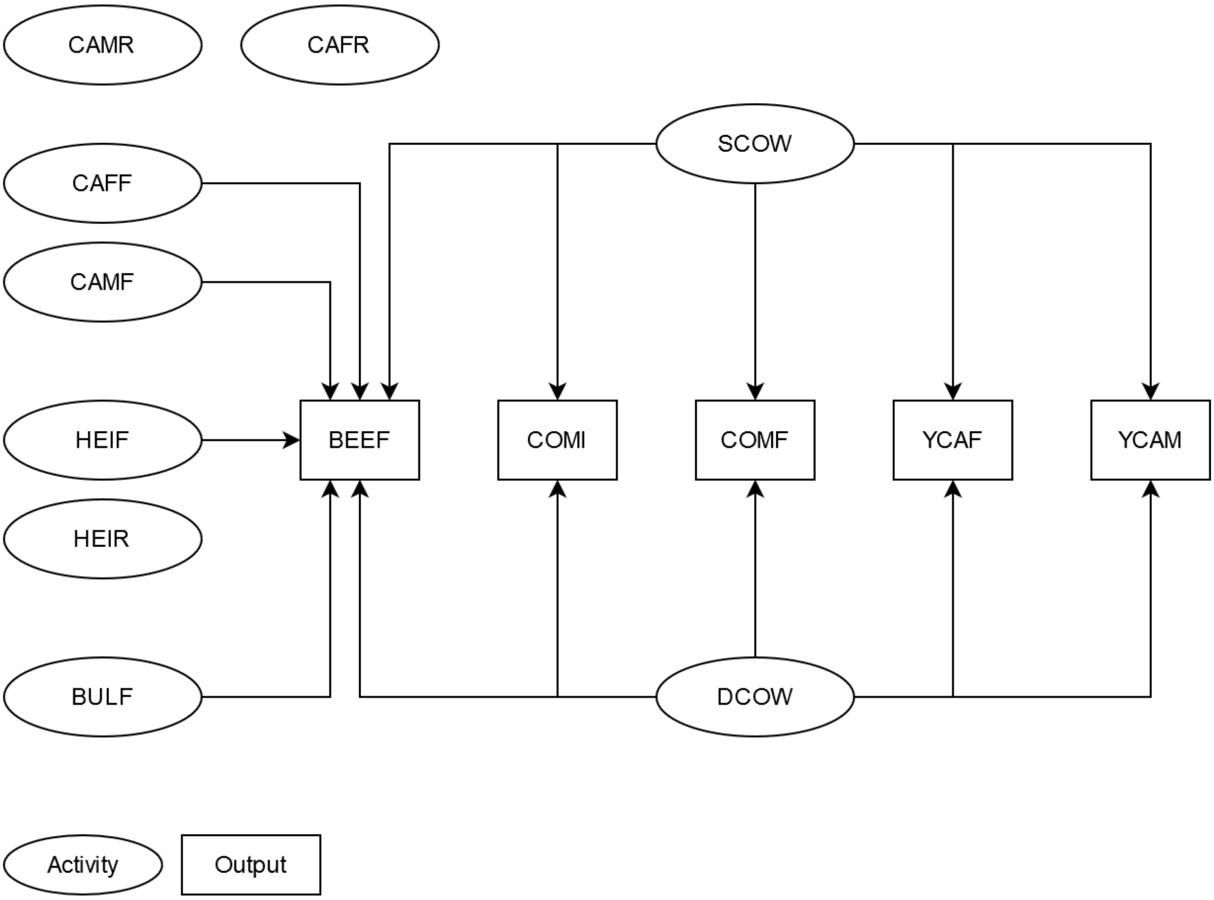


Figure 4. Relations of livestock activities and output for dairy and cattle sector

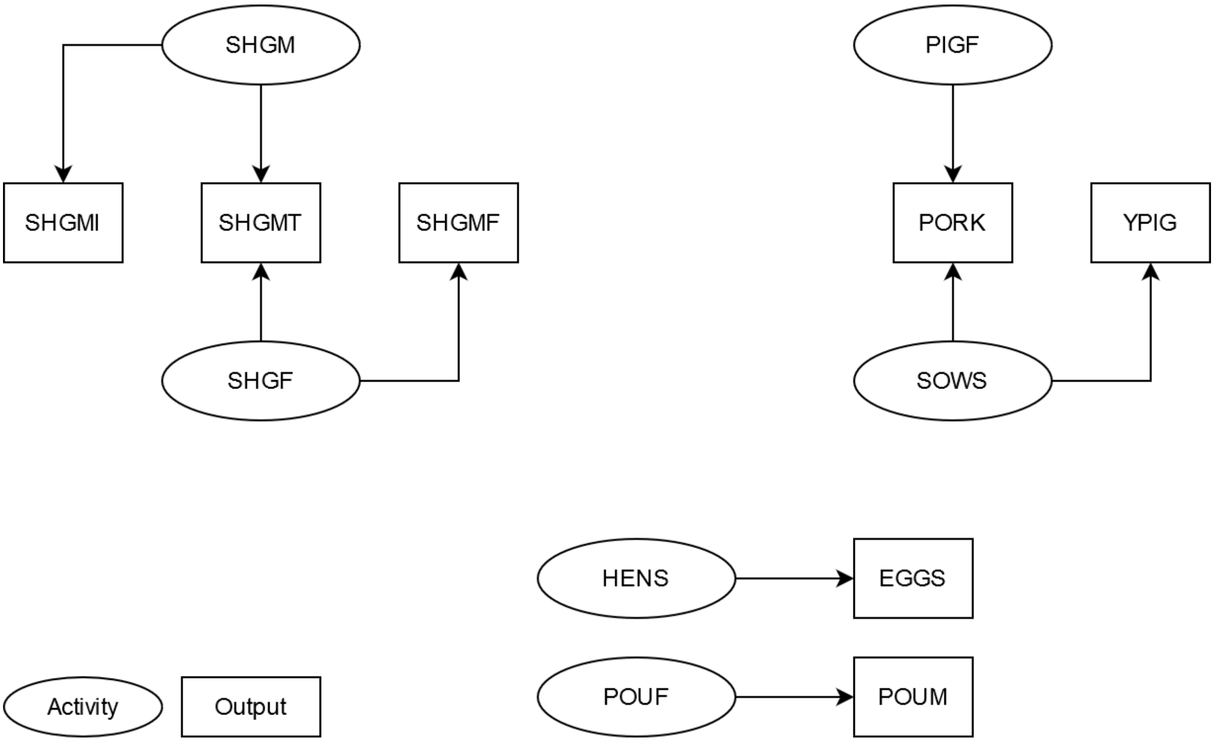


Figure 5. Relations of livestock activities and output for sheep and goats and granivore sector

Table 14. Comparison of total EU livestock production between IFM-CAP and Eurostat

		Eurostat	IFM-CAP	
Milk production	Cows	143100	COMI	150128
	Sheeps and goats	5500	SGMI	5603
Meat	Beef carcass (bovine meat)	6900	BEEF	6737
	Sheep and goat meat carcass,	500	SGMT	895
	Pork carcass	22000	PORK	20168
	Poultry meat carcass	12000	POUM	8339
	Eggs weight	6700	EGGS	3826

Note: For 2014 and after, LBOV0 and LBOVFAT are computed from LBOV1, as they are discontinued

2.2.2.4.Feed requirements and allocation of feed resources

Most of the mathematical programming models applied in the literature represent the interactions between crop and animal activities through feed supply and demand balances. The feed balances guarantee that animal-specific nutrient demands (requirements) are met from internally produced or purchased feed (e.g. forage, grains, concentrates). Balancing feed supply (availability) and demand (requirements) is done through nutrient values. The physical quantities of feed, as well as the animal feed requirements, are expressed in nutrient values such as energy, dry matter, protein, fibre and essential amino acids such as lysine. The feed demand depends on the feed requirement per animal and the number of animals; the feed supply depends on the nutrient content of each feed component and its available (on-farm produced and/or purchased) quantity (e.g., De Cara and Jayet, 2000; Alford et al., 2004; De Cara et al., 2005; Crosson et al., 2006; Britz and Witzke, 2014; Heckelei et al., 2012; Arata et al., 2013).

Following the literature, we have developed a specific module within IFM-CAP to endogenously match feed availability and feed requirements for the livestock in each farm. The structure of this feed module is depicted in **Figure 6**. It consists of two main components: feed availability and feed requirements. Feed availability represents the supply of different types of feed, such as grass, fodder, cereals and concentrates. The list of individual feed products considered in IFM-CAP, as well as their corresponding feed category, is reported in Table 15. Feed requirements depend on livestock type (species and purpose) and are determined by, among other things, animal productivity (e.g. weight, milk production), duration of animal raising and keeping activities and farm herd size.

Feed availability and feed requirements are then converted into nutrient values and balanced by animal category at farm level. Table 16 lists the set of nutrients considered in IFM-CAP. Feed availability has to meet the protein (CRPR) and energy (ENNE) needs of each animal category (i.e. supply equals demand for CRPR and ENNE). In addition, for certain animal categories, additional minimum and/or maximum intake constraints are introduced. Maximum intake constraints concern dry matter (DRMX) and fibre (FIDI, FICO, FICT, FISM, FISF, FILG), while the minimum constraints are set for dry matter (DRMN) and lysine (LISI).

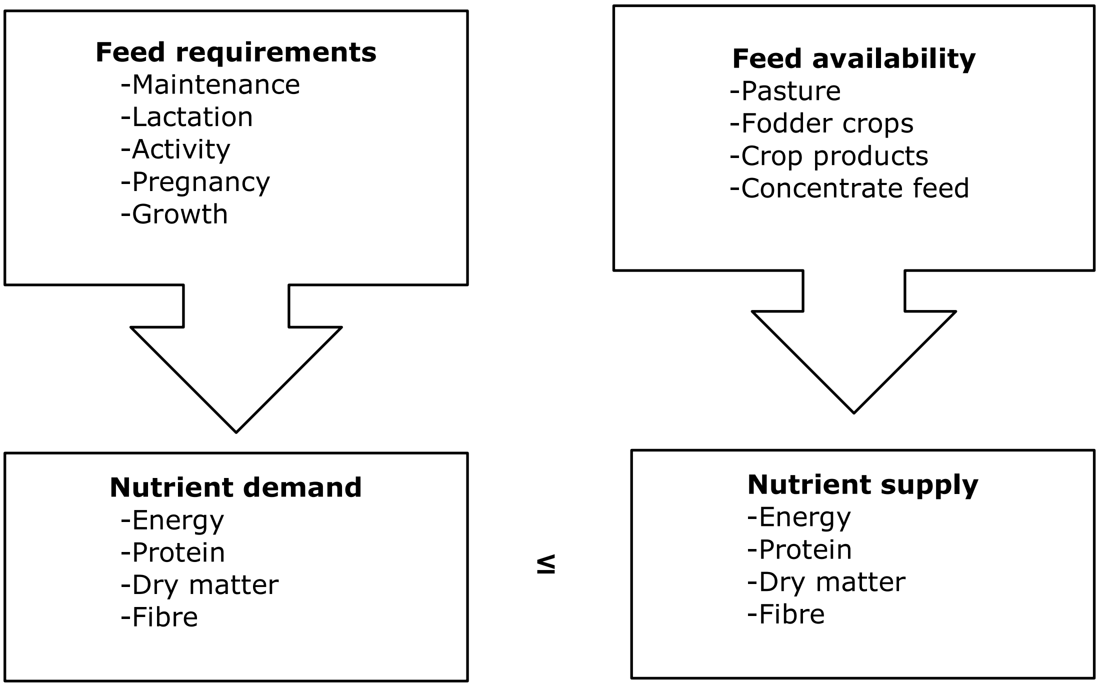


Figure 6. Feed module in IFM-CAP

Moreover, minimum and maximum thresholds of feed groups (e.g. cereals, fodder) in animal diets expressed in dry matter intake are defined for each animal category. The thresholds ensure that the allocation of feed does not result in overuse or underuse of certain feed groups and matches animals’ physiological requirements.

Table 15. List of feed products in IFM-CAP

Name of feed	Feed code	Name of aggregated feed	Aggregated feed grouping	Feed category		
				Fodder feed	**Concentrate feed**	**Other feed**
Soya cake	SOYC	Feed rich protein	FPRO		X	
Olive cake	RAPC				X	
Sunflower cake	SUNC				X	
Soya oil	SOYO				X	
Distillers dried grains with solubles	DDGS				X	
Pulses	PULS				X	
Fodder maize	MAIF	Fodder maize	FMAI	X		
Permanent grassland	PGRA	Grass	FGRA	X		
Rough grazing	RGRA	Grass	FGRA	X		
Other fodder on arable land	OFAR	Other fodder	FOFA	X		
Soft wheat	SWHE	Feed cereals	FCER		X	
Durum wheat	DWHE				X	
Rye and meslin	RYEM				X	
Barley	BARL				X	
Oats	OATS				X	
Grain maize	MAIZ				X	
Other cereals	OCER				X	
Sheep and goat milk feeding	SGMF	Sheep and Goat Milk for feeding	FSGM			X
Milk for feeding	COMF	Cow Milk for feeding	FCOM			X
Molasse	MOLA	Feed rich energy	FENE		X	
Straw	STRA	Straw	FSTR	X		
Table grapes	TAGR	Feed other	FOTH			X
Table olives	TABO					X
Olive oil	OLIV					X
Other fruits	OFRU					X
Citrus fruits	CITR					X
Apples and pears	APPL					X
Potato	POTA					X
Other oil	OOIL					X
Sunflower	SUNF					X
Soya	SOYA					X
Rapeseed	RAPE					X

FADN data do not contain all the information needed to parameterise the feed module in IFM-CAP. FADN contains farm aggregated economic data on feed availability and costs. However, disaggregated feed data by activity (e.g. feed use by animal category), prices and yields of certain feed crops, nutrient content of feed and animal requirements are not available in FADN. To fill this gap, we supplement FADN data with external sources such as other official statistical sources (e.g. Eurostat), scientific literature or other models (e.g. CAPRI). The external sources utilised in the current version of the model are documented below. The disadvantage of using external data is that they may be inconsistent with FADN data and may provide unreliable information, in particular when MS or regional data are used at farm level. To reduce this problem, we employ the HPD approach to estimate farm-level data and external data are used only as prior information in the estimation approach. The estimation approach combines these different data sources by taking into consideration the minimisation of deviation of estimated data values from the available prior information, the minimisation of feed costs (this component was included in the HPD objective function), balancing between feed requirements and availability, and data constraints to ensure that the sum of activity feed costs is as close as possible to the aggregated cost values reported in FADN (see section 2.2.2).

Feed availability is represented by the physical quantity of feed, as well as its nutrient content and costs (i.e. prices, time, quantity). Farms can use feed produced on-farm or purchased on the market. The purchased feed is estimated based on farm level feed costs and feed prices in the HPD approach. The on-farm production of feed during the base year is obtained from FADN. For straw, we use the residue-to-crop ratio (RCR) as a function of grain yield (SWHE, DWHE, RYEM, BARL and OATS) to obtain straw production (see section 2.2.1.6). The data on yields of fodder crops (OFAR, GRAS, ROOF and MAIF) are not fully reported in FADN. We give priority to FADN data when available. We use CAPRI data only if information is not available in FADN or if the yield difference between FADN and CAPRI data is greater than 25 %. For grassland production is missing in the vast majority of cases and impute it from CAPRI (see section 2.2.1.6).

Feed prices are derived from FADN, except for fodder and concentrates, for which data come from the CAPRI.

Table 16. List of nutrients in IFM-CAP

Nutrient	Description (unit)
ENNE	Net energy (MJ/kg)
ENMR	Metabolisable energy ruminants (MJ/kg)
ENMC	Metabolisable energy chicken (MJ/kg)
ENMH	Metabolisable energy horses (MJ/kg)
ENMP	Metabolisable energy pigs (MJ/kg)
DRMN	Minimum dry matter (kg/kg)
DRMX	Maximum dry matter (kg/kg)
CRPR	Crude protein (kg/kg)
LISI	Lysine (kg/kg)
FIDI	Fibre (kg/100 kg)
FICO	Fibre dairy cows (fill unit system)
FICT	Fibre cattle (fill unit system)
FISM	Fibre sheep and goat milk (fill unit system)
FISF	Fibre sheep and goat fattening (fill unit system)
FILG	Fibre long

For the *nutrient content of feed*, we rely exclusively on external sources, as this type of data is not available in FADN. In the literature, nutrient values (e.g. regional averages) are most often taken from technical books and/or are based on expert knowledge. For example, in their FADN-based farm model for Emilia-Romagna in Italy, Arata et al. (2013) collected nutrient content data from regional rule books and from personal communications from a local animal nutritionist. Similarly, De Cara et al. (2005) and De Cara and Jayet (2000) extracted nutrient data from the literature (Jarrige, 1988; Jarrige, 1989) and combined this with expert knowledge for their FADN-based representative EU farm model. The CAPRI model relies on nutrient contents from the Institut National de la Recherche Agronomique (INRA) and the SPEL/EU-Base Model (Wolf, 1995). In the IFM-CAP model, we use the nutrient content of feed at MS level from CAPRI.

The *feed requirements* are critical for an accurate representation of crop–animal interactions. They describe how much nutrients (i.e. energy, crude protein, fibre and dry matter) each animal activity requires for its main biological functions. The full set of underlying data needed to calculate feed requirements including nutrients and physical quantities is not available in FADN. To overcome this lack of data, we use the requirement functions combined with FADN and external data, as is usual in the literature (e.g. De Cara and Jayet, 2000; De Cara et al., 2005; Arata et al., 2013). More precisely, we used the requirement functions as implemented in CAPRI (Nasuelli et al., 1997; IPCC, 2006; Britz and Witzke, 2014) and from other sources (e.g. GfE, 2006; LfL, 2014, NRC, 1994) to calculate an approximate value of animal requirements (see Annex B). These values are then used as prior information to estimate the final nutrient requirements by animal category, which guarantees that feed availability equals feed requirement at farm level in both physical and nutrient terms.

The prior values of animal requirements are determined by predefined coefficients and animal productivity parameters. The predefined coefficients are extracted from FADN data or other sources (e.g. CAPRI; Eurostat), or are calculated based on the combination of both sources. However, the predefined coefficients and animal productivity parameters (e.g. milk fat content, animal live weight, start/end of day of production process) may depart significantly from the actual values observed at farm level. To account for this uncertainty, we consider variation of these coefficients and parameters (e.g. by using the standard deviation) to derive lower and upper bounds of animal requirements. The lower and upper bounds demarcate the most likely interval within which the actual values of animal requirements lie.

The main productivity parameters that determine the nutrient requirements include live weight of animal, raising/fattening period, milk and/or meat production, daily animal growth rate; fat content of milk, and start and end date of animal raising/fattening process. These values are obtained from FADN, calculated based on the combination of FADN data and other sources (e.g. CAPRI; Eurostat) or are assumed to be as in CAPRI. For example, the fat content of milk is extracted from Eurostat, whereas the live animal weight of dairy cows is obtained by dividing the selling value of cows available from FADN by the cows' live weight price obtained from Eurostat.

2.2.2.5. The HPD program for estimating feed requirements and feed resources

The feed module aims to balance feed requirements and feed availability at farm level as described in section 2.2.2.4. It describes quantity of certain feed categories (cereals, rich protein, rich energy, feed based on dairy products, other feed) or single feeding stuff (fodder maize, grass, fodder from arable land, straw, milk for feeding) used per animal activity level (cows, heifers, calves, etc.). It also ensures that the total energy, protein, dry matter and fibre requirements of animals are met by the own-produced and purchased quantities of feed. The feed requirements can be covered by roughage produced on-farm or purchased (hay, straw, silage, etc.) and own-produced or purchased concentrates.

Assuming that the feed contents are accurately known, the objective is to estimate, at given animal herd sizes and prices, the quantity of feeding stuffs needed to meet animal requirements, in physical units and nutrient values, at the minimum feed costs. In addition, the minimum relative squared deviation between estimated animal requirements and prior information (including the deviation from lower/upper bounds) is assured, as is the minimum relative squared deviation between estimated on-farm produced feed quantity, purchased feed quantity, feed from common land (i.e. grass) other use of on-farm consumed crops and feed costs and their observed values in FADN data. This is performed with the HPD approach using information on feeding costs and on-farm produced feed reported in the FADN database, feed content, feed prices, prior information on animal requirements functions, a set of constraints for balancing feed requirement and feed availability (energy, crude protein, fibre, dry matter) and a set of consistency constraints.

The model results provide estimates on nutrient requirements, physical quantity of feed for each feed and each animal activity, quantity of on-farm produced and purchased feed, feed from common land, fodder yields, and other use of on-farm consumed crops.

The general formulation of the HPD problem is the following:

$$\min\,(b_{f,i,n}^{F}, db_{f,i,n}^{F+}, db_{f,i,n}^{F-}, \zeta_{f,i,m}, z_{f,m}, z_{f,m}^{own}, z_{f,m}^{other}, z_{f,m}^{comn}, dz_{f,m}^{comn}, \zeta_{f,i,m}, \zeta_{f,i,m}^{own}, \zeta_{f,i,m}^{purc}, \zeta_{f,i,m}^{comn}, \qquad (54)$$

$$dc^{own}_{f,AG,FG}, dc^{purc}_{f,AG,FG}, b^{DM}_{f,i,FG^-}, db^{DM}_{f,i,FG^-}, y_{f,i,m}, d\tau_{f,i,m})\ HPD_f =$$

$$+\sum_{i\in A,n}\left(\frac{b^F_{f,i,n}-\bar{b}^F_{f,i,n}}{\sigma^b_{f,i,n}}\right)^2$$

Minimization of the deviation of animal requirements from the prior value

$$+\sum_{i\in A,n}\left(\frac{db^{F+}_{f,i,n}}{\sigma^{b-}_{f,i,n}}\right)^2+\sum_{i\in A,n}\left(\frac{db^{F-}_{f,i,n}}{\sigma^{b+}_{f,i,n}}\right)^2$$

Minimization of the deviation of animal requirements from the lower and upper bounds

$$+\sum_{m\in F}\left(\frac{p_{f,m}z_{f,m}}{2}\frac{1}{\sum_{m\in M}\left[p_m\left(\sum_{i\in A}y_{i,m}x_i\right)\right]}\right)^2$$

Feed cost minimization weighted by the revenue of livestock activities

$$+\sum_{m\in NFODR}\left(\frac{z^{own}_{f,m}-z^o_{f,m}}{0.3\cdot z^o_{f,m}}\right)^2$$

Minimization of the deviation of on-farm produced feed quantity ($z^{own}_{f,m}$) from the FADN observed on-farm use of the own produced crop output ($z^o_{f,m}$)

$$+\sum_{i\in C}\sum_{m\in FODR}\left(\frac{y_{f,i,m}-\bar{y}_{f,i,m}}{0.3\cdot\bar{y}_{f,i,m}}\right)^2$$

Minimization of the deviation of fodder yields from the prior values

$$+\sum_{m\in F}\left(\frac{dz^{comn}_{f,m}}{0.3\cdot\bar{z}^{comn}_{f,m}}\right)^2$$

Minimization of the deviation of common land feed quantity ($z^{comn}_{f,m}$) from the prior value ($\bar{z}^{comn}_{f,m}$)

$$+\sum_{i\in A}\sum_{FG}\left(\frac{d\psi^{min}_{f,i,FG}}{2}\right)^2$$

Minimization of the deviation of the dry matter minimum share of the specific feed group in livestock diet

$$+\sum_{i\in A}\sum_{m\in CONC}\left(\frac{d\tau_{f,i,m}}{2}\right)^2$$

Minimization of the deviation of the maximum concentrate feed share in the total concentrate feed diet

$$+\sum_{AG,FG}\left[\frac{\left(\frac{dc^{own}_{f,AG,FG}}{0.3\cdot c^{o,own}_{f,AG,FG}}\right)^2+\left(\frac{dc^{purc}_{f,AG,FG}}{0.3\cdot c^{o,purc}_{f,AG,FG}}\right)^2}{\sum_{FG}\left[p_{FG}\left(\sum_{AG}y_{AG,FG}x_{AG}\right)\right]}\right]$$

Minimization of the deviation of the estimated aggregate own produced/purchased feed costs from the aggregated own produced/purchased feed costs observed in FADN weighted by the revenue of livestock activities

Subject to

$\forall i\in A\ ,\ \forall n\in N^+$

$$\sum_{m\in F}\zeta_{f,i,m}\cdot\delta_{f,n,m}\geq b^F_{f,i,n}$$

The sum of nutrient content n of different types of feed m given to animal i (sum of variable $\zeta_{i,m}\delta_{f,n,m}$ over m) is greater than the minimum requirement of nutrient n (55)

$\forall i\in A,\ \forall n\in N^o$

$$\sum_{m\in F}\zeta_{f,i,m}\cdot\delta_{f,n,m}=b^F_{f,i,n}$$

The sum of nutrient content n of different types of feed m given to animal i (sum of variable $\zeta_{f,i,m}\cdot\delta_{f,n,m}$ over m) exactly satisfies the requirement for nutrient n (56)

$\forall i\in A,\ \forall n\in N$

$$b^F_{f,i,n}+db^{F-}_{f,i,n}\leq\bar{b}^{F-}_{f,i,n}$$

The estimated animal requirement n ($b^F_{f,i,n}$) must be lower than the upper (57)

		bound ($b_{f,i,n}^{o,F-}$) adjusted by the deviation ($db_{f,i,n}^{F-}$)	
$\forall i \in A$, $\forall n \in N$	$b_{f,i,n}^{F} + db_{f,i,n}^{F+} \geq \bar{b}_{f,i,n}^{F+}$	The estimated animal requirement *n* ($b_{f,i,n}^{F}$) must be greater than the lower bound ($b_{f,i,n}^{o,F+}$) adjusted by the deviation ($db_{f,i,n}^{F+}$)	(58)
$\forall AG$, $\forall FG$	$\sum_{i \in AG} \sum_{m \in FG} p_{f,m} \zeta_{f,i,m}^{purc} x_{f,i}^{o} + dc_{f,AG,FG}^{purc} = c_{f,AG,FG}^{o,purc}$	The estimated purchased feed costs (sum per animal feed costs for each animal group) should be as close as possible to the aggregated purchased feed costs observed in FADN at farm level ($c_{f,AG,FG}^{o,purc}$) across all animal groups AG and feed groups FG	(59)
$\forall AG$, $\forall FG$	$\sum_{i \in AG} \sum_{m \in FG} p_{f,m} \zeta_{f,i,m}^{own} x_{f,i}^{o} + dc_{f,AG,FG}^{own} = c_{f,AG,FG}^{o,own}$	The estimated own produced feed costs (sum per animal feed costs for each animal group) should be as close as possible to the aggregated own produced feed costs observed in FADN at farm level ($c_{f,AG,FG}^{o,purc}$) across all animal groups AG and feed groups FG	(60)
$\forall m \in F$	$z_{f,m} = z_{f,m}^{own} + z_{f,m}^{purc} + z_{f,m}^{comn}$	The estimated quantity of feed *m* at farm level ($z_{f,m}$) equals estimated own-produced feed quantity ($z_{f,m}^{own}$), estimated purchased feed quantity ($z_{f,m}^{purc}$) and estimated feed from common land (i.e. grass) ($z_{f,m}^{comn}$)	(61)
$\forall m \in NFODR$	$z_{f,m}^{o} = z_{f,m}^{own} + z_{f,m}^{other}$	FADN observed on-farm use of the own produced non-fodder crop output *m* ($z_{f,m}^{o}$) equals estimated feed use plus estimated other on-farm uses of the crop output *m*	(62)
$\forall m \in F$	$z_{f,m}^{comn} + dz_{f,m}^{comn} = \bar{z}_{f,m}^{comn}$	The estimated feed quantity from common land (grass) ($z_{f,m}^{comn}$) is equal to its prior value ($\bar{z}_{f,m}^{comn}$) and the deviation variable ($dz_{f,m}^{comn}$)	(63)
$\forall i \in C$, $\forall m \in FODR$,	$z_{f,m}^{own} = y_{f,i,m} x_{f,i}^{o}$	Estimated on-farm use of the own produced fodder output $z_{f,m}^{own}$ equals the estimated yield $y_{f,i,m}$ times the observed activity level $x_{f,i}^{o}$	(64)
$\forall m \in F$	$z_{f,m}^{own} = \sum_{i \in A} \zeta_{f,i,m}^{own} x_{f,i}^{o}$	Total own-produced feed *m* at farm level $z_{f,m}^{own}$ equals the sum of the per animal own-produced feed quantities multiplied by livestock activity levels	(65)
$\forall m \in F$	$z_{f,m}^{purc} = \sum_{i \in A} \zeta_{f,i,m}^{purc} x_{f,i}^{o}$	Total purchased feed *m* at farm level $z_{f,m}^{purc}$ equals the sum of the per animal purchased feed quantities multiplied by livestock activity levels	(66)

$\forall m \in F$	$z^{comn}_{f,m} = \sum_{i\in A} \zeta^{comn}_{f,i,m} x^{o}_{f,i}$	Total feed *m* feed from common land at farm level $z^{comn}_{f,m}$ equals the sum of the per animal feed obtained from common land multiplied by livestock activity levels	(67)
$\forall i \in A,\ \forall m \in F$	$\zeta_{f,i,m} = \zeta^{own}_{f,i,m} + \zeta^{purc}_{f,i,m} + \zeta^{comn}_{f,i,m}$	Total amount of feed *m* per animal *i* $(\zeta_{f,i,m})$ equals own-produced feed quantity per animal $(\zeta^{own}_{f,i,m})$, purchased feed quantity per animal $(\zeta^{purc}_{f,i,m})$ and feed from common land per animal (i.e. grass) $(\zeta^{comn}_{f,i,m})$	(68)
$\forall i \in A,\ \forall FG$	$\sum_{m\,\in\,FG^{+}} \zeta_{f,i,m} \cdot \delta^{DM}_{f,m} \geq \left(\psi^{min}_{f,i,FG} + d\psi^{min}_{f,i,FG}\right) b^{F}_{f,i,DM}$	The contribution to dry matter of feeds belonging to the specific feed type FG must be greater than the minimum allowed $\psi^{min}_{f,i,FG} b^{F}_{f,i,DM}$ adjusted by the deviation $d\psi^{min}_{f,i,FG} b^{F}_{f,i,DM}$	(69)
$\forall i \in A,\ \forall FG$	$\sum_{m\,\in\,FG^{-}} \zeta_{f,i,m} \cdot \delta^{DM}_{f,m} \leq \psi^{max}_{f,i,FG} b^{F}_{f,i,DM}$	The contribution to dry matter of feeds belonging to the specific feed type FG must be lower than the maximum allowed $\psi^{max}_{f,i,FG} b^{F}_{f,i,DM}$	(70)
$\forall i \in A,\ \forall m \in CONC$	$\zeta_{f,i,m} \leq \left(\tau_{f,i,m} + d\tau_{f,im}\right) \sum_{m\,\in\,CONC} \zeta_{f,i,m}$	The different concentrate feed *m* given to animal i $(\zeta_{f,i,m})$ must be lower than the maximum share of the total concentrate feed quantity $(\tau^{o}_{f,i,m})$ adjusted by the deviation variable $(d\tau_{f,i,m})$	(71)
$\forall m \in F$	$z^{purc}_{f,m} \cdot z^{other}_{f,m} = 0$	Consistency constraint: ensures that purchase and other on-farm use of feed *m* does not occur at the same time	(72)
$\forall m \in FODR$	$z^{purc}_{f,m} \left(dz^{comn}_{f,m} + \sum_{m\in FODR} z^{other}_{f,m} \right) = 0$	Consistency constraint: excludes purchase of fodder *m* if fodder other on-farm use or the deviation of the feed from common land from its prior value are positive for at least one fodder output	(73)
$\forall m \in CERE$	$z^{purc}_{f,m} \sum_{m\in CERE} z^{other}_{f,m} = 0$	Consistency constraint: excludes purchase of cereal *m* if other on-farm use is positive for at least one cereal output	(74)
$\forall m \in PCONC$	$z^{purc}_{f,m} \sum_{m\in PCONC} z^{other}_{f,m} = 0$	Consistency constraint: excludes purchase of concentrate feed *m* not produced on-farm if other on-farm use is positive for at least one cereal output	(75)

where o indexes observed value of a given variable; n denotes the different types of nutrient or energy requirements $(n \in N)$; AG defines the set of animal activity groups (e.g., ruminants, pigs, and poultry) $(AG \subset A)$; FG is set of feed groups (e.g., concentrated, and rough types of feed) $(FG \subset F)$; $FODR$ and $NFODR$ is set of fodder outputs $(FODR \text{ and } NFODR \subset F)$; $CERE$ is set of cereal outputs $(CERE \subset F)$; and $CONC$, and $PCONC$ are sets of concentrate feed and purchased concentrate feed not produced on-farm, respectively ($CONC \text{ and } PCONC \subset F$). $b_{f,i,n}^{F}$ is the animal requirement for nutrient n and livestock activity i; $b_{f,i,DM}^{F}$ is animal requirement for dry matter; $db_{f,i,n}^{F+}$ and $db_{f,i,n}^{F-}$ are the deviations of animal nutrient requirements from their lower and upper bounds, respectively; $z_{f,m}$ is total quantity of farm use of feed m; $z_{f,m}^{own}$, $z_{f,m}^{purc}$ and $z_{f,m}^{comn}$ are total quantity of own-produced feed, purchased feed, and feed obtained from common land (i.e. grass), respectively; $z_{f,m}^{o}$ is FADN observed on-farm use of the own produced crop output m; $dz_{f,m}^{comn}$ is the deviations of the common land feed (grass) from its prior value; $\zeta_{i,m}$ is the amount of feed m given to animal activity i; $\zeta_{f,i,m}^{own}$, $\zeta_{f,i,m}^{purc}$ and $\zeta_{f,i,m}^{comn}$ are the amount of on-farm produced, purchased and common land feed m given to animal activity i, respectively; $c_{f,AG,FG}^{o,own}$ and $c_{f,AG,FG}^{o,purc}$ are costs reported in FADN for own-produced and purchased feed, respectively; $dc_{f,AG,FG}^{own}$and $dc_{f,AG,FG}^{purc}$ are the error terms for the estimated costs relative to the costs reported in FADN for own-produced and purchased feed, respectively; $\psi_{f,i,FG}^{min}$ and $\psi_{f,i,FG}^{max}$ are the minimum and maximum share in total feed consumption (represented in dry matter) from feed group FG for animal activity i, respectively; $d\psi_{f,i,FG}^{min}$ is the deviation of the minimum dry matter share; $\tau_{f,i,m}$ is the maximum concentrate feed share in the total concentrate feed diet (represented in tonnes); and $d\tau_{f,i,m}$ is the deviation of the concentrate feed share from the maximum concentrate feed share. Regarding prior information, $\bar{b}_{f,i,n}^{F}$and $\sigma_{f,i.n}^{b}$ are the mean and standard deviation of the animal nutrient requirements, respectively; $\bar{b}_{f,i,n}^{F+}$ and $\bar{b}_{f,i,n}^{F-}$ are the lower and upper bounds of the animal nutrient requirements, respectively; $\sigma_{f,i,n}^{b-}$ and $\sigma_{f,i,n}^{b+}$ are the standard deviations of the lower and upper bounds of the animal nutrient requirements, respectively; $\bar{y}_{f,i,m}$ is the mean yield (yield prior value); and $\bar{z}_{f,m}^{comn}$ is prior value of feed obtained from common land (i.e. grass).

The first component of the objective function (54) is linked to the minimisation of the normalised squared deviation of estimated animal requirements from the prior information; the second component is related to the minimisation of normalised squared deviation estimated animal requirements from lower and upper bounds. The aim is to impose a higher penalty if requirements are outside the bounds. The third component ensures cost minimisation of feed; the fourth component minimises the relative squared deviation between the estimated on-farm produced feed and observed FADN on-farm use of the own produced crop output quantity; the fifth component minimises the relative squared deviation between the estimated fodder yield and its prior value; the sixth component minimises the relative squared deviation between the estimated common land feed quantity from the prior value; the seventh component minimise the relative squared error of the feed group share from the minimum share in the total feed consumption (measured by dry matter), whereas the eighth component does the same for the individual concentrate feed shares in total concentrate consumption but measured in physical quantities instead of dry matter. The final component minimises the relative squared error of the estimated (on-farm produced and purchased) feed costs from the FADN recorded feed costs. Because all components in the objective function except the cost minimisation element are differences, we scale the function by the livestock production value.

Equations (55) and (56) balance the feed requirement with the feed availability in nutrient values. Equations (57) and (58) constrain the deviation of animal requirements to be within or around the lower and upper bounds of animal requirements. The bounds are used to account for the uncertainty in data determining the level of animal requirements. Equations (59) and (60) constrain the estimated costs of on-farm produced feed and purchased feed, respectively, to be as close as possible to their observed values in FADN. Equations (61) and (62) ensure that physical quantity of feed is balanced at farm level for the total quantity of feed (the total feed equals on-farm produced feed, purchased feed and feed from common land) and on-farm consumed of the own-produced crop output (on-farm consumed crop output equals feed use and other use), respectively. Equation (63) constrains the feed obtained from common land to be as close as possible to its prior value obtained from FADN. Equation (64) establishes physical quantity balance for fodder production. Equations (65), (66) and (67) sum the feed use over all animal activities for on-farm produced feed, purchased feed and feed from common land, respectively. Equations (68) ensures that physical quantity of feed is balanced per animal: the total feed per animal equals on-farm produced feed per animal, purchased feed per animal and feed per animal from common land. The minimum share constraint (69) ensures that a given feeding stuff (or group of feed) represents at least a certain amount of total feed consumption (measured in dry matter), whereas the maximum share constraint (70) ensures that a given feeding stuff (or group of feed) does not exceed a certain limit in the total feed consumption for a given animal activity. These two constraints ensure certain feed management practices and prevent overuse or underuse of certain feeds. The constraint described by equation (71) ensures that a feed concentrate does not exceed a certain maximum limit in the total concentrate feed consumption for a given animal activity (measured in physical quantity). This constraint aims to ensure that the composition of the concentrate feed corresponds as closely as possible to the observed data (21). Equations (72) to (75) are consistency constraints that aim to exclude certain feeding practices perceived unlikely or unreasonable (i.e. exclude feed purchase and other on-farm use at the same time for different feed types).

Figure 7, Figure 8 and Figure 9 compare the HPD estimated IFM-CAP costs with the actual FADN costs aggregated at MS level for purchased feed, farm produced feed and total feed costs, respectively. We also report the slope for the estimated linear model between IFM-CAP costs and the FADN costs (22). A slope value equal to 1 implies that, on average, the estimated IFM-CAP costs correspond to the FADN costs at MS level. A slope lower than 1 implies that estimated costs are on average lower than the FADN costs. As shown in the figures, the slope between the estimated and FADN costs is highest for the total feed costs (99 %), followed by purchased feed (99 %) and the own feed (90 %). At farm level, the slopes are 77 %, 70 %, and 72 % for total feed costs, purchased feed costs and the own feed costs, respectively (not shown in the figures). Because the slopes are somehow lower than 1 (particularly at farm level), our model underestimates the FADN costs. The use of external data and regional aggregates for nutrient feed content, feed prices and, to some extent, for fodder yields may have led to differences between the estimated and the observed costs. The discrepancy arises because, in reality, these data will probably vary across farms and thus may depart from the regionally aggregated values. In addition, imposing cost minimisation of feed mix may have led to an underestimate of the feed cost given that, in reality, strict cost minimisation may not always hold, particularly in the presence of market imperfections (e.g. transaction costs, uncertainty).

---

(21) The maximum limits are available from FEEDMOD.

(22) Linear model is specified as follows: *IFM-CAP estimated costs = Slope * FADN costs + error*

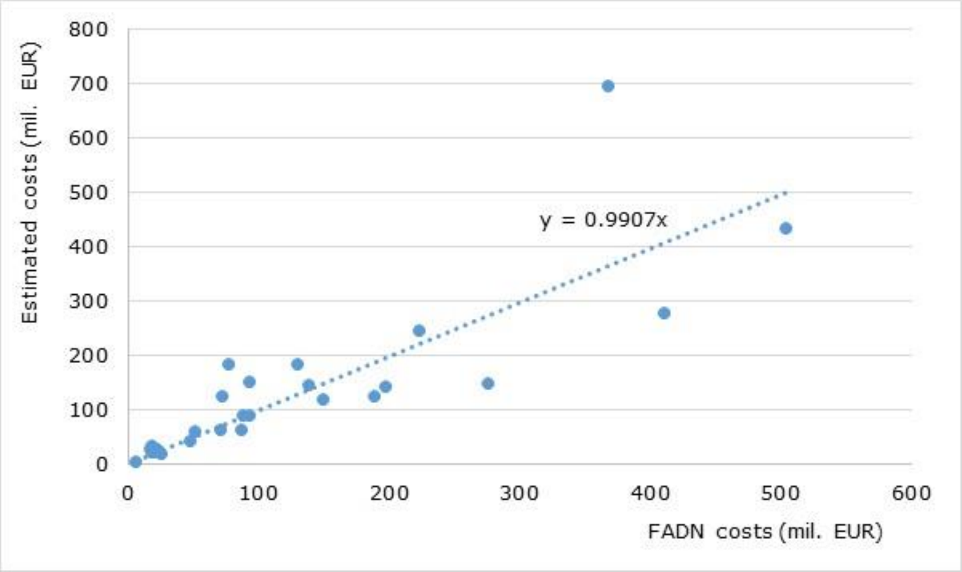


Figure 7. FADN purchased feed costs versus IFM-CAP estimated purchased feed costs aggregated at MS level (EUR/farm)

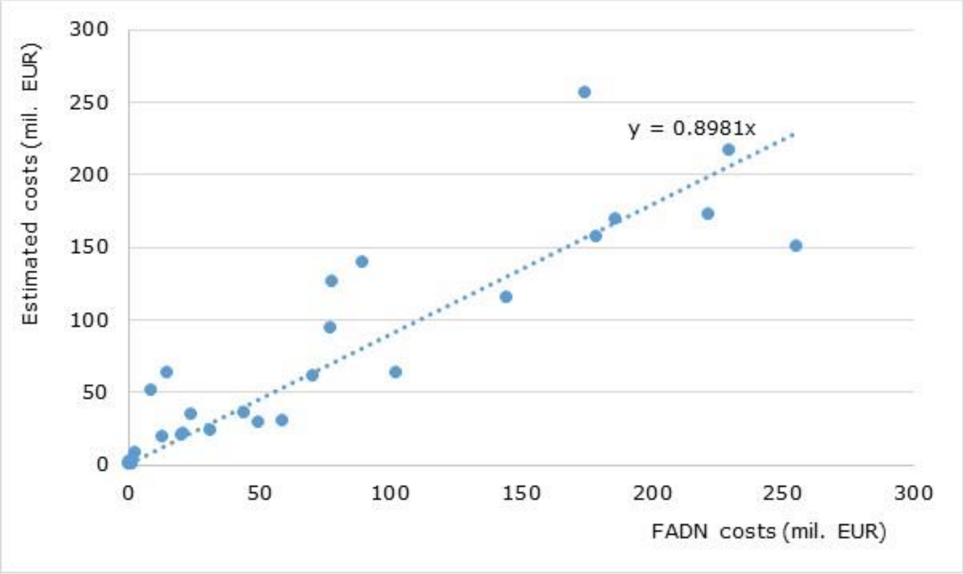


Figure 8. FADN costs of farm produced feed versus IFM-CAP estimated costs of farm produced feed aggregated at MS level (EUR/farm)

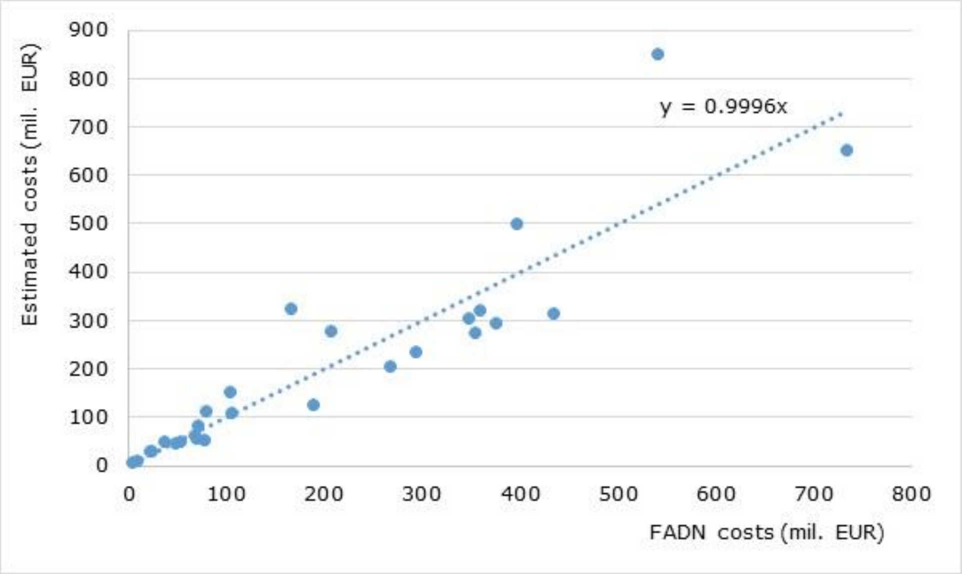


Figure 9. Total FADN feed costs versus IFM-CAP estimated total feed costs aggregated at MS level (EUR/farm)

Figure 10 to Figure 16 compare the HPD estimated animal intake requirements, with their lower and upper bounds, for dairy cows (DCOW), fattening of pigs (PIGF) and sheep and goat activity (SHGM) for selected nutrients for all farms modelled in IFM-CAP. As reported in Figure 10 and Figure 11, the estimated energy (ENNE) and protein (CPRP) intakes of dairy cows are around the lower bound for most farms. In contrast, the estimated intake of fibre (FIDI) is in between upper and lower bounds for most farms (Figure 12). The main explanation for the underestimation of energy and protein is that the HPD estimation model cannot balance them within the bounds for the given set of feeds (determined by the constraints described by equations (55) to (75)). The ratio of the energy and protein content of the available feed cannot be matched with the ratio of these requirements for dairy cows such that they remain within the lower and upper bounds. This could be because the nutrient contents of feed in our model are not farm level specific but are provided at MS level, and thus may depart from the actual values. This is particularly problematic for fodder feed, the nutrient content of which may vary widely across regions and farms. Similarly to the balancing problem of dairy cows, the estimates for energy requirements for fattening of pigs (PIGF) are around the lower bound for most farms, whereas the estimates for protein intake are around the upper bound (Figure 13 and Figure 14). In contrast to dairy cows and fattening of pigs, the energy and protein requirements for sheep and goat activity (Figure 15 and Figure 16) are mostly within lower and upper bounds. This could be because of a less heterogeneous diet of sheep and goat activity thus allowing the HPD model to balance the supply and demand of feed such that animal requirements are within the bounds.

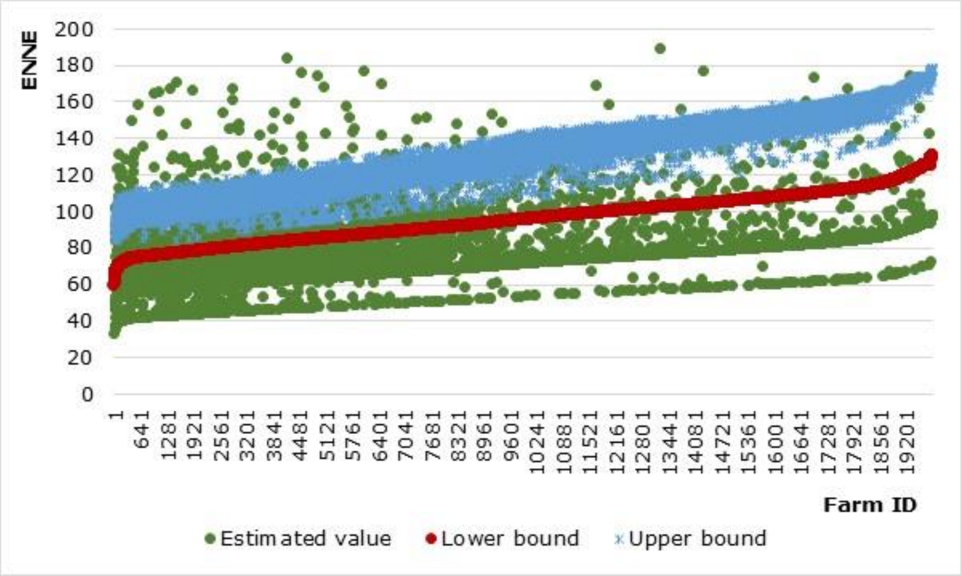


Figure 10. Estimated ENNE intake for DCOW for all IFM-CAP farms

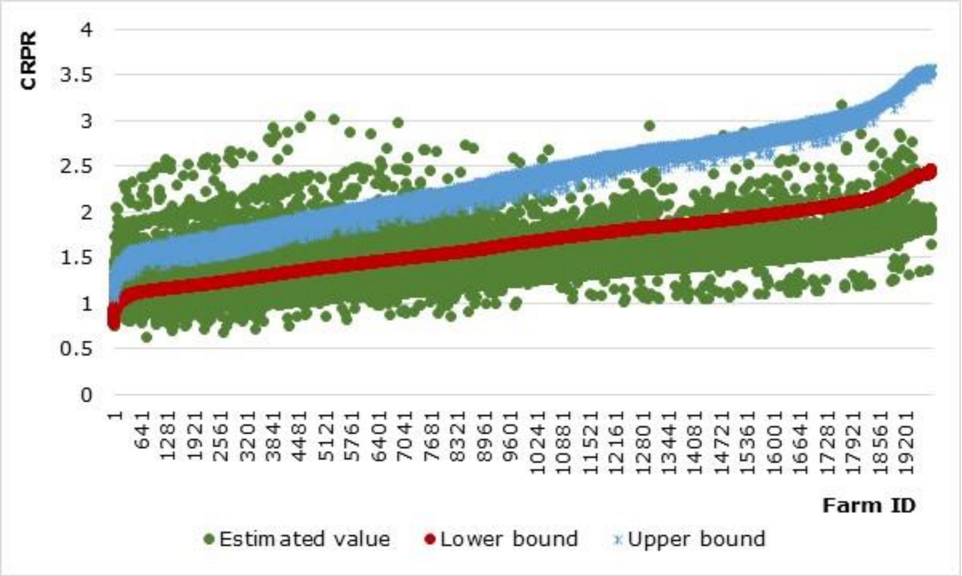


Figure 11. Estimated CRPR intake for DCOW for all IFM-CAP farms

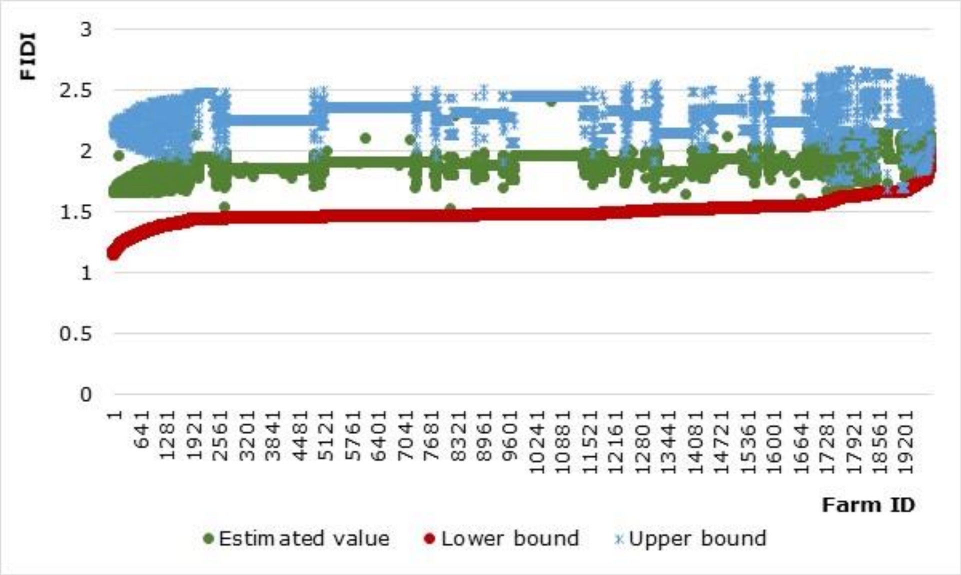


Figure 12. Estimated FIDI intake for DCOW for all IFM-CAP farms

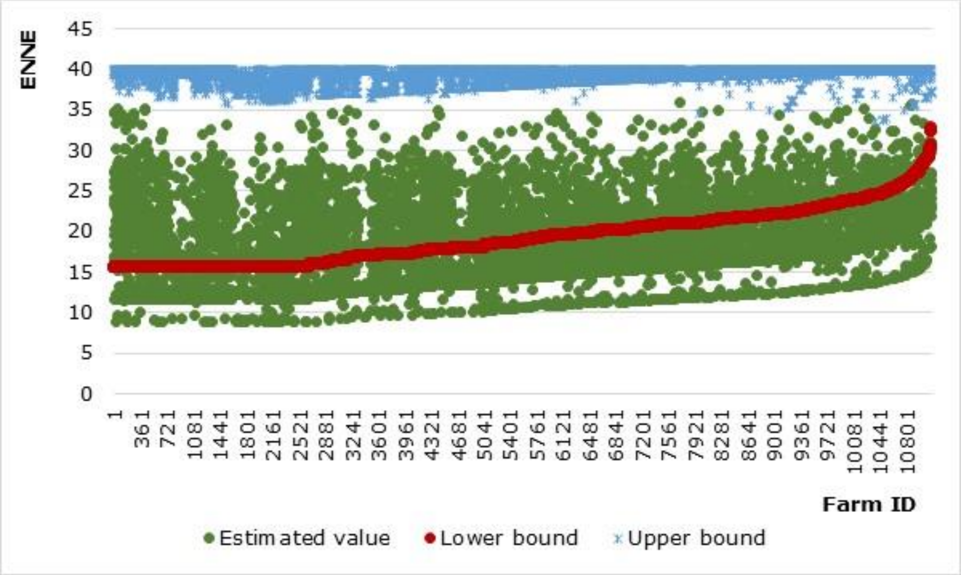


Figure 13. Estimated ENNE intake for PIGF for all IFM-CAP farms

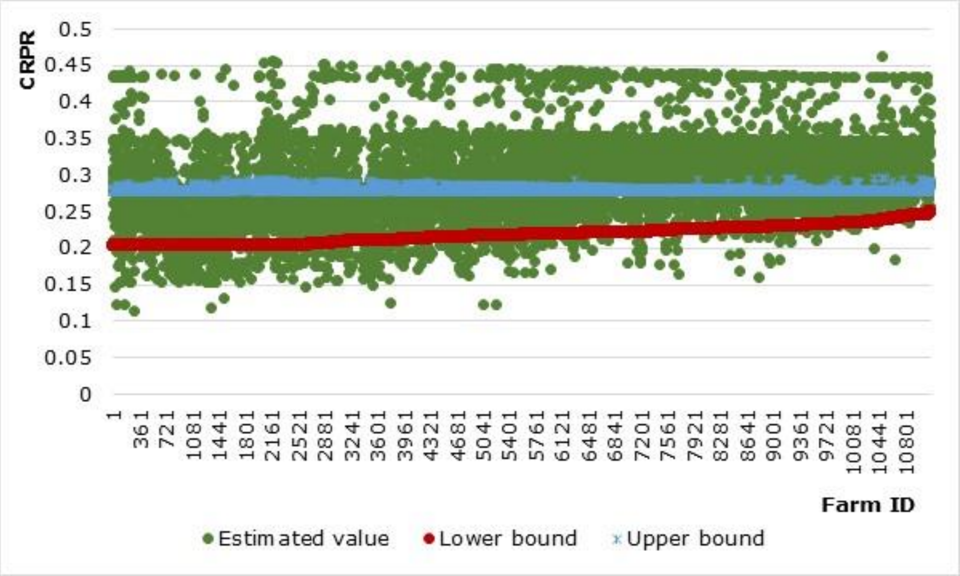


Figure 14. Estimated CRPR intake for PIGF for all IFM-CAP farms

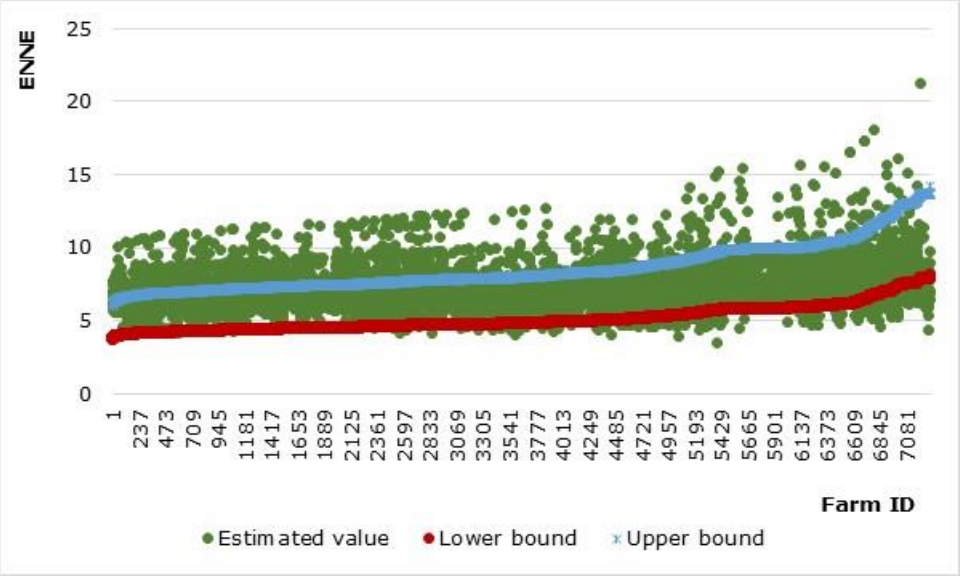


Figure 15. Estimated ENNE intake for SHGM for all IFM-CAP farms

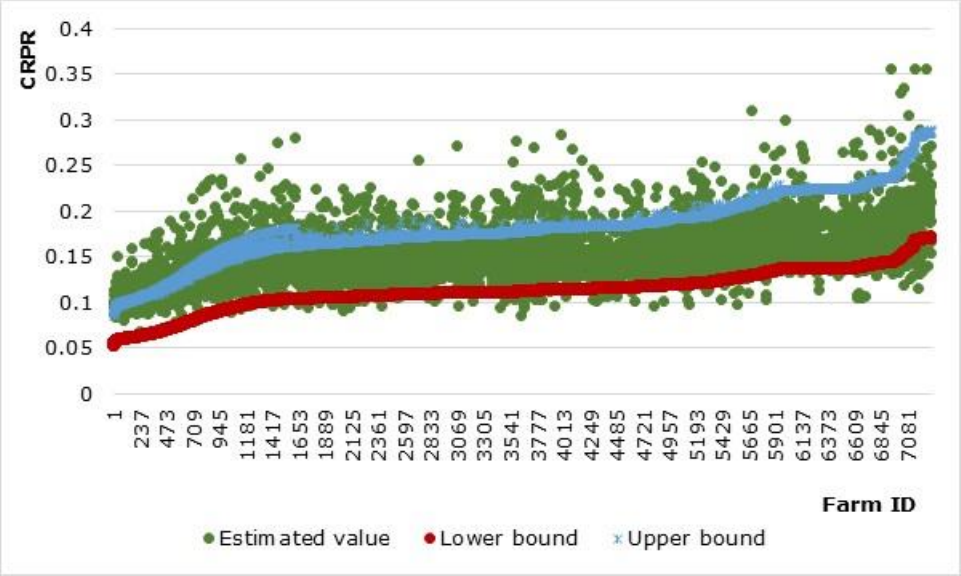


Figure 16. Estimated CRPR intake for SHGM for all IFM-CAP farms

2.2.3. Model calibration

2.2.3.1. Calibration theory

Over the past decade, several PMP approaches have been developed to accurately calibrate programming models (23). The original PMP approach, introduced by Howitt (1995), relied on the introduction of non-linear terms (a non-linear behavioural function) to an otherwise linear optimisation model. Through such additional non-linear terms, the objective function of the (profit-maximising) model becomes concave and allows for interior solutions that were not previously feasible for a given constraint set (24). PMP calibration can be succinctly described as the process of recovering the parameters of the behavioural function so that the necessary first-order conditions (FOCs) of the final model are exactly satisfied at the observed levels of the related variables (i.e. we assume that the optimal solution coincides with the observations). As detailed in Section 2.1, the non-linear behavioural function for IFM-CAP is associated with the variables for crop and animal activity levels (through parameters d_i and $Q_{i,j}$) and for the feed input coefficient (through parameters $d_{i,m}^F$ and $Q_{i,m}^F$). Effectively, calibration of IFM-CAP aims to recover these unknown parameters, together with the CARA coefficient φ, so that the optimisation model exactly reproduces the observed activity levels $\bar{x}_i$ and the observed value of the feed input coefficient $\bar{\zeta}_{i,m}$ in the base year.

For simplicity of exposition, we break down the FOCs of the optimisation program in IFM-CAP into those related to (i) crop activities, (ii) animal activities and (iii) feed input coefficients. Using the notation introduced in Section 2.1, the FOCs for the i-th crop activity, when evaluated at the optimal level of all unknown variables, can be written as:

$$\sum_{m\in M} p_m y_{i,m} - \sum_k c_{i,k} - \sum_{m\in F} p_m \xi_m y_{i,m} - \left(d_i + \sum_j \bar{x}_j Q_{i,j}\right) - \varphi \sum_j \bar{x}_j \Omega_{i,j} - \sum_t A_{i,t}\bar{\theta}_t = 0 \tag{76}$$

$$b_t - \sum_i A_{t,i}\bar{x}_i = 0 \tag{77}$$

$$\forall\, i \in \text{ set of crop activities}$$

Similarly, the FOCs for animal activities can be written as:

$$\begin{aligned}\sum_{m\in M} p_m y_{i,m} - 10^{-3}\sum_{m\in F} p_m \bar{\zeta}_{i,m} - \sum_{m\in F} p_m \xi_m y_{i,m} - \left(d_i + \sum_j \bar{x}_j Q_{i,j}\right) - \sum_{m\in F} d_{i,m}^F \bar{\zeta}_{i,m} \\ - 0.5\sum_{m\in F} \bar{\zeta}_{i,m} Q_{i,m}^F \bar{\zeta}_{i,m} - \varphi \sum_j \bar{x}_j \Omega_{i,j} - \sum_t A_{t,i}\bar{\theta}_t = 0\end{aligned} \tag{78}$$

$$b_t - \sum_i A_{i,t}\bar{x}_i = 0 \tag{79}$$

$$\forall\, i \in \text{ set of animal activities}$$

For the feed input coefficients:

---

(23) For a review on PMP models see de Frahan et al. (2007), Heckelei and Britz (2005), Heckelei et al. (2012), Mérel and Bucaram (2010) and Paris (2011).

(24) PMP approaches have also been applied to intrinsically non-linear models, including non-linear E-V models (Petsakos and Rozakis, 2015) and supply models with explicit production relationships, such as constant elasticity of substitution or other production functions (e.g. Mérel et al., 2011). As in the typical PMP approach for linear models, calibration in this case focuses on recovering, or adjusting, the non-linear parameters so that the necessary first-order conditions are satisfied at the observed activity allocation.

$$-10^{-3}p_m\bar{x}_i - d_{i,m}^F\bar{x}_i - \bar{\zeta}_{i,m}Q_{i,m}^F\bar{x}_i - \sum_{v}\left(\sum_{n}A_{n,m,v}^F\bar{\theta}_{i,n,v}^F\right) = 0 \qquad (80)$$

$$\forall\, i\ \in\ \text{set of animal activities, and}\ \forall\, m\ \in\ \text{set of feeds}$$

Recovering the PMP terms (d_i, $Q_{i,j}$, $d_{i,m}^F$ and $Q_{i,m}^F$) in the above equations is not a trivial exercise and cannot be carried out using standard econometric estimation techniques because the number of unknowns parameters is greater than the number of available equations, that is, the FOCs that need to be satisfied. This constitutes an undetermined equation system with no unique solution (an 'ill-posed' problem). The ill-posedness of the above system of equations is exacerbated by the presence of the shadow prices of the binding constraints ($\bar{\theta}_t$ and $\bar{\theta}_{i,n,v}^F$) in the FOCs, which are themselves unknown and need to be recovered as well. For this reason, most early PMP approaches, applied to LP models, used the traditional PMP procedure, which ignored the shadow prices from the final calibrating equation. Such approaches typically set all off-diagonal $Q_{i,j}$ elements to zero and calculated the remaining parameters using ad hoc assumptions (25). Although keeping only the diagonal $Q_{i,i}$ elements is often an acceptable simplification for quadratic-cost PMP models (e.g. Mérel and Bucaram, 2010) and it is also adopted for the current version of IFM-CAP, the traditional PMP calibration procedure has been heavily criticised because the transition from a linear model to a non-linear one (by adding the non-linear behavioural function) leads to inconsistent shadow prices in the final non-linear model, when calibration is examined from an econometrician's viewpoint (Heckelei and Wolff, 2003). Furthermore, the theoretical justification of the accompanying ad hoc assumptions is weak, and the overall calibration process does not control for the model's response to parameter changes (i.e. the model's second-order properties), which may sometimes lead to erratic model behaviour.

To ensure a non-arbitrary parameter specification, recent applied programming models have either (i) estimated model parameters with cross-sectional data (Heckelei and Britz, 2000) or time series data when multiple observations were available (Britz and Arata, 2019; Buysse et al., 2007; Jansson and Heckelei, 2011) or (ii) used exogenous information on supply elasticities (Britz and Witzke, 2014; Mérel and Bucaram, 2010) and/or on shadow prices of resources (de Frahan et al., 2007). The objective of calibrating an optimisation model by forcing it to replicate, or at least approximate, exogenous information on price elasticities of supply is to ensure that the response of the model with respect to changes in at least one parameter of interest (in this case, prices) is consistent with the available exogenous information about what this response may be (the elasticity values).

For the calibration of IFM-CAP we use both multiple observations (cross-sectional data) and prior information on NUTS 2 price supply elasticities to calibrate the base year farm activity plans. Price elasticities of supply for crops are taken from available econometric studies at NUTS 2 level (Jansson and Heckelei, 2011) (26). Elasticities of 1 and 0.1 are used for annual crops and permanent crops, respectively, when prior information is unavailable. Supply elasticities for livestock activities, as well as feed demand elasticities, are taken from CAPRI.

When using supply elasticity priors for calibrating mathematical programming models of agricultural supply, the elasticity formula needs to account not only for the impact of prices on supply but also for the impact of prices on the dual values of the model's binding constraints. In other words, and using vector notation for simplicity (in bold), the

---

(25) A presentation of these early PMP approaches can be found in Heckelei (2002).

(26) The use of supply elasticities from Jansson and Heckelei (2011) is motivated by the fact that they provide estimates at EU regional level; there are no other studies available that would provide better regional resolution and/or estimates of land allocation elasticities across EU regions. Moreover, IFM-CAP assumes fixed yields, meaning that land allocation elasticities correspond to supply elasticities.

optimal value $\bar{\mathbf{x}}$ of the activity vector is a function of both prices and dual values, while the resulting duals $\bar{\boldsymbol{\theta}}$ are themselves a function of prices. These relationships can be expressed using the functions $\bar{\mathbf{x}} = X(\mathbf{p}, \bar{\boldsymbol{\theta}})$ and $\bar{\boldsymbol{\theta}} = \Theta(\mathbf{p})$, so that the total effect of $\mathbf{p}$ on $\bar{\mathbf{x}}$ is given by the total derivative evaluated at $\{\bar{\mathbf{x}}, \bar{\boldsymbol{\theta}}\}$:

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}\mathbf{p}} = \frac{\partial X}{\partial \mathbf{p}} + \frac{\partial X}{\partial \boldsymbol{\theta}} \frac{\partial \Theta}{\partial \mathbf{p}} \tag{81}$$

Mérel and Bucaram (2010) call this approach 'non-myopic', or 'exact', as opposed to 'myopic' calibration, which focuses only on how $\mathbf{p}$ affects $\bar{\mathbf{x}}$ through the partial derivative $\partial F / \partial \mathbf{p}$, and ignores the impact of $\mathbf{p}$ on $\bar{\boldsymbol{\theta}}$ (i.e. the myopic approach does not consider the second term in the right hand side of equation (81)). Under a non-myopic calibration approach, and setting $\mathbf{T} = \mathbf{Q} + \varphi \boldsymbol{\Omega}$, the price elasticity of supply $\mathbf{E}^s$ for IFM-CAP crop activities can be expressed as ([27]):

$$\mathbf{E}^s \equiv \frac{\mathrm{d}\mathbf{x}}{\mathrm{d}\mathbf{p}} \cdot \frac{\mathbf{p}}{\bar{\mathbf{x}}} = [\mathbf{T}^{-1} - \mathbf{T}^{-1}\mathbf{A}'(\mathbf{A}\mathbf{T}^{-1}\mathbf{A}')^{-1}\,\mathbf{A}\mathbf{T}^{-1}][(1 - \boldsymbol{\xi})' \cdot \mathbf{y}] \cdot \frac{\mathbf{p}}{\bar{\mathbf{x}}} \tag{82}$$

Equation (82) represents a system of non-linear equations that needs to be solved for $\mathbf{T}$, given an exogenously defined value of $\mathbf{E}^s$. Because of the numerical complexity of this task, (which also includes two matrix inversion operations of $\mathbf{T}^{-1}$ and $(\mathbf{A}\mathbf{T}^{-1}\mathbf{A}')^{-1}$), solving for $\mathbf{T}$ relies on an HPD program, which is presented in Section 2.2.3.2.

There are two important observations related to equation (82). The first is that the linear PMP terms d_i do not affect the model's supply response ([28]). As a result, d_i can be recovered separately from the non-linear terms $Q_{i,j}$, which are part of matrix $\mathbf{T}$ in the above elasticity formula. From a modelling viewpoint, this means that the calibration of IFM-CAP can be split into two steps: the first step relies on equation (82) with exogenous information on $\mathbf{E}^s$ (priors) to recover $Q_{i,j}$, while the second step uses the estimated $Q_{i,j}$ to calculate d_i so that the reference activity vector is reproduced (Garnache and Mérel, 2015). Specifically, once the quadratic terms are recovered using exogenous supply elasticities, d_i is calculated from the FOCs and then further adjusted, if needed, to ensure that the model replicates the observed situation. This latter adjustment involves running the model with additional constraints that bind all activities to their observed levels, and then adding the dual values of these constraints to d_i ([29]).

The other important observation is that non-myopic calibration against supply elasticity priors is possible only when at least one of the constraints is binding for the observed situation (i.e. $\bar{\boldsymbol{\theta}} \geq 0$). This condition assumes a priori knowledge of the correct specification of matrix $\mathbf{A}$, or, more precisely, the partition of $\mathbf{A}$ in each farm that corresponds to binding constraints. If the said partition of $\mathbf{A}$ is a zero matrix, then $\bar{\boldsymbol{\theta}} = 0$ and the term $\mathrm{d}\mathbf{x}/\mathrm{d}\mathbf{p}$ in the elasticity equation reduces to $\partial X / \partial \mathbf{p}$ (equivalent to myopic calibration).

The first task when calibrating IFM-CAP against supply elasticities is thus to determine the binding constraints and to reconstruct matrix $\mathbf{A}$ in each farm. In IFM-CAP, matrix $\mathbf{A}$ is always non-zero, as the land constraint is modelled as an equality and therefore it is

---

([27]) This equation is similar to the elasticity equation 11 in Jansson and Heckelei (2011, p. 145). The only difference is the inclusion of losses $\boldsymbol{\xi}$, which are calculated in the IFM-CAP feed module and are part of the model's objective function.

([28]) The same applies for the shadow prices $\bar{\boldsymbol{\theta}}$ of the model's binding constraints (Mérel and Bucaram, 2010). More precisely, the elasticity formula is equivalent to a value function as it has been derived under the assumption that shadow prices are evaluated at their optimal level.

([29]) The procedure for adjusting d_i resembles the first step of the traditional PMP algorithm in the sense that it uses additional calibration constraints to recover some linear parameters (the dual values of the additional constraints) that allow the model to calibrate. IFM-CAP uses these dual values to simply adjust d_i while retaining the $\mathbf{T}$ matrix previously recovered with the supply elasticity priors, which means that the model's supply response with respect to price changes (the model's implied elasticities) is not affected.

always binding (equation (6) in Section 2.1.2). Drawing on experience with previous versions of IFM-CAP, the VCS constraint (equation (11) in Section 2.1.2) is also assumed to be binding for both crop and animal activities. In addition to the above constraints, we followed a 'reverse engineering' approach to examine if any of the diversification (greening) constraints (equations (18) and (19) in Section 2.1.2) are also binding at farm level under the base year FADN data. This exercise revealed that the 75 % diversification constraint is binding for only 31 farms in the entire 2017 FADN sample, while only three farms satisfied the 95 % diversification constraint as an equality. More importantly, about 4 % of all farms in the base year FADN sample do not comply with the diversification requirements, as these are modelled in IFM-CAP. This can be explained by the existence of aggregate commodities in the model whose components count as individual crops for the purposes of greening by sampled FADN farms (e.g. other cereals activity). In other words, as IFM-CAP cannot fully capture in detail all the available farm options for complying with greening, some farms may fail to calibrate because the diversification constraints in the model are inconsistent with the observed crop allocation in the base year. Similarly, IFM-CAP does not capture all the implementation options for EFAs, for example field margins and buffer strips. For the above reasons, greening constraints were dropped and the only constraints considered in the calibration of crop activities were those related to total land endowment and to VCS payments.

Because of the complexity of the base year model and the inability to determine the constraints that are indeed binding, the full model will not always be consistent with the exogenous supply elasticities used for its calibration. Nevertheless, although a reduced constraint set may lead to a full model whose true implied elasticities are different from the elasticities calculated with equation (82) –unless land endowments and VCS are indeed the only binding constraints in the farm – the approach remains useful for calibration purposes. The reason is that it provides a minimum level of second-order structure to the calibrated model by setting bounds to its supply response.

In the case of feed PMP terms, a similar 'reverse engineering' approach to reconstruct a single matrix of feeding constraints cannot be implemented because constraints are split into those related to nutrient requirements (equations (7) and (8)) and those related to aggregate feeds (equations (9) and (10)). For this reason, calibration against feed price elasticities of demand from CAPRI can only be carried out myopically. However, the principle of recovering $Q_{i,m}^F$ separately from the linear term $d_{i,m}^F$, which was explained above in the case of activities, can still be applied for the feed PMP terms. This means that we use priors for feed demand elasticity to calculate $Q_{i,m}^F$, and then we impute this value in equation (80) – the FOCs for the feed input coefficient variable – in order to obtain an estimate of $d_{i,m}^F$.

The myopic calibration of $Q_{i,m}^F$ involves solving the FOCs (80)) for $\bar{\zeta}_{i,m}$:

$$\bar{\zeta}_{i,m} = -\frac{1}{Q_{i,m}^F \bar{x}_i}\left[10^{-3} p_m \bar{x}_i + d_{i,m}^F \bar{x}_i + \sum_v \left(\sum_n A_{n,m,v}^F \theta_{i,n,v}^F\right)\right]$$

and then differentiating with respect to p_m:

$$\frac{d\zeta_{i,m}}{dp_m} = -\frac{10^{-3}}{Q_{i,m}^F}$$

The own feed price elasticity of demand at $\bar{\zeta}_{i,m}$ is then defined as:

$$E_{i,m}^F = \frac{d\zeta_{i,m}}{dp_m}\frac{p_m}{\bar{\zeta}_{i,m}} = -\frac{10^{-3}}{Q_{i,m}^F}\frac{p_m}{\bar{\zeta}_{i,m}} \tag{83}$$

Parameter $Q_{i,m}^{F}$, which corresponds to the quadratic part of the feed PMP terms, can be directly calculated from equation (83):

$$Q_{i,m}^{F} = -\frac{10^{-3}}{\bar{E}_{i,m}^{F}} \frac{p_m}{\bar{\zeta}_{i,m}}$$

Finally, $d_{i,m}^{F}$ can be obtained by imputing $Q_{i,m}^{F}$ in equation (80). However, because the dual values of any binding feeding constraints are not known, $d_{i,m}^{F}$ is obtained from equation (80) without considering the term $\sum_{v}\left(\sum_{n} A_{n,m,v}^{F} \bar{\theta}_{i,n,v}^{F}\right)$.

2.2.3.2. The highest posterior density program for calibrating against supply elasticities

For each farm f, the HPD model minimises the weighted sum of normalised squared deviations, from prior information, of the estimated (i) farm-specific own-price supply elasticities; (ii) farm-specific CARA coefficient; and (iii) a farm-specific $Q_{i,j}$ matrix of the behavioural function (diagonal). The estimation procedure is applied for both crops and livestock activities. Using the algebraic notation introduced in Section 2.1, the general formulation of the HPD problem is as follows:

$$\min_{\varphi_f,\, E_{f,i,i},\, Q_{f,i,j} \geq 0} HPD_f = \left(\frac{\varphi_f - \bar{\varphi}_f}{\sigma_f^{\varphi}}\right)^2 + \sum_{i} \omega_{f,i} \left(\frac{E_{f,i,i} - \bar{E}_{f,i,i}}{\sigma_{f,i.i}^{E}}\right)^2 + \sum_{i} \omega_{f,i} \left(\frac{Q_{f,i,i} - \bar{Q}_{f,i,i}}{\sigma_{f,i,i}^{Q}}\right)^2 \tag{84}$$

which is subject to:

$$\begin{gathered} E_{f,i,j} = \left\{ \left(T_{f,i,j}\right)^{-1} - \sum_{t} \left[\sum_{i'} \left(T_{f,i',i}\right)^{-1} A_{f,i',t} \left(\sum_{i',j'} A_{f,t,i'} \left(T_{f,i',j'}\right)^{-1} A_{f,j',t} \right)^{-1} \sum_{i'} A_{f,t,i'} \left(T_{f,i',j}\right)^{-1} \right] \right\} \\ \times \left[\sum_{m} \frac{p_{f,m}}{\bar{x}_{f,i}} (1 - \xi_{f,m}) y_{f,j,m} \right] \\ \text{where } i', j' \in \text{set of production activities and are aliases of } i \end{gathered} \tag{85}$$

$$\begin{gathered} T_{f,i,j} = Q_{f,i,j} + \varphi_f \Omega_{f,i,j} \\ \text{where } Q_{f,i,j} = 0 \ \forall\, i \neq j \ (Q \text{ is diagonal}) \end{gathered} \tag{86}$$

$$\begin{gathered} \sum_{i'} T_{f,i,i'} \left(T_{f,i',j}\right)^{-1} = 1 \ \forall\, i = j \\ \sum_{i'} T_{f,i,i'} \left(T_{f,i',j}\right)^{-1} = 0 \ \forall\, i \neq j \\ \text{where } i' \in \text{set of production activities and is an alias of } i \end{gathered} \tag{87}$$

$$\begin{gathered} \sum_{i,j} A_{f,t,i} \left(T_{f,i,j}\right)^{-1} A_{f,j,t} \left(\sum_{i,j} A_{f,t',i} \left(T_{f,i,j}\right)^{-1} A_{f,j,t'} \right)^{-1} = 1 \ \forall\, t = t' \\ \sum_{i,j} A_{f,t,i} \left(T_{f,i,j}\right)^{-1} A_{f,j,t} \left(\sum_{i,j} A_{f,t',i} \left(T_{f,i,j}\right)^{-1} A_{f,j,t'} \right)^{-1} = 0 \ \forall\, t \neq t' \\ \text{where } t' \in \text{set of binding farm constraints and is an alias of } t \end{gathered} \tag{88}$$

where $\bar{\varphi}_f$ and σ_f^{φ} are the mean and the standard deviation for the CARA coefficient used as prior information; $\bar{Q}_{f,i,i}$ and $\sigma_{f,i,i}^{Q}$ are the mean and the standard deviation for the elements of matrix $\mathbf{Q}$ (assumed to be diagonal) used as prior information; and $\bar{E}_{f,i,i}$ and $\sigma_{f,i.i}^{E}$ are the mean and standard deviation of farm own-price elasticities of supply used as prior information. The normalised squared deviations for the supply elasticities and for the elements of matrix $\mathbf{Q}$ are weighted by the proportion of observed activity gross margins to total farm profits, $\omega_{f,i} = gm_{f,i}/\sum_i gm_{f,i}$, to allow the more profitable activities to dominate.

Equation (84) defines the HPD measure to be minimised, equation (85) is the elasticity formula (the algebraic equivalent of equation (82)), equation (86) defines matrix $\mathbf{T}$ as the sum of $\mathbf{Q} + \varphi\mathbf{\Omega}$, and equations (87) and (88) correspond to the inversion operations for matrices $\mathbf{T}$ and $\mathbf{AT}^{-1}\mathbf{A}'$, respectively, as required by the elasticity formula.

The prior value of the CARA coefficient, $\bar{\varphi}_f$, is specified relative to a common prior value of the relative risk aversion (RRA) across all farms. RRA is a measure of risk aversion that is invariant to the level of income and therefore it can be compared among farms of different economic size. Specifically, the prior value of RRA is set as equal to 1, so that $\bar{\varphi}_f = 1/z_f$, where z_f is the farm income observed in the base year. This follows the Arrow–Pratt definition of the ARA coefficient $\varphi_f = -U''(z_f)/U'(z_f)$ (Pratt, 1964), where $U(z_f)$ is the utility of income that is assumed to be exponential in IFM-CAP, as explained in Section 2.1. The selected prior RRA value of 1 is a typical average value often suggested by authors (Hardaker et al., 2004).

The elasticity prior $\bar{E}_{f,i,i}$ is set as equal to the CAPRI supply elasticity for the region where farm f belongs. Although using regional values as farm-specific priors is a strong assumption, the approach presents several advantages compared with the HPD model applied for the calibration of IFM-CAP v.1, which sought to approximate the prior as the weighted aggregate elasticity across all farms (Louhichi et al., 2018b). A first obvious advantage is that it is numerically simpler and computationally faster, as the program minimises the HPD metric per individual farm rather than for a group of farms. A second advantage is that the regional supply elasticity prior is better approximated, as the resulting farm elasticities are distributed more closely around the regional prior. In other words, the revised approach approximates the regional elasticity indirectly, by controlling the individual farm supply response, whereas the previous approach directly targeted the regional elasticity prior but did not control for the farm-specific supply response. However, because of the computational complexity of the HPD model that ran over a group of farms, the deviation from the regional prior was often significant, which also resulted in erratic farm-specific elasticities.

The prior information for $\bar{Q}_{f,i,i}$ is based on a simple regression model that uses all farms at NUTS 2 level to estimate a common (diagonal) matrix of implicit activity costs $B_{i,i}$, in which the activity FOCs for each farm–activity combination are the estimating equations:

$$Y_{f,i} = B_{i,i} + \varepsilon_{f,i} \tag{89}$$

where $\varepsilon_{f,i}$ are the residual terms, which can also be interpreted as the linear activity PMP terms $d_{f,i}$. The dependent variable $Y_{f,i}$ corresponds to the non-PMP part of the activity FOCs and is given by:

$$Y_{f,i} = \mathrm{E}\left[gm_{f,i}\right] - \bar{\varphi}_f \sum_j \bar{x}_{f,j}\Omega_{f,i,j} - \sum_t \mathrm{A}_{f,i,t}\bar{\theta}_{f,t} = 0 \qquad \forall\ i\ \in\ \text{set of crops activities} \tag{90}$$

$$Y_{f,i} = \mathrm{E}\left[gm_{f,i}\right] - \sum_{m \in F} d^F_{i,m}\bar{\zeta}_{i,m} - 0.5 \sum_{m \in F} \bar{\zeta}_{i,m} Q^F_{i,m} \bar{\zeta}_{i,m} - \bar{\varphi}_f \sum_j \bar{x}_{f,j} \Omega_{f,i,j} - \sum_t \mathrm{A}_{f,i,t} \bar{\theta}_{f,t} = 0$$

$$\forall\, i \,\in\, \text{set of animal activities}$$

The term $\mathrm{E}\left[gm_{f,i}\right]$ is the expected activity gross margin, specified in Section 2.1.1, while the feed PMP terms $d^F_{i,m}$ and $Q^F_{i,m}$ are calculated as explained in the previous section, using price elasticities of feed demand from CAPRI. The matrix of binding constraints for each farm, $\mathrm{A}_{f,i,t}$, refers to land endowments and VCS payments (the latter are only for animal activities) and the shadow prices $\bar{\theta}_{f,t}$ are set as equal to the land rental value reported in the FADN database and to the unitary VCS payment for i activity, respectively.

The regression model, defined by equations (89) and (90), requires that $B_{i,i}$ be non-negative. It is estimated in GAMS as a least squares process, that is, as an unconstrained program that minimises the sum of the squared residuals $\varepsilon_{f,i}$. Parameter $\bar{Q}_{f,i,i}$ for each farm is then calculated by scaling $B_{i,i}$ by the inverse of the observed activity vector at farm level – $\bar{Q}_{f,i,i} = B_{i,i}/\bar{x}_{f,i}$ (30) – while the linear PMP terms $d_{f,i}$ are set as equal to the residuals $\varepsilon_{f,i}$. This value of $d_{f,i}$ constitutes only an initial estimate, as it is later adjusted with a constrained model run, which adds the following calibration constraints to the IFM-CAP model that bound each activity level within the interval $[\bar{x}_{f,i} - Z^-, \bar{x}_{f,i} + Z^-]$:

$$x_{f,i} \leq \bar{x}_{f,i} + Z^- \; [\lambda_{f,i}] \tag{91}$$

$$x_{f,i} \geq \bar{x}_{f,i} - Z^- \; [\lambda_{f,i}] \tag{92}$$

where Z^- is a very small number that is used to disentangle the calibration constraints from the land constraint. As the model is non-linear, upper and lower bounds of activity levels are set to ensure that the observed situation is reproduced exactly. A similar set of calibration constraints can be found in Kanellopoulos et al. (2010) and Petsakos and Rozakis (2015). Every activity i is bound only by one of the two constraints, resulting in a dual value $\lambda_{f,i} \geq 0$ if the constraint in equation (91) is binding, and $\lambda_{f,i} \leq 0$ in the case of equation (92). This run modifies $d_{f,i}$ upwards or downwards, depending on which constraint is binding, so that its updated value is given by $d'_{f,i} = d_{f,i} + \lambda_{f,i}$.

---

(30) This scaling of $B_{i,i}$ to calculate the farm-specific $\bar{Q}_{f,i,i}$ is equivalent to $\bar{Q}_{f,i,j} = \left(1/\sqrt{\bar{x}_{f,i}}\right) B_{i,j} \left(1/\sqrt{\bar{x}_{f,j}}\right)$, which was used in IFM-CAP v.1 for the fully specified matrices $\mathbf{Q}$ and $\mathbf{B}$.

2.2.4. Baseline construction

The baseline is a reference situation that represents a mid-term projection of the European farming sector under the current policy setting. The IFM-CAP baseline is in line with the baseline of the Directorate-General for Agriculture and Rural Development, which refers to projections of the situation of the agricultural sector for 2030. Scenarios are counterfactuals of the baseline; the bulk of the data and the assumptions in the scenarios are identical to those of the baseline.

Below we present the various components of the baseline model. Each of these components contains a number of assumptions. Changing any of these assumptions will result in a different baseline; for example, instead of using the single FADN base year of 2017, the average of multiple years (e.g. 2014–2017) could be used, or additional activities could be more loosely defined.

2.2.4.1. Farm accountancy data network base year data

Some data are considered immutable across time (e.g. weights, total UAA and observed activities) and are loaded directly from the FADN database without any modifications. Other data are the result of more complex estimations, namely the data resulting from the calibration procedure (e.g. PMP terms and risk aversion coefficients) and those related to the feed requirements module (e.g. animal requirements and nutrient content). The production year that the above data refer to is termed the 'base year'; this is currently 2017.

2.2.4.2. Additional activities

Given that the baseline is about a mid-term projection, we need to assume that the activities of the farm may expand relative to the activities we observe in the base year. We make the following assumption: a farm can be engaged in a non-observed activity if that activity has been observed in its NUTS 3-TF14-ORGANIC cell (31) in more than 5 % of the total UAA of the cell. We also assume that all farms can be engaged in the FALL, PULS and SOYA activities, so as to follow the greening policy restrictions. The non-observed yields, prices and PMP parameters are estimated as the averages of the corresponding parameters of the NUTS 3-TF14-ORGANIC cell (32). The PMP linear terms of the additional activities are further adjusted so that in the base year their levels are zero.

2.2.4.3. CAPRI trends

To be in line with the Directorate-General for Agriculture and Rural Development projections for 2030, we incorporate the CAPRI trends and adjust the expected prices and yields for the base year. More specifically, we use output prices and yields from the CAPRI projections for the year 2030, which was taken as the time horizon for CAP greening scenario simulations. These projections are used to update the price and yield

---

(31) Two farms belong to the same NUTS3-TF14-ORGANIC cell if their properties with regard to the NUTS 3 unit, the type of farming (FADN code TF14) and organic status are the same for the base year.

(32) Without this modelling decision, farms would not be allowed to choose activities that are not observed in the same region and farm type, which may restrict their choices and thus also the simulated results. However, the set of activities observed in the same farm type and region is indicative of the probable feasible options that a farm faces when choosing a production structure. If activities are not observed in other similar farms, it indicates that they were probably not economically feasible because various unobserved factors (e.g. experience, skills, fixed costs and natural constraints) that are not accounted for in our model would make such an activity choice unprofitable. Hence, our approach for modelling non-observed activities partially accounts for unobserved factors that may impact farms' choices. A similar approach, but with more restrictive selection criteria, was used by Mahy et al. (2015), who consider the closest peers to address the self-selection problem. They select the closest peers based on the total farm area, crop area allocation, number of crops, geographical distance between farms and permanent grassland sgare; our approach, in contrast, is based on only two criteria, NUTS 2 region and farm type.

data of the base year, based on the trends. The CAPRI baseline is the best option for this adjustment for two reasons. First, it is developed in conjunction with the European Commission baseline and the projections present a consistent set of market and sectoral income prospects defined on the basis of specific policy and macroeconomic assumptions (Britz and Witzke, 2014; Himics et al., 2013). Second, the CAPRI nomenclature is consistent enough with that of IFM-CAP that a minimum number of assumptions is necessary for the transfer of the CAPRI projections to our model.

2.2.4.4.Inflation and technological change

We also adjust input costs based on inflation and technological change predictions Given the medium-term outlook on the prospects for agricultural markets and income (Directorate-General for Agriculture and Rural Development) (33), we assume an inflation rate of 1.6 % for input prices. We also assume a negative adjustment of those prices of 0.6 % due to the productivity growth of the technical factors (European Commission, 2016a).

2.2.4.5.Baseline policy

The baseline policy refers to the current 2014–2020 CAP policy. This section describes the key assumptions made for the baseline construction regarding the CAP policy

All FADN farms are assumed to **comply** with the definition of 'active farmer' and are therefore eligible for subsidies.

Entitlements and decoupled payments per hectare for individual farms are directly drawn from the 2017 FADN data set, as it contains payments according to the 2014–2020 CAP configuration. That implies that, although some CAP options are not endogenously modelled, they are captured by the heterogeneity of payments across farms. For example, although there are no explicit data to model the CAP regulation establishing that afforested areas and areas under the water framework directive, birds directive or habitats directive are eligible for decoupled payments if they had the right to a payment in 2008, the actual eligibility of individual farms will be reflected in the data for decoupled payments. The same applies for 'capping', as the decoupled payment data contain the 'capped' payments, if any. For this, for the construction of the baseline scenario, no alteration is made regarding the distribution of the decoupled payments to farms. For MSs that are applying convergence to the decoupled unit payment values, this may lead to small deviations, as the end of the convergence period is 3 years after the base year (2020 is the end of the convergence period and the base year is 2017).

Regarding **coupled payments**, as extensively discussed in Section 2.2.1.8, these are drawn from the FADN data set and it is assumed that their distribution will remain the same in the baseline scenario.

The **young farmers scheme** and the **small farmers scheme** are not modelled in IFM-CAP. There are no data available in the FADN database to accurately assess the young farmers scheme (34), while the small farmers scheme is a voluntary measure that is not easily modelled in the current version of IFM-CAP. In terms of budget, these two schemes are minor. The young farmers scheme may represent only up to 2 % of the direct payment budget, while the small farmers scheme is applied in 15 MSs and is disbursed only to (small) farmers applying for it, which is difficult to assess (European Commission, 2017).

Under the **CAP greening measures**, 30 % of direct payments are conditional on complying with three mandatory requirements: (i) crop diversification for arable crops;

---

(33) https://ec.europa.eu/info/food-farming-fisheries/farming/facts-and-figures/markets/outlook/medium-term_en

(34) The FADN database does not contain information on whether a farmer has become the head of the farm for the first time.

(ii) maintenance of permanent grassland; and (iii) allocation of land to EFAs. Following EU regulations (EU, 2015; European Parliament and Council, 2014a), the modelling of greening measures assumes full compliance with the three greening measures without allowing farmers to trade off income reductions with full compliance against direct payment reductions as a consequence of partial or full non-compliance. Most studies in the literature model full compliance with CAP greening requirements (e.g. Cortignani and Dono, 2015; Gocht et al., 2017; Mahy et al., 2015; Was et al., 2014); a few allow farmers to choose the level of (non-)compliance (e.g. Cortignani et al., 2017; Solazzo and Pierangeli, 2016; Vosough-Ahmadi et al., 2015). As shown in Table 17, the crop diversification measure applies only to farms with an arable area greater than 10 ha. Farms with more than 75 % of their total eligible land covered by grassland and farms with 75 % of their arable area cultivated with forage are also not subject to the crop diversification measure. Furthermore, there are stricter requirements for farms having more than 30 ha of arable land (group 2) than for farms with between 10 and 30 ha of arable land (group 1). Farms in the latter group need to have at least two different crops and the main crop should not exceed 75 % of the arable land. Farms in the former group are required to have at least three crops; the main crop should not cover more than 75 % of the arable land and the two main crops together should not cover more than 95 % of the arable land.

Table 17. Crop diversification measure as implemented in IFM-CAP

	Exempted farms	**Farms group 1**	**Farms group 2**
Arable land (*AL*)	< 10 ha*	10–30 ha	≥ 30 ha
Minimum number of cultivated crops	—	2	3
Maximum proportion of main crop in *AL* (%)	—	75 %	
Maximum proportion of two main crops in *AL* (%)	—	—	95 %

NB: *Farms are excluded if (i) fodder area + fallow area ≥ 75 % of *AL* and *AL* – (fodder + fallow) < 30 ha; (ii) grassland + other herbaceous fodder crops > 75 % of the UAA and *AL* – other herbaceous crops < 30 ha; or (iii) the farming is organic.

Source: Compiled based on Regulation No 1307/2013 (European Parliament and Council, 2013), Commission Delegated Regulation No 639/2014 (European Parliament and Council, 2014a) and Commission Delegated Regulation No 640/2014 (European Parliament and Council, 2014b).

Under the maintenance of permanent grassland measure, the ratio of grassland to total agricultural area cannot decrease by more than 5 % compared with the reference ratio in 2017. Moreover, under this measure, farms are prevented from ploughing and converting permanent grassland in areas designated by MSs as environmentally sensitive (35).

The calculation of the reference ratio can be applied at national, regional or subregional level: 23 MSs apply it at national level, four MSs do so at regional level and one MS is without permanent grassland (Malta). If the ratio of grassland to total agricultural area has decreased by more than 5 % at the national or regional level (depending on implementation), the obligation needs to be imposed at farm level (European Parliament and Council, 2013, 2014a,b).

We use 2017 as the reference year for modelling the grassland measure, as this is the IFM-CAP base year. That is, we calculate the ratio of grassland to total agricultural area

---

(35) These areas can be located inside or outside Natura 2000 areas.

for 2017 and compare it with the ratio at baseline (2030). If in an MS or region (depending on implementation) the ratio decreases by more than 5 % at baseline relative to the base year, we impose the obligation at farm level in the greening scenario.

Two categories of grassland are modelled in IFM-CAP: permanent grassland and rough grazing area. Permanent grassland is assumed to be fully replaceable with arable land if relative returns change, while the rough grazing area is assumed to be fixed, as this type of land is usually of low quality. Both grassland categories are assumed to be subject to the grassland measure in the greening scenario.

In the case of environmentally sensitive areas, we consider that grassland located in a Natura 2000 area is subject to the grassland measure of no conversion to arable land.

The EFA measure requires farms with more than 15 ha of arable land to allocate at least 5 % of that land (excluding areas under grassland) to an EFA. The areas that qualify as EFAs include land left fallow, terraces, landscape features, buffer strips, agroforestry, areas with short rotation, afforested areas, catch crops and nitrogen-fixing crops (Table 18) (European Parliament and Council, 2013, 2014a). MSs can choose which land elements they classify as eligible for EFA status. As reported in Table 18, land cultivated with nitrogen-fixing crops is the most common type of EFA-eligible area across MSs (27 MSs), followed by fallow land (26 MSs) and areas with short rotation (20 MSs). The eligible land elements have different weights in contributing to EFA levels (varying between 0.3 and 30), depending on their conversion and weighting factors (36).

The EFA measure is the most challenging measure to model, as no data are available that enable us to capture different eligible land elements. Because of missing data in the FADN database, only the following four elements of EFAs are considered in IFM-CAP: fallow land (including voluntary set-aside), afforested areas, catch crops and nitrogen-fixing crops. Fallow land and nitrogen-fixing crops are endogenous activities in IFM-CAP. Forests and catch crops are not endogenously modelled in IFM-CAP and, therefore, their areas are set as equal to the base year level. EU regulations specifies the list of crops that can be considered catch crops / green cover or nitrogen-fixing crops. Given that in the IFM-CAP model some minor activities are aggregated, they cannot be mapped exactly to this list of eligible crops. Therefore, we assume that all cereals and pulses can be considered catch crops and that pulses and soya can be considered nitrogen-fixing crops. MSs with more than 50 % of their land surface area covered by forest may decide that the EFA measure will not be applied in areas in which more than 50 % of the land surface area at local administrative unit (LAU) 2 level (37) (or other contiguous geographical area) is covered by forest and the ratio of forest land to agricultural land is higher than 3:1. This forest exemption is applied in Estonia, Latvia, Finland and Sweden. Given that the FADN (and IFM-CAP) does not include any information at LAU 2 level, the forest exemption is assessed at farm level, but only for farms located in NUTS 3 regions in which the exemption is applied (European Commission, 2016b).

It is important to note that MSs can change the elements that are eligible to be counted as EFAs on a yearly basis. Table 18 reports the notifications applied in 2016, which correspond to the assumptions used in the modelling of CAP greening in IFM-CAP.

---

(36) As in the case of the crop diversification measure (Table 17), farms with more than 75 % of their total eligible land covered by grassland and farms with 75 % of their arable area cultivated with forage are not subject to the EFA measure.

(37) An LAU is a low-level administrative division of an MS. There are two levels of LAUs: LAU 1 and LAU 2.

Table 18. Land elements eligible as EFAs

Eligible area	No of implementing MSs	Conversion factor	Weighting factor	Modelling in IFM-CAP
Fallow land	26	n.a.	1	Yes
Terraces	8	2	1	No
Hedges or wooded strips	13	5	2	No
Isolated trees	13	20	1.5	No
Trees in line	16	5	1.5	No
Trees in groups	17	n.a.	1.5	No
Field margins	16	6	1.5	No
Ponds	12	n.a.	1.5	No
Ditches	15	3	2	No
Traditional stone walls	7	1	1	No
Other landscape features under GAEC or SMR	11	n.a.	1	No
Buffer strips	17	6	1.5	No
Agroforestry	11	n.a.	1	No
Strips along forest edges (no production)	9	6	1.5	No
Strips along forest edges (with production)	6	6	0.3	No
Areas with short rotation	20	n.a.	0.3	No
Afforested areas	14	n.a.	1	Yes
Catch crops or green cover	19	n.a.	0.3	Yes
Nitrogen-fixing crops	27	n.a.	0.7	Yes

NB: GAEC, good agricultural and environmental conditions; n.a., not applicable; SMR, statutory management requirements.

Source: Compiled based on EU regulations (European Parliament and Council, 2014a, 2015).

2.2.4.6. Validation of the baseline scenario

We have performed a validation of the baseline model, examining the effect of each of the above assumptions to the baseline results. More specifically, In Table 19, we are summarizing the scenarios run for validating the baseline. The 'by/cl' scenario contains only the baseyear and the calibration data. This is expected to be almost identical to the baseyear data. The 'by/cl/a', 'by/cl/c' and 'by/cl/g' scenarios isolate the additional activities, CAPRI trends and greening assumptions. These scenarios will provide the effect of each of the above assumptions to the baseyear data. The rest of the scenarios are combinations of the previous scenarios and are useful for seeing the interaction effects of the baseline assumptions. The 'by/cl/a/c/g' scenario is the actual baseline model that includes all three elements.

In Table 20 and Table 21, we provide the results of the model for the baseline assumptions. As expected, the calibrated model 'by/cl' provides almost identical results to the baseyear ('by') values. Also, the inclusion of additional activities (by/cl/a) and of greening constraints (by/cl/g) does not have a significant impact in the results. Results change in comparison to the baseyear when we introduce price changes from CAPRI trends. Nevertheless, the changes are reasonable and thus we conclude that the baseline construction is valid.

Table 19. Scenarios run for the validation of the baseline (2017 baseyear)

Scenario	Additional Activities	CAPRI Trends	Greening Constraints
by/cl			
by/cl/a	+		
by/cl/c		+	
by/cl/g			+
by/cl/a/g	+		+
by/cl/c/g		+	+
by/cl/a/c	+	+	
by/cl/a/c/g	+	+	+

Note: 'by'=2017 baseyear / 'cl'=calibration / 'a'=additional activities / 'c'=capri trends / 'g'=greening constraints

Table 20. Estimated areas for the various stages of the baseline in EU (th. ha)

	by	by/cl	by/cl/g	by/cl/a	by/cl/a/g	by/cl/c	by/cl/c/g	by/cl/a/c	by/cl/a/c/g
DWHE	2,560	2,560	2,136	2,543	2,452	2,348	2,035	2,186	2,109
SWHE	22,166	22,166	17,413	21,729	19,705	20,945	16,672	19,255	17,661
BARL	11,961	11,961	9,731	11,925	11,371	13,694	11,048	13,747	13,019
MAIZ	7,795	7,795	6,374	7,855	6,809	8,311	6,604	8,606	7,433
RYEM	2,109	2,109	1,770	2,156	2,106	2,321	1,922	2,239	2,164
OATS	2,558	2,558	2,118	2,564	2,478	2,293	1,923	1,825	1,762
PARI	504	504	548	510	519	509	547	540	560
OCER	4,243	4,243	3,510	4,195	4,068	3,008	2,620	2,178	2,084
PULS	3,019	3,019	2,348	3,071	2,846	2,771	2,274	2,777	2,554
RAPE	7,173	7,173	5,499	7,432	6,658	6,703	5,110	7,653	6,723
SUNF	4,535	4,535	3,677	4,508	3,604	4,388	3,506	4,326	3,312
SOYA	915	915	732	944	790	742	637	680	552
OOIL	251	251	207	270	258	340	278	471	450
SUGB	1,760	1,760	1,519	1,670	1,598	907	904	802	778
POTA	1,217	1,217	1,120	1,314	1,257	1,474	1,309	1,782	1,697
TEXT	372	372	344	380	350	338	318	314	296
TOBA	85	85	78	96	101	88	80	108	104
OCRO	418	418	322	460	302	491	379	556	386
OIND	448	448	388	643	594	598	493	858	793
MAIF	5,692	5,692	3,724	5,686	5,409	8,560	5,063	11,416	10,230
PGRA	29,331	29,331	36,248	28,910	29,579	26,346	34,244	23,020	24,214
RGRA	6,498	6,498	6,498	6,498	6,498	6,498	6,498	6,498	6,498
OFAR	14,676	14,676	22,671	14,675	16,445	17,389	25,057	19,439	21,348
VGOF	1,109	1,109	993	1,225	1,099	1,109	996	1,197	1,085
VGMG	346	346	314	436	314	357	321	435	332
VGUG	119	119	161	134	138	122	164	130	134
APPL	595	595	679	610	617	620	704	674	682
CITR	410	410	421	409	414	428	440	430	435
BERR	252	252	282	252	253	266	296	271	272

NUTS	1,175	1,175	1,280	1,186	1,197	1,033	1,146	934	945
PEAC	230	230	237	235	236	206	215	170	171
OLIV	3,184	3,184	3,415	3,181	3,230	2,973	3,221	2,511	2,549
OFRU	436	436	512	444	457	424	501	445	459
TABO	255	255	274	266	266	288	306	440	440
TAGR	74	74	97	76	84	81	104	115	123
TWIN	2,734	2,734	3,034	2,759	2,811	2,951	3,231	2,997	3,047
FLOW	59	59	53	57	40	59	53	58	40
NURS	69	69	69	68	68	69	69	68	68
FALL	4,349	4,349	4,883	4,311	8,660	3,632	4,391	3,529	8,172
	by	by/cl	by/cl/g	by/cl/a	by/cl/a/g	by/cl/c	by/cl/c/g	by/cl/a/c	by/cl/a/c/g

Table 21. Estimated number of animals for the various stages of the baseline in EU

ACT	by	by/cl	by/cl/g	by/cl/a	by/cl/a/g	by/cl/c	by/cl/c/g	by/cl/a/c	by/cl/a/c/g
DCOW	22,300	22,099	22,094	22,103	22,111	30,816	30,775	30,826	30,833
SCOW	11,304	11,230	11,232	11,228	11,229	12,143	12,142	12,141	12,141
HEIR	9,002	8,984	8,984	8,984	8,984	9,795	9,792	9,795	9,795
HEIF	6,609	6,582	6,583	6,599	6,598	8,569	8,551	8,586	8,586
BULF	7,024	7,001	7,004	7,041	7,042	8,458	8,435	8,498	8,498
CAMR	18,682	18,433	18,432	18,433	18,432	16,361	16,378	16,361	16,361
CAFR	4,562	4,553	4,552	4,553	4,553	4,143	4,145	4,143	4,143
CAMF	1,024	1,004	1,005	1,004	999	863	868	863	864
CAFF	1,024	1,000	1,005	999	1,004	986	989	986	986
SOWS	12,566	10,191	10,191	10,190	10,190	18,945	17,855	18,941	18,937
PIGF	85,849	74,612	74,627	74,595	74,599	87,501	85,004	87,475	87,476
SHGM	61,664	58,583	58,415	58,524	58,503	60,047	60,007	60,118	60,103
SHGF	26,504	23,621	23,554	23,609	23,555	39,390	39,236	39,403	39,397
HENS	276,224	189,138	188,069	188,911	189,653	336,970	331,503	336,566	337,530
POUF	967,896	729,652	729,386	731,016	730,388	936,570	923,113	938,173	937,811
OANI	3,164	2,873	2,873	2,873	2,873	2,873	2,873	2,873	2,873
ACT	by	by/cl	by/cl/g	by/cl/a	by/cl/a/g	by/cl/c	by/cl/c/g	by/cl/a/c	by/cl/a/c/g

2.2.5. Environmental Indicators

The environmental indicators are used to further process the model's results and add the environmental effects of the evaluated policies. Below we briefly describe the relevant indicators for greenhouse gas emissions, nitrogen, input use, biodiversity and soil quality. More detailed information can be found in Bielza et al. (2015), Bielza et al. (2017) and Bielza et al. (2021).

2.2.5.1. GHG and N indicators

The following indicators are currently implemented in IFM-CAP:

- **CH4TOT:** GHG methane emissions = CH4ENT + CH4MAN + CH4RIC (in t CO2eq).
- **N2OTOT:** GHG Nitrous oxide emissions = N2OGRA + N2OAPP + N2OMAN +N2OSYN + N2OHIS + N2OCRO + N2OLEA + N2OAMM (in t CO2eq)
- **GWPA:** GHG total emissions = GH4+N2O (in t CO2eq).
- **NH3:** Ammonia emissions =NH3GRA + NH3APP + NH3MAN + NH3SYN (in t N)
- **FB:** Farm balance (N surplus) = FBH + FBA (in t N)

All of them can be expressed in total amount, t per hectare and kg per kg.

2.2.5.2. Input use indicators

The FADN database for the year 2012 lacks of data on intermediate inputs in physical quantities such pesticides and fertilizers(38), so that proxies on input use(39) have to be applied in the structure of the IFM-CAP model. The following four indicators are currently implemented (Bielza, 2017):

- **EINP:** Average annual expenditure on inputs per ha - proxy of agricultural intensification
- **EPLA:** Pesticide expenditure per ha - proxy of pesticide risk
- **EFER:** External fertilizer expenditure per ha - proxy of fertilizer consumption
- **FERQ:** Quantity of fertilizers and soil improvers – proxy of fertilizer application (kg per ha)

The first three indicators resulting from dividing the corresponding input expenditure on total inputs, pesticides or fertilizers by the UAA, while FERQ approximates input quantities by using a standard fertilizer price (Bielza, 2017).

2.2.5.3. Biodiversity

The IFM-CAP model contains three indicators related to biodiversity on position: CRICH, CDIVE and CDIVE_ha. The indicators are calculated using functional crop groups. A functional crop group is defined as a set of crops that are homogeneous from a biodiversity support perspective.

The biodiversity indicators are calculated as follows:

- **CRICH**: Number of functional crop groups cultivated on a farm on a per hectare basis. Crops considered in the indicator are annual and perennial crops. Crops considered in the indicator are annual and perennial activities. Permanent grasslands and rough grazing pastures are excluded.

---

(38) From 2014 onwards, the FADN unit started to collect data on NPK input use, which is still being processed and validated.

(39) The use of these types of proxies were suggested by Westbury et al. (2011) and the FLINT project (http://www.flint-fp7.eu/).

  - The indicator is calculated by dividing the number of functional crop groups per farm (M) by the sum of the respective areas (Aj):

$$\mathrm{CRICH} = \frac{\mathrm{M}}{\sum_{\mathrm{j=1}}^{\mathrm{M}} \mathrm{Aj}}$$

- **CDIVE**: Shannon diversity index calculated on crop groups shares at farm level. Crops considered in the indicator are annual and perennial crops. Permanent grasslands and rough grazing pastures are excluded. Crops grown under shelter are also excluded.
  - The indicator is estimated as follows

$$\mathrm{CDIVE} = -\sum_{\mathrm{f=1}}^{\mathrm{M}} (\mathrm{cj} \cdot \ln \mathrm{cj})$$

    where M is the number of functional crop groups in the farm and where *cj* is the share of the j-th functional crop group calculated as:

$$\mathrm{cj} = \frac{\mathrm{A_j}}{\sum_{\mathrm{j=1}}^{\mathrm{M}} \mathrm{A_j}}$$

- **CDIVE_ha:** Shannon diversity index per hectare. This indicator can be computed at farm level and should be aggregated at regional / country level by weighted average using total farm areas and FADN weighing factors.

The last two indicators are complementary. In order to make scenario comparison or geographic comparison at the aggregate level, both indicators should be higher (or lower) to determine that the new situation is better (or worse) than the baseline. If one indicator is higher and the other is lower, then it will not be possible to draw clear conclusions for biodiversity.

2.2.5.4.Soil quality and management indicators

The IFM-CAP model at its current stage includes two indicators referring to soil quality and management:

Crop cover effect on soil erosion: Conventionally, soil erosion is estimated using the well known Revised Universal Soil Loss Equation (RUSLE) equation, which estimates soil losses as a function of several biophysical factors such as erosivity of the eroding agents (mainly water), erodibility of the soil, slope steepness and slope length of the land, land cover, stoniness and human practices designed to control erosion. There is no farm-level information at EU-28 level about most of these factors (e.g. rainfall, slope, soil type, soil erodibility…), so that the RUSLE equation cannot be fully used. To overcome this limitation, a proxy based on crop cover was proposed by Bielza (2017).

The crop cover erosion indicator (**EROSF_CP**) is calculated as the product of RUSLE cover-management factor (C-factor) and soil conservation and support practices factors (P-factor), which are part of the full RUSLE equation. EROSF_CP is calculated as follows:

$$\mathrm{EROSF\_CP^f} = \sum_{\mathrm{i=1}}^{\mathrm{I}} \left[\mathrm{C_{ci}^{NUTS2}} \times \frac{\mathrm{Aar_i^f}}{\mathrm{TEA^f}} \times \mathrm{C_M^{NUTS2}} \times \mathrm{P^{NUTS2}} \right] + \sum_{\mathrm{j=1}}^{\mathrm{J}} \left[\mathrm{C_{cj}^{NUTS2}} \times \frac{\mathrm{Aper_j^f}}{\mathrm{TEA^f}} \times \mathrm{P^{NUTS2}} \right]$$

where the $\mathrm{Aar_i}$ is the arable crops area (including fallow land but not crops under shelter) in farm f, I the number of arable crops in the farm, $\mathrm{Aper_j}$ is the area of permanent crops and permanent grasslands (including rough grazing) in farm f, J the number of permanent crops and grasslands in the farm, TEA the total erodible area (the utilised agricultural area of the farm except those of crops under shelter), $\mathrm{C_C}$ the crop-specific cover factor(40), P the support practices factor and C_M the management factor, which is calculated as follows:

$$\mathrm{C}_{\mathrm{M}}^{\mathrm{NUTS2}} = \mathrm{C}_{\mathrm{Mtillage}}^{\mathrm{NUTS2}} \times \mathrm{C}_{\mathrm{Mcatchcrop}}^{\mathrm{NUTS2}} \times \mathrm{C}_{\mathrm{Mresidues}}^{\mathrm{NUTS2}}$$

The IFM-CAP model also has two other intermediate erosion factors on position:

- **EROSF_Cc** is only based on C- crop-specific factors
- **EROSF_Ccm** also includes soil management factors defined at NUTS2 level.

Bielza et al. (2017) reports the full description and values of the factors applied in the IFM-CAP model. Most of the factors have been obtained from the European Soil Data Centre (ESDAC(41)) and further information can be found in Panagos et al. (2015a and b).

- The share of permanent grassland, **SGRA**, is used as a proxy for soil organic carbon content, and it is calculated as follows:

$$SGRA = \frac{Arg + Aug + Amp}{UAA}$$

Using FADN on rough grazing area (Arg), unused permanent grassland with DP (Aug), Area of meadows and permanent pastures (Amp) being defined as the grassland grown for 5 years or more on cultivated land and Total Utilised Agricultural Area (UAA). Note that rough grazing areas are fixed in the IFM-CAP model.

---

(40) Crops have been grouped depending on their physical and phenological features which can increase or decrease soil erosion (i.e. soil coverage, root structure, phenological stages, etc.).

(41) esdac.jrc.ec.europa.eu, European Commission, Joint Research Centre

3. Code design and implementation

The IFM-CAP model is a large-scale farm model. It consists of more than 100 code files and several gigabytes of data. Additionally, it was developed for continuous use and its size is expected to grow as requests for policy evaluations accumulate.

For these two reasons, efficient organisation of the code is necessary to produce a model that is error-free, easy to maintain and extend. An efficient design is a prerequisite for having a model that can grow while keeping the complexity manageable.

In this section, initially we describe the design principles for coding IFM-CAP and then we provide guidance on how to install, run and further develop the model.

3.1. Conceptual basis

3.1.1. The farm model abstraction

We have based the IFM-CAP2 coding on a farm model abstraction as in Figure 17. The farm **optimisation problem** is at the core of the design. It contains all parameters, variables and equations that the modeller assumes that represent in a satisfactory way the decision-making process of the farm. The **parameter values** (i.e. the data) are at its perimeter. They include individual characteristics of the farm (e.g. the costs of the farm's activities) and general parameters (e.g. policy parameters that apply to all farms). A part of those values will be the result of **data estimation procedures**. These procedures form the outer layer.

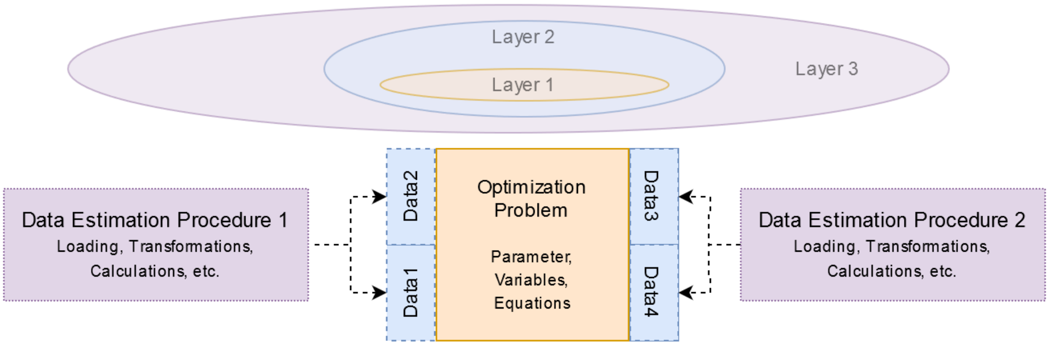


Figure 17. An abstract view of a farm model

It is a core design choice that this logical partitioning of a farm model is also represented in the IFM-CAP code. The reflection of this representation in the code makes the model more transparent and facilitates its verification. '**What the model does**' can be understood by reviewing the behavioural component (layer 1), disregarding the other two layers. '**What data are the model using**' can be easily answered by examining the data (layer 2), disregarding the optimisation problem and/or data estimation procedures. '**How the data were derived**' can be assessed by focusing on the data operations (layer 3). Thus, the comprehension of the model can be broken into smaller parts and is easier to do. Furthermore, when these three layers are kept separate, it is easier to undertake debugging and error tracing.

3.1.2. The optimisation problem concept

We define an optimisation problem using the following **conceptual elements** (42).

1. **Variables.** This represents the set of variables (i.e. which variables are contained in the problem).
2. **Equations.** This represents the set of equations and their mathematical formulation (i.e. which equations are contained in the problem and their exact mathematical form).
3. **Parameters.** This represents the set of parameters (i.e. which parameters the problem contains without any reference to the actual data that are assigned to them).
4. **Parameter data.** These are the values assigned to the parameters (43).
5. **Solve.** This represents the final matrix of the problem. It is the actual numerical matrix that is passed to the solver.

Changing one of those elements (e.g. adding a new variable or modifying the parameter data) will result in a different optimisation problem (i.e. a different final matrix). However, some elements will trigger more extended changes to the problem and other will have limited effect.

More specifically, altering the set of variables or the set of parameters (i.e. adding or removing a variable or a parameter from the problem) will necessarily trigger a change in the equation element (adding or altering an equation; otherwise, it was meaningless to add or remove a variable or parameter). However, a change in the equation element (e.g. adding, removing or changing the definition of an equation) will not necessarily require a change in the set of variables or the set of parameters. In addition, a change in the parameter set will require a change in the parameter data, while the reverse does not hold; changing the value of a parameter does not necessarily require a new parameter. In addition, any change in the equations or the parameter data will result in a different final matrix of the problem. The same holds for a change in the set of variables or the set of parameters, as this will trigger a change in the equations and/or the parameter data.

The above relationships regarding the changes triggered to the optimization problem by changes in its element can be organised in a hierarchical relationship in Figure 18. An arrow from (A) → (B) signifies that any change in the element (A) will require a change in element (B) (if we add a new variable, we need to add or change an equation), although the opposite is not necessarily true (if we add an equation, it does not necessarily mean we need a new variable)

---

(42) Without a loss of generality, the naming convention is related to the GAMS modelling language.

(43) This definition also contains the numerical values of any attributes of the variables and the equations (upper and lower bounds, etc.).

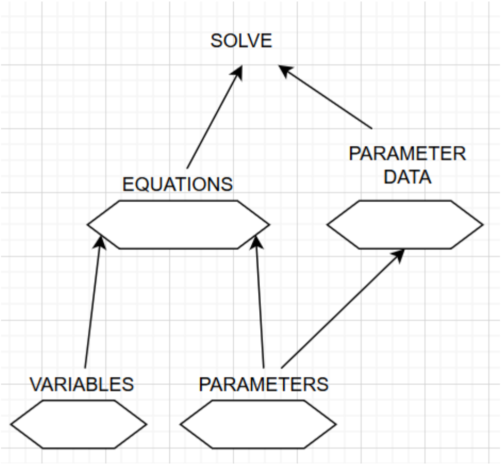


Figure 18. The hierarchical relationship of the elements of the optimisation problem. An arrow from (A) → (B) signifies that any change in the element (A) will require a change in element (B) (if we add a new variable, we need to add or change an equation), although the opposite is not necessarily true (if we add an equation, it does not necessarily mean we need a new variable).

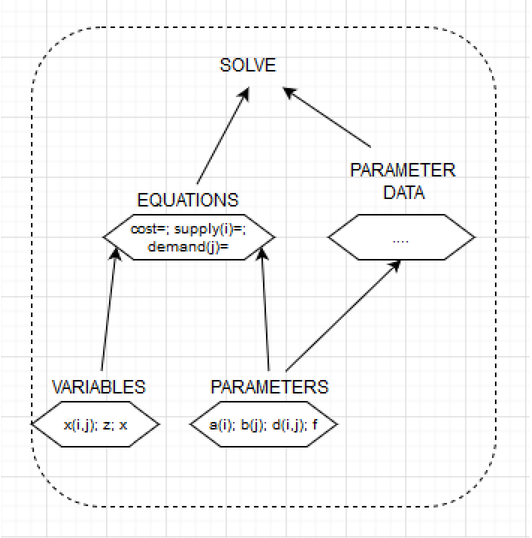
To make the above definitions more concrete, we provide the example of the classical transport problem in Box 1. The code is shown in part A and the conceptual elements of the optimisation problem are shown in part B.

We use this hierarchy in order to partition IFM-CAP code more efficiently, as explained in a later section.

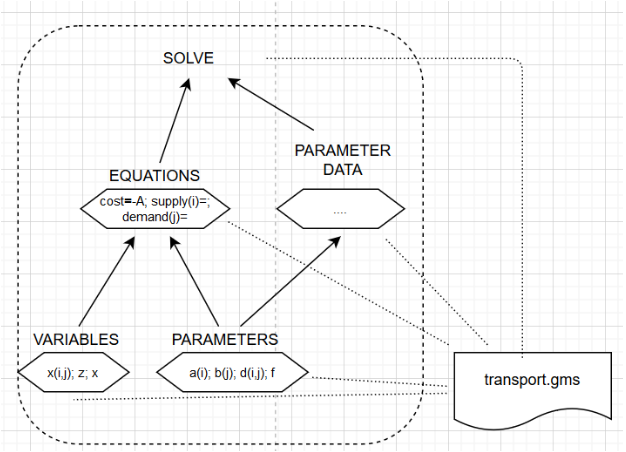
Box 1. Example of a dependence diagram of a language program

```
Set
i 'canning plants' / seattle, san-diego /
j 'markets' / new-york, chicago, topeka /;

Parameter
a(i) 'capacity of plant i in cases'
/ seattle 350
 san-diego 600 /

b(j) 'demand at market j in cases'
/ new-york 325
 chicago 300
 topeka 275 /;

Table d(i,j) 'distance in thousands of miles'
 new-york chicago topeka
seattle  2.5 1.7 1.8
san-diego 2.5 1.8 1.4;

Scalar f 'freight in dollars per case per thousand miles' / 90 /;

Parameter c(i,j) 'transport cost in thousands of dollars per case';
c(i,j)  = f*d(i,j)/1000;

Variable
x(i,j) 'shipment quantities in cases'
z 'total transportation costs in thousands of dollars';

Positive variable x;

Equation
cost 'define objective function'
supply(i) 'observe supply limit at plant i'
demand(j) 'satisfy demand at market j';

cost.. z =e = sum((i,j), c(i,j)*x(i,j));

supply(i).. sum(j, x(i,j)) =l = a(i);

demand(j).. sum(i, x(i,j)) =g = b(j);

Model transport / all /;

solve transport using lp minimizing z;

display x.l, x.m;
```

A: Code of the transport problem



Part B: Conceptual elements



Part C: File organisation

3.1.3. Reproducibility

Reproducibility refers to the ability of the scientific community to obtain the same results as the originators of some specific scientific findings (Lusoli, 2020).

In the case of IFM-CAP, one obstacle for reproducibility is the availability of data. Our data are based on individual FADN farm data and there is a confidentiality agreement between the JRC and the Directorate-General for Agriculture and Rural Development that prohibits the sharing of the data. This makes it impossible for a third party to reproduce the results. However, we are sharing the model's code and, in the future, we may release some artificial farm data that will allow independent researchers to experiment and validate the model's operation.

Beyond this, we apply three general principles for making a research project reproducible, as described in Kitzes et al. (2017):

1. clearly separate, label and document all data, files and operations that occur on data and files;
2. document all operations fully, automating them as much as possible and avoiding manual intervention in the workflow when feasible;
3. design a workflow as a sequence of small steps that are glued together, with intermediate outputs from one step feeding into the next step as inputs.

A related issue is that of **literate programming**, a programming paradigm introduced by Donald Knuth. Knuth's key message was that code should be made not only machine readable (or executable by the computer) but also human readable. This means that the logical structure of the code should reflect a narrative that resembles that of a paper, as opposed to a set of somewhat ad hoc instructions that reflect the order in which the analysis took place (de la Guardia and Sturdy, 2019). We adhere to this approach by applying sufficient commenting and also employing the R Markdown language for several data-related operations.

3.1.4. Modularity

Modularity is a concept that has gained popularity in recent decades across a variety of professional and disciplinary settings. Russell (2012) provides a thorough description of the concept.

> *Modularity describes specific relationships between a whole system and its particular components. A modular system consists of smaller parts (modules) that fit together within a predefined system of architecture. Modules feature standardized interfaces, which facilitate their integration with the overarching system architecture. A key feature of each module is that it should encapsulate (or 'black box') its messy internal details, thus masking technical, organizational, cultural, and political conflicts to display only a consistent interface. The designers of modular systems are therefore able to swap modules in a 'plug-and-play' manner, which increases the system's flexibility.*

In the software engineering domain, modularisation was embraced from the beginning and has been the subject of extensive and ongoing research (van der Hoek and Lopez, 2011). Quite early, Parnas (1972) established the fundamental principle of **information hiding**, which refers to reducing the information that a module allows other modules to access. The second related principle is that of **low coupling and high cohesion**, introduced by Stevens et al. (1974), which refers to the dependence of one module on the others to the least extent possible (coupling) and the existence of strong dependence in the elements of a module (cohesion). A derived principle is that of **separation of concerns**, which refers to decomposing a computer program in such a way that different concerns or aspects of the problem at hand are addressed in distinct modules (Dijkstra, 1982).

Obviously, modularisation is a desired property of any design but there are two obstacles to achieving it. The first is related to the **cross-cutting aspects/concerns** that a problem has. There is a limit to the degree of low coupling related to the nature of the problem. The second is that **there is an unlimited set of possible modularisations** and the designer has to select one. Thus, it is a design challenge to identify the optimal set of modules.

The need for modularity comes from the fact that, usually, a model is continuously evolving. If we had to develop a one-time-use model, a non-modular solution would be more efficient. However, the time and effort involved in continuously adjusting a monolithic model grows exponentially and, after a certain point, rewriting the model from scratch may take less time than modifying it. On the other hand, a modular model is easily adaptable to extensions and modifications.

It is often the case that the word 'modular' is misunderstood. Calling a model 'modular' merely on the basis that it provides varied functionality or that the code lies in different files is not correct. These 'modules' are usually split into different files but their 'software logic' is still very much interlinked, so that from a software point of view they are a single monolithic entity. A non-modular model, although it may have 'modules', requires too much effort and time to understand, maintain and extend. The **modularity** of the model is a qualitative property of logic design, not of how the files are split.

A model that is, in fact, modular has a structure such that understanding it is easy and extending or modifying it does not create **ripple effects** (i.e. the need to change the code in many other parts of the model), and thus extending or modifying it does not require too much effort. The properties of a modular model are as follows.

1. **Transparency.** The model can be reviewed module by module, facilitating overall comprehension and quality control.
2. **Maintainability.** Code and database updates of a module do not affect other modules.
3. **Extensibility.** Modules can be extended or added to the core model without

affecting other parts of the model.

4. **Distributed development.** Modellers focus on specific modules, which facilitates coordination of the coding efforts.

There are already some well-recognised and mature software development methods available for increasing the modularity of software code.

1. **Divide and conquer.** This principle dictates the decomposition of a large and extended problem into smaller self-sustained subproblems. Handling the series of smaller subproblems is less complex than dealing with the whole problem at once.
2. **Separation of concerns.** This is a design principle for separating a computer program into distinct sections, so that each section addresses a separate concern.
3. **Data flow-oriented design.** In the data flow-oriented design approach, the design is information driven. The program structure follows that of the information flow and the emphasis is on the processing or operations performed on the data.

3.2. Coding guidelines

The organisation and the structure of the code have the objective of making the model modular, transparent and easy to debug. More specifically, considering the abstractions of the farm models and the optimisation problem, as analysed in the previous section, and the specifications of modularity and reproducibility, the structure of the code adheres to the guidelines found in the following subsections.

3.2.1. Model versions and projects

3.2.1.1. Versions

As discussed earlier, the IFM-CAP model will be used for different assessment exercises and some will probably require a significantly restructured model, that is, changing, adding or deleting sets, variables, equations and parameters. For example, the present IFM-CAP model has a fixed input–output relationship for farm production; a request to examine the effect of a nitrogen tax will require the introduction of a functional input–output relationship for nitrogen and output yield. In creating the restructured model the current model will not be deleted, as it is possible that it will be reused. Therefore, it is most probable that more than one **version** of the model will reside in the codebase (44).

There are two ways to handle the parallel existence of two or more versions of the model.

1. **Try to fit all versions in the same codebase.** The variables, equations and parameters of the different models are kept in the same file/files, with the common elements used by all models and the individual elements used only by the relevant model. This approach requires the utilisation of 'smart' and ad hoc solutions to avoid the code of one model interfering with the use of the other model. For example, in the case of a 'shared equation', i.e. has one form in the first model and a different form in the second model, there may be different versions of the equation or a coefficient in front of the terms that are relevant to only one model that will be zeroed out depending on the model employed.
2. **Clearly separate the versions into different folders/files.** In this approach the two versions do not share any code at all. The common code is duplicated in the different model folders (but still separated so that it can be edited independently) and the individual code relates only to the relevant models. Developing a new version does not imply that the workload will be the same as the workload for developing the current version, as the current model will probably be copied to a different folder and only the required changes will be applied. If a bug is found in the predecessor model (the starting point of the new version), it will need to be manually corrected in the descendant model.

As modularity and transparency are of prime concern, we chose to follow the second approach. Although it leads to greater code redundancy, it also provides greater code clarity and is far more straightforward to read, understand and debug (45). In the first approach, the redundant elements (sets, variables, equations and parameters that are used by one model but not by the other) add noise to the process of understanding the code. In addition, the ad hoc solutions required to make the mixed code work with the different models make this approach more error prone and harder to read and understand. Given that reading and understanding the code is a task that is exercised

---

(44) This leads to redefining the notion of the 'model' in the IFM-CAP case. As parallel versions of the IFM-CAP optimisation problem will exist, and all are potentially utilisable, IFM-CAP is converted to a 'modelling framework' that encompasses the design principles and the knowledge base of the developers and can be quickly used to develop a 'model' that responds to a particular policy question.

(45) IFM-CAP v.1 was coded following the first approach; comparing IFM-CAP v.2 with v.1 shows convincingly the advantages of the second approach regarding code clarity and transparency.

more frequently than writing the code (46), the second approach also leads to higher productivity in terms of model development.

The various data preparation processes are also version specific. For example, calibration is expected to be different between the version of the model for a fixed yield and that for a functional yield; in this case each version will have its own calibration code. For data preparation code that is similar between two versions, again for the sake of code clarity, we will follow the code duplication approach. For example, if between two versions the feed module is identical, we will duplicate the code across the two versions.

However, creating a different version of the model whenever a new equation or a new parameter is required would bloat the codebase unnecessarily. Thus, a version should be spawned only in the case of significant changes to an existing version. For dealing with limited modifications to an existing model, the notion of a '**project**' should be used, as presented in the following section.

3.2.1.2. Project

The standard use case is that a version of the IFM-CAP model is employed for answering one or more research questions, for example what the production and environmental effects of a specific configuration of a new CAP policy are or what the production effects of a nitrogen tax are. Answering different research questions with the same version of a model requires slight modifications to the code, to the input data and to the analysis of the results.

Thus, for organising efficiently both the code and the data, the notion of a 'project' is introduced. A project is the isolated code (47) of a particular version of the model that has responsibility for using that version's logic and data and, if necessary, modifying it in order to answer a specific research question. One version of the model will contain many projects, as research questions that utilise that version will accumulate.

For example, the main version of the model contains the **legal_proposal** and the **trade** projects. The first was created for evaluating the European Commission's legal proposal for the CAP post 2020; the second was created for evaluating the production effects of the FTA trade agreements[48] based on the assumptions of the Aglink/Cosimo computable general equilibrium model. The two projects differ in the following aspects:

- **Different output prices.** The legal_proposal inherits the baseline prices while the trade project reads prices from an external file.
- **Different scenario assumptions.** The legal_proposal project has 20 scenarios that vary in terms of the CAP budget, the Right Hand Side (RHS) of CAP constraints and the equations included. On the contrary, the trade project has only three scenarios, which differ only in terms of prices and use the same CAP setting (same RHS constraints and equations).
- **Different reporting needs.** For the legal_proposal project, Qlik reports are created that provide production, income and environmental effects by NUTS region, type of farming, economic size and eco-scheme adopters. For the trade project, we deliver a single Excel file providing production and income effects by type of farming and economic size only.

---

(46) Once the coder produces a part of the code (the code chunk), it is very often used in 'reading mode'. The coder himself or other members of the team 'read' this code chunk with the purpose of understanding it. Two exemplary cases of the 'reading mode' are, when a new developer is incorporated into the team and he needs to understand how the model works, and when the model needs to be extended and this code chunk interacts with the part of the model that is to be extended.

(47) Isolated code in the sense that, whatever happens to the code (change, delete, add logic), the version's code is not affected.

[48] FTA stands for 'Free Trade Agreements'

For consistency with our priority of code clarity, we do not try to meet these different requirements inside the same files of the version code. Instead, we split the code into the different projects. We provide more details on how the versions and the project code are organised in the following section.

3.2.1.3.Version and project organisation

Figure 19 provides an overview of the organisation of different model versions and projects. Different model versions are kept in separate folders within the 'versions' folder of the IFM-CAP code root folder. In each version's folder there is a 'projects' folder. In this folder, each project resides in its own folder. Code that is common to all projects of a version resides in folders and files inside that version's folder.

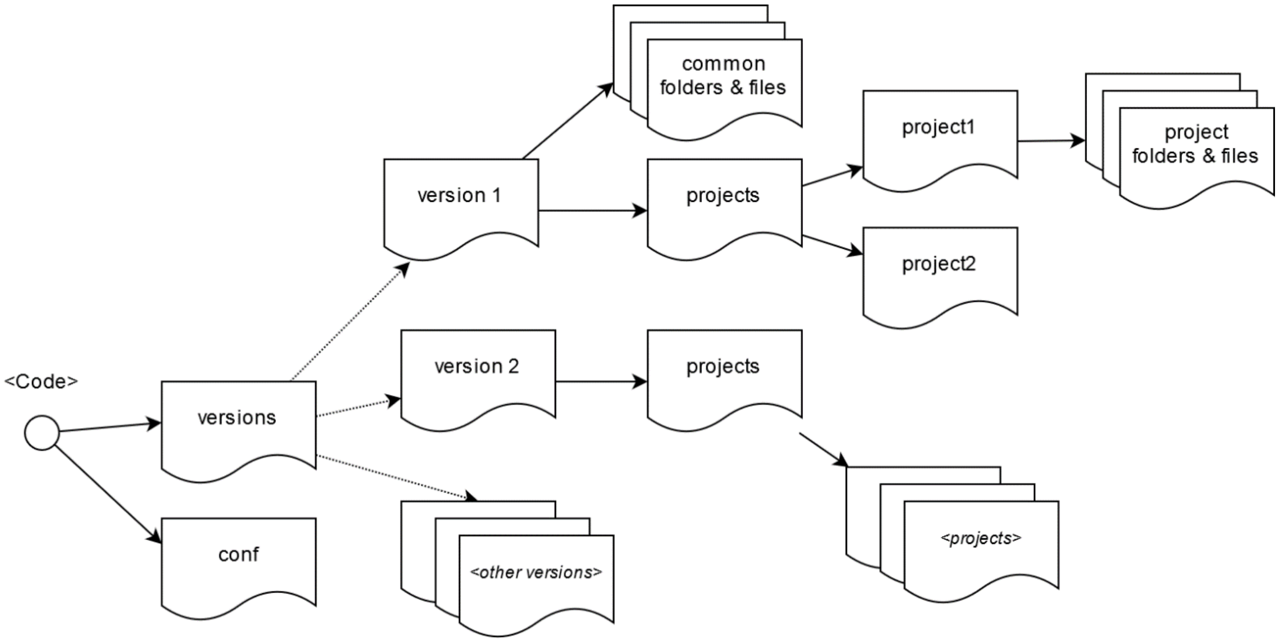


Figure 19. Organisation of versions and projects in the IFM-CAP model

A real-life example is shown in Figure 20. This figure shows the current status of the IFM-CAP model (October 2021).

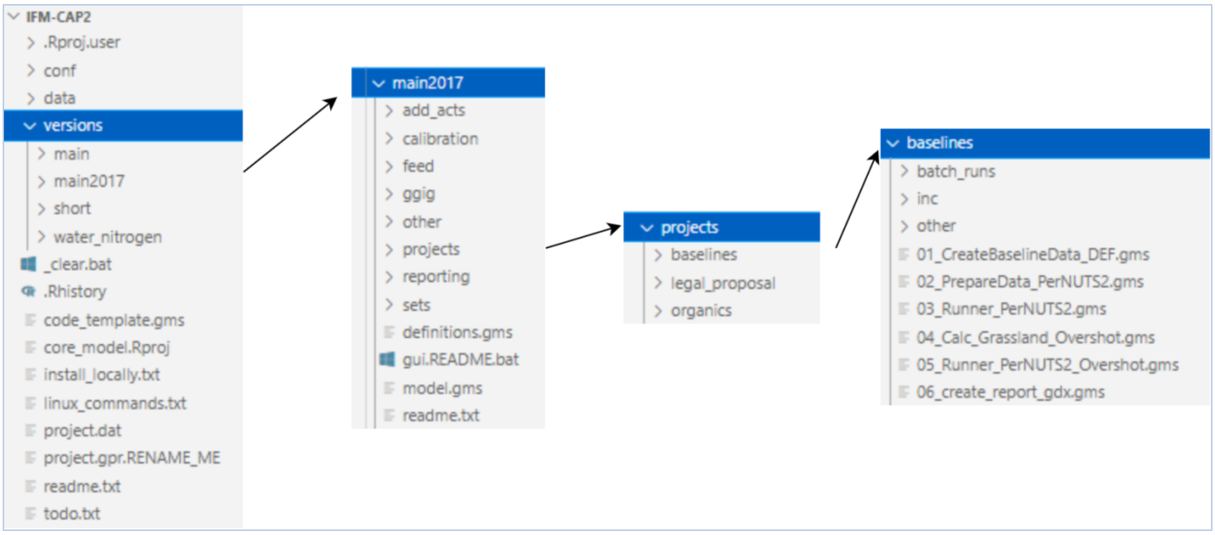


Figure 20. Real-life example of the organisation of the versions and projects in the IFM-CAP model

The 'versions' folder resides in the code root folder. Four versions of the model exist in this folder. The 'main' version is the version that was set up to work with the 2012 base year data, which was mainly prepared using IFM-CAP v.1 code. The 'water_nitrogen' version is under development and will incorporate yield responses based on water and nitrogen.

The 'main2017' version is the standard version of the model related to the 2017 base year. Inside this version, alongside the 'projects' folder, there are several other files and folders. The model.gms and the definitions.gms files are where the equations, the variables and the parameters of the optimisation problem reside, that is, these files define the optimisation problem. The 'sets' folder keeps the version's set definitions and their elements. All are expected to be common to all projects and thus they are defined at this level. The 'ggig' folder contains code related to the GUI; again, as it is common to all projects, it is defined at the same level as the 'projects' folder. The 'calibration', 'feed' and 'add_acts' folders keep the version-related data processes for calibration, the estimation of feed requirements and the elicitation of additional activities, respectively. The 'reporting' folder keeps code that can be used for parsing the results of a run and producing a report. It contains reporting code that is common to all projects, for example for parsing a report GAMS Data eXchange (GDX) file and aggregating its data for uploading to a Business Intelligence web application. Any reporting-related code that is specific to a project will reside inside the folder of that project.

Inside the 'projects' folder are the three existing projects of the main2017 version of the model. The baselines project is responsible for creating the baseline data. The legal_proposal project contains the simulation runs for a paper submitted to a journal regarding the post-CAP 2020 European Commission proposal. The organics project contains simulation runs related to evaluating the impacts of converting 25 % of the land to organic farming. The individual projects can load data from the 'calibration', 'feed' and 'add_acts' folders and can also include code from the parent version folder (sets, definitions and model). However, the projects do not share code between them.

Within each project folder, the organisation of the files and folders is left to the discretion of the modeller responsible for the project. The project code is allowed to load the version's data and to modify them and also to add or remove equations, variables and parameters or change sets, but it should not include code from other projects. In addition, the project code is responsible for running the simulations and saving the results. This setting provides maximum flexibility while maintaining the necessary independence between projects, allowing for a modular model.

The most important principle behind the versions–project organisation is that code that is more generic should be higher in the folder hierarchy. On the contrary, code that is more specific should be deeper in the folder hierarchy. For example, the model.gms file is more generic in the sense that is applicable to any project of a specific version; thus, it is placed above all projects in the folder hierarchy. A file that is running a simulation for a project should reside inside the folder of the project, as it is specific to the project. Analogously, an include file related to a specific project should reside either inside the specific project's folder or deeper in the folder hierarchy.

A corollary is that a file can only include other files that are above it in the folder hierarchy. A project's code file is allowed to include the model.gms file that is located two levels above. However, a version's file should not include a project's file. This stems from the fact that more specific files (which are deeper in the folder hierarchy) can include more generic files (which are higher in the hierarchy), but not vice versa (Figure 21).

Adhering to those two principles will lead to a model that is modular and scalable, as altering any specific file will not lead to other specific code being broken. Any change to a specific level's file will not propagate to the more generic code (as the generic code is not allowed to include specific code) and thus will not create ripple effects (i.e. potentially breaking the functioning of the code in other levels).

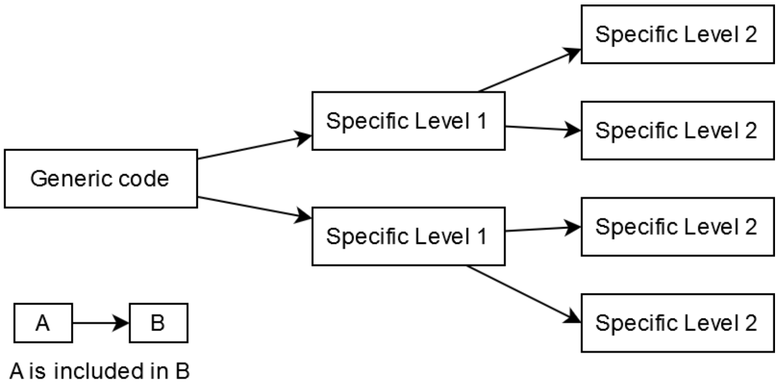


Figure 21. Abstract representation of the 'file include hierarchy' in IFM-CAP

3.2.2. Code and data

3.2.2.1. Static and dynamic code

Within a particular version of the model, the code can be distinguished into two classes that serve different purposes.

1. One class of code, has a static functionality, defining parameters, variables and equations. The model.gms, definitions.gms and sets are all static code. They do not employ the computer's central processing unit (CPU) but are there for defining the elements of the optimisation problem. Although they do not produce any data, they have a central role and correspond to the optimisation problem entity in the farm model abstraction.
2. The other class of code has a dynamic functionality, meaning that it is actually 'doing' something, for example running a model or calculating parameters, by utilising CPU cycles.

A code file can contain either only static code or both static and dynamic code (49).

In IFM-CAP, a design principle is that the optimisation problem itself should reside in a file(s) that contains only static code. In this way, it will be easy to verify that the conceptual model has been coded well. For this reason, we will use model.gms and definitions.gms files in each version of the model.

Files that contain dynamic code will probably include static code from other files and will also contain additional in-line static code. The rule detailed in the previous section on how files should be included applies here too (files should only include other files from the same or a higher level of folder hierarchy).

In addition, as a rule of thumb, we prefer in-line static code to including it from outside files. Including files makes reading the code difficult and complicates debugging. For this reason, including files should be done only when a significant number of lines of code is identical for two or more code files that reside in the same folder level or deeper.

3.2.2.2. Data organisation

As described in Section 3.1.1, the optimisation problem and the data used should be distinct entities. Thus, in IFM-CAP we adhere to the following principles regarding data organisation:

1. The data needed to run the model are prepared and saved before the running of the optimisation model. This means that running the model should not include any data calculation/transformation procedures. Sometimes, it may look redundant to have two files, one for saving the input data and another for loading it and running the simulation, but it greatly improves the reliability of the run itself as run data can be reviewed easily. More details are provided in Section 3.2.2.4.
2. The data operations that produce the data reside in separate folders with independent code. For example, as shown in Section 3.2.1, the calibration process

---

(49) Although it does not produce any results, a file that contains only static code can compile. On the contrary, a dynamic file is required to do something with a static code; thus, it will necessarily contain some static code. In GAMS terms, it is analogous to the code that is relevant for the compilation phase and the execution phase. Parameters, variables and equations can be defined and are parsed in the compilation phase. They are not actually doing anything (i.e. utilising CPU cycles), but they can exist as standalone code. On the contrary, assignments, calculations and solve statements, although they utilise the CPU and change the memory state, cannot be run until the relevant parameters, variables and equations have been defined.

resides in a separate folder inside the version folder and contains code that is not connected to any other part of that version. In this way, we achieve two things. First, one can review the data operation more easily and verify that it does what it is supposed to do. Second, any change to the data operation code will not affect the rest of the version's code and vice versa. The binding of the data operation logic to the rest of the version's logic (i.e. how the calibration procedure is connected to the version's optimisation problem) should be realised only through data exchange. In practice, the data operation will output a data file and this in turn will be the input for another version's file. This facilitates the independence of the data operation code because, as long as the structure of the data file remains the same, the data operation and the rest of the version can be developed independently.

3. As explained in more detail in Section 3.3.2, the code files and the data files are kept in different locations. The data folder structure should mirror the code folder structure. We provide an illustrative example in Figure 22. The two versions that exist in the code folder are reflected in the two different folders that reside in the database folder. As the path of the versions in the code folder always contains a 'versions' folder, we do not mirror this in the database folder; instead, we add a 'v.' prefix to denote that the folders are related to a version. Inside each version folder in the code folder there is a project folder; again, this is reflected inside the 'v.version1' and 'v.version2' folders. The 'p.' prefix denotes that each folder is related to a project. Note that the data operation procedure that resides in the '<Code>/versions/version2/dataOper1' folder is located in the '<Database>/v.version2/dataOper1' folder. The mirroring principle allows two things: first, it enables the numerous files that exist in IFM-CAP to be organised efficiently; second, it enables the code that was responsible for creating a data file to be easily traced.

Within a project or a data operation folder, other folders can be created on demand, to facilitate clear data storage, but in general the mirroring principle should also be respected. Section 3.2.2.4 provides an example of this.

A related principle is that of naming raw data files with an underscore ('_') prefix. Raw data files are those that are not the result of any code execution. For example, an Excel file containing the CAP budget provided by the Directorate-General for Agriculture and Rural Development could have the name '_dg_agri_2021_budget.xls'. In this way one can easily spot the data files that are the starting point for producing other data.

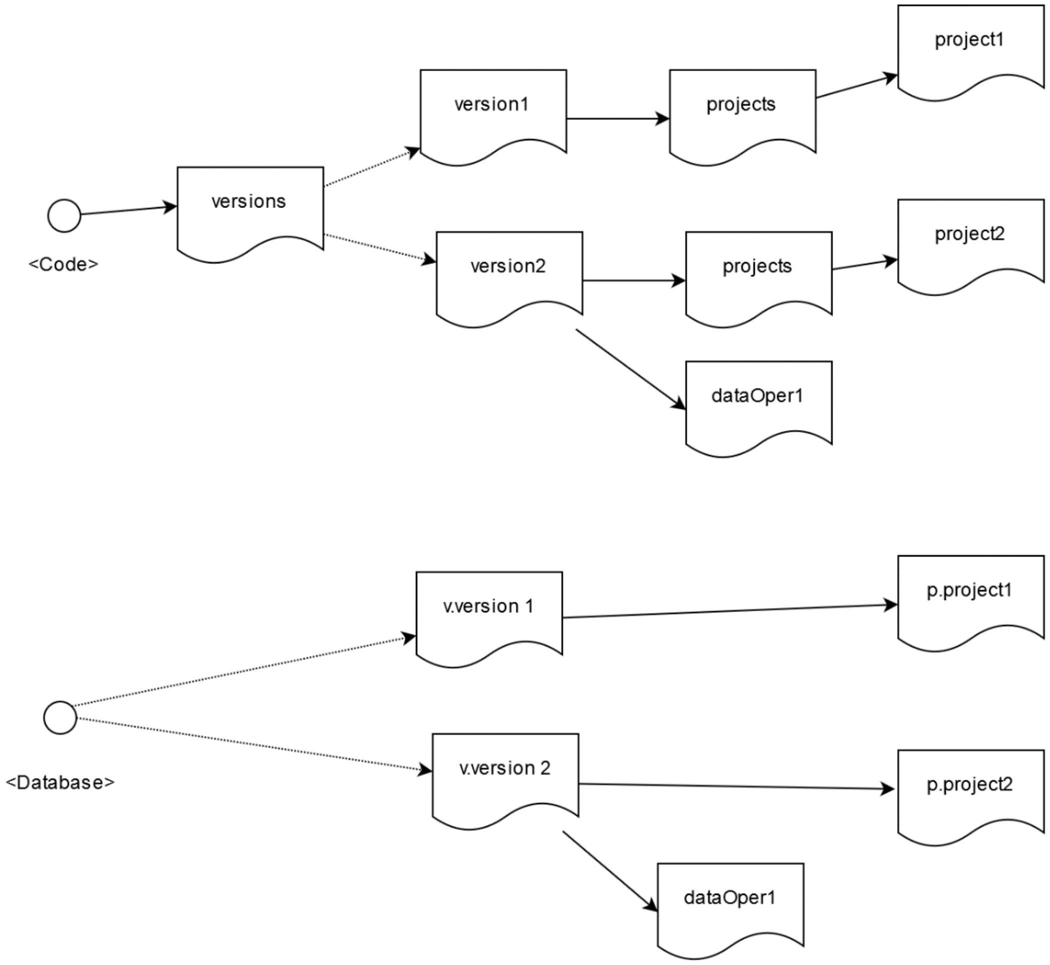


Figure 22. Example of how organisation of the code is mirrored in organisation of the data

3.2.2.3. The small sequential steps pattern

Following the modularisation principle of 'divide and conquer', we decompose the logic of the dynamic code into smaller steps. The amplitude of the decomposition is at the discretion of the version's modeller, but a balance between reproducibility, traceability and code efficiency should be sought.

For this, the version's dynamic code is split into files that start with a number that is indicative of the order of the steps. The code in each file has a self-contained logic that completes a specified task. In practice, self-contained logic means that a file reads an input data file, applies calculations and saves an output data file; in this way the calculations can be verified easily by comparing the changes between the input and the output data files and the intentions of the algorithm contained in the file.

Some concrete examples of this pattern are provided in Figure 23. In 'calibration on v.main2017', the 00 step gathers all required data from other parts of the version and compiles a big GDX file; the 01 step then splits this file into GDX files at NUTS 2 level that are read by the 02 step, which estimates priors, etc.

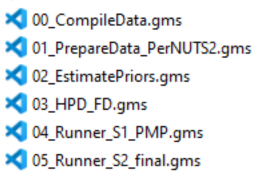
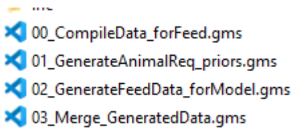
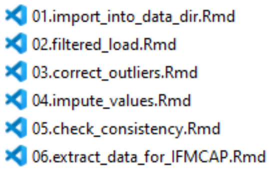
Calibration on v.main2017	Feed in v.main2017	FADN module
 00_CompileData.gms 01_PrepareData_PerNUTS2.gms 02_EstimatePriors.gms 03_HPD_FD.gms 04_Runner_S1_PMP.gms 05_Runner_S2_final.gms	 00_CompileData_forFeed.gms 01_GenerateAnimalReq_priors.gms 02_GenerateFeedData_forModel.gms 03_Merge_GeneratedData.gms	 01.import_into_data_dir.Rmd 02.filtered_load.Rmd 03.correct_outliers.Rmd 04.impute_values.Rmd 05.check_consistency.Rmd 06.extract_data_for_IFMCAP.Rmd

Figure 23. Examples of the small sequential steps pattern in IFM-CAP

Especially for data operations, we have moved a significant part of the GAMS code of IFM-CAP v.1 into R Markdown. R Markdown is a dynamic reporting framework that combines explanatory text, R code and the results of the code execution in the same file. The use of R Markdown enhances transparency in the following ways.

1. Documentation of the code is straightforward and efficient. One can describe the logic of the code and its assumptions in the same place that the code resides, using figures, tables and formatted text. In addition, having the documentation together with the code allows the documentation to be easily updated when updating the code, reducing the cost of maintaining the documentation and coupling the code with its documentation.
2. Inspection of what the code does is straightforward. One can easily produce summary statistics of the code execution. For instance, in a script with the purpose of loading and transforming the raw comma-separated values (CSV) files of the FADN database into an R-compatible format, one can display the names of the files that were loaded, the number of records, some plots with the distributions of the variables of interest, etc.
3. The workflow of the file is naturally split into smaller steps. This allows for easy debugging.

3.2.2.4. The prepare–run–report pattern

As already discussed, the data and the dynamic code that issues solve commands for the optimisation problem should be distinct. For this we employ the ′prepare–run–report′ pattern, as shown in Figure 24.

The ‘prepare’ procedure is responsible for collecting the data for the optimisation problem and preparing a GDX file that contains these data. The ‘run’ procedure includes (using $include) the optimisation GMS file and loads the data (using $GDXIN or execute_load) and runs the model. Running the model can mean running the model for one farm or running the model for a whole NUTS 3 area, for example. The ‘run’ procedure is also responsible for saving the solution results into a GDX file. The solution results contain only the levels of the variables and the equations at the optimum solution. Any further calculations that are based on the solution are carried out during the ‘report’ procedure, which loads the solution file, processes it and creates a GDX file that contains the results of the calculations.

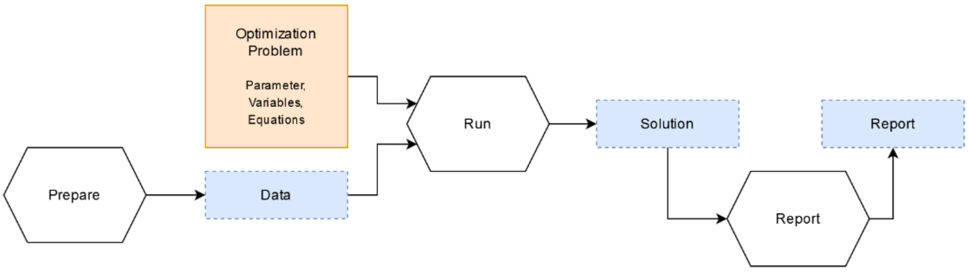


Figure 24. The prepare–run–report sequence

3.3. Using and developing the model

3.3.1. Installation

3.3.1.1. Software and hardware requirements

The farm decision model is written in GAMS (50). It is solved using the CONOPT (51) and SBB (52) solvers. Thus, in order to run the model, a licensed version of GAMS software with a CONOPT and SBB licence is required. We also suggest using GAMS distribution 28.2 or higher.

The creation of the model database uses both GAMS and R. Thus, in case one needs to run the data preparation procedures, an R (53) installation is required. Rstudio (54) is also recommended as it provides the best way to open and handle R code.

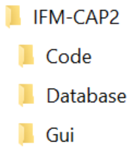
For running the GUI, any freely available Java (55) distribution is required. In addition, the binaries of the GAMS graphical user interface generator (56) (GGIG) are necessary.

Installing TortoiseSVN (57) is recommended. It facilitates downloading of the latest version of the code and updating of the model.

There are distinct hardware requirements for preparation of the data and for running the model. Regarding the preparation of IFM-CAP data, a personal computer with at least 16 GB of memory is required. Regarding the running of the farm decision model, an acceptable hardware is a personal computer with multiple CPUs so that farms can be solved in parallel. At the JRC, a Windows machine with 64 GB of memory and 40 CPUs allows us to prepare data without any memory problems and also to solve one scenario for the whole of the EU in approximately 3 hours.

3.3.1.2. How to install

It is advisable to create a folder named 'IFM-CAP2' and inside this folder to create three subfolders: 'Code', 'Database' and 'Gui'.



The IFM-CAP SVN repository (58) should then be checked out to the 'Code' folder. A username and password provided by the IFM-CAP team at the JRC.D.4 Economics of Agriculture Unit are required to complete this step. More details on how to use SVN to check out a repository are provided on the TortoiseSVN website (59).

When the code has been downloaded to the 'Code' folder, the file 'install_locally.txt' will be available. This file contains detailed instructions on how to download the initial database and the GGIG Java binaries for the GUI.

---

(50) https://www.gams.com/
(51) https://www.gams.com/latest/docs/S_CONOPT.html
(52) https://www.gams.com/latest/docs/S_SBB.html
(53) https://cran.r-project.org/bin/windows/base/
(54) https://rstudio.com/products/rstudio/download/
(55) https://www.java.com/en/download/
(56) https://www.ilr.uni-bonn.de/em/rsrch/ggig/ggig_e.htm
(57) https://tortoisesvn.net/
(58) https://svn.jrc.es/repos/IFM-CAP/branches/IFM2
(59) https://tortoisesvn.net/docs/nightly/TortoiseSVN_en/tsvn-quick-start.html

3.3.2. Folder structure

There are four main folders in the IFM-CAP model (Figure 25). The **GIGG executables** contain the Java binaries for the GUI. The **scratch directory** is the directory where temporary files are created during the execution of the model (e.g. the GAMS scratch directories when solving the model). The **database** contains all input data files required for running the model and all output data files produced by the model. The **code** directory contains all the model code.

The locations of these folders are independent from each other and are configurable by the user. For example, the <ScratchDir> folder can reside in a RAM disk or a fast SSD disk, the <GGIGexe> folder can be in a network shared drive, and the <Database> and the <Code> folders can be in another drive. The main configuration file is located in '<Code>/conf/inc.gms'. In this file we explicitly define the location of the <ScratchDir> and <Database> folders. An example folder configuration is provided in Box 2.

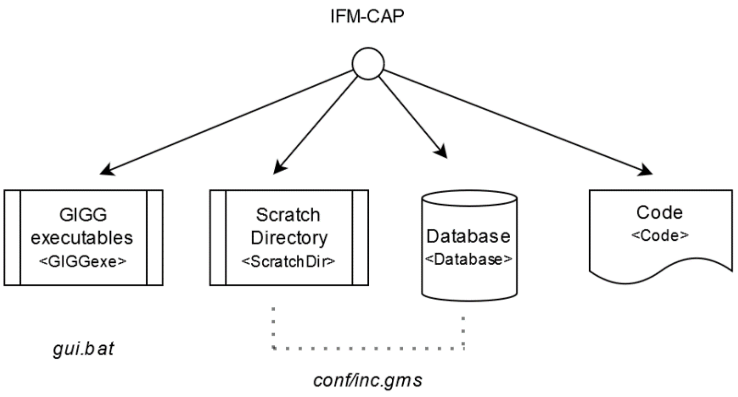


Figure 25. IFM-CAP folder structure (level 1)

Box 2. Example folder configuration

```
$ontext
In this file set any Model-wide Compile-time Variables
This file must be the only place where such Model-wide variables can be defined.
$offtext


*----------------------------
* DatabaseDirectory
* the directory that data operations are taking place.
* This facilitates easy changing of input/output data
* It shall NOT have a trailing "\"
*----------------------------
$SETGLOBAL DatabaseDirectory "E:/IFM-CAP2-database"

*----------------------------
* ScratchDirectory
* The directory that temporary GAMS directories will be created (255a, 255b, etc.)
* Use a location of a fast disk (SSD, Ramdisk, etc.)
* It shall NOT have a trailing "\"
*----------------------------
$SETGLOBAL ScratchDirectory "F:"
```

For <Code>, we do not explicitly define its location. As we run the model through the GAMS environment, there is an implicit definition of the code directory as equal to the working directory of the GAMS project we are running. This implies that, when including a GMS file, we should always start from the root of the code file but without explicitly giving a file location. We provide an example in Box 3. We load the file calling '$INCLUDE versions/main/definitions.gms'.

Box 3. Example of %DatabaseDirectory% usage

```
$INCLUDE conf/inc.gms


$INCLUDE versions/main/definitions.gms


$GDXIN '%DatabaseDirectory%/ifm1/DATA_2012_bis_FDALL.gdx'
$ LOAD FD
$GDXIN

$GDXIN '%DatabaseDirectory%/ifm1/p_data_baseline_2030_IA_20122018.gdx'
$ LOAD p_DataFD
$GDXIN

EXECUTE_UNLOAD "%DatabaseDirectory%/ifm1/FDALL_Clean.gdx" FDALL;
```

Developers of the application should respect the concern of each folder. Most importantly, data, whether input or output, should reside only in <Database>. In addition, whatever type of code is used should reside in <Code> (60).

In GAMS, when loading or saving data, the folder where data are located should always be referred to using the %DatabaseDirectory% control parameter. In the example in Box 3, at the beginning of the script we include 'conf/inc.gms', which contains the values of the folder locations; then, for $GDXIN or EXECUTE_UNLOAD, we always provide the file path in relation to %DatabaseDirectory%. This allows two or more database folders to be maintained. Switching the model from using one database to using another database is very easy by changing the %DatabaseDirectory% control variable in 'inc/conf.gms'.

3.3.3. Running the model

In this section, we provide details on how to run the model. There are three ways to run the model.

1. **Using the GMS files.** This option provides finer control over the model and can be used either for debugging or for further development.
2. **Using the GUI.** This option is preferable for novice users or for working with a few NUTS 2 or NUTS 0 regions, especially if there is no need to carry out debugging.
3. **Using the GGIS's batch facility.** This is the most efficient way to run the model for the whole of the EU. The model is run in parallel, utilising as many CPUs as possible.

3.3.3.1. Using the GAMS source files

GMS script organisation

There are three GMS script types.

---

(60) For example, the GGIG XML definition files do not reside in the <GGIGexe> folder but in '<Code>/versions/main/ggig/ifmcap_default.xml'. These files define the GUI elements and are part of the IFM-CAP code, as they are expected to evolve (new work steps and tasks, new options, etc.).

1. **Model definition scripts.** The first category includes the scripts containing the optimisation problem. These are the 'model.gms', which contains the variables and equations of the model, and the 'definitions.gms', which contains the parameter definitions.
2. **Runnable scripts.** The second category includes the GMS scripts that either load data and transform them or load data and solve the model and produce results.
3. **Included scripts.** The last category of scripts contains GMS files that are not runnable but instead contain code that is reusable by two or more runnable scripts. The strategy is to create included scripts only if their content is identical between two or more runnable scripts.

For runnable scripts, a template is provided in Box 4. It contains a header with information on the purpose of the script, the control parameters it contains, and its input and output data files.

Box 4. Template for a runnable script

```
********************************************************************************
$ontext
<The purpose of the Script>

Control Parameters:
<List of Control Parameters>

Input data Files:
<List of Data input files>

Output data Files:
<List of Data output files>
$offtext
********************************************************************************


* CONTROL PARAMETERS
********************************************************************************

* If running in controlled state then the GUI overrides these options
$IfThenI.controlled NOT %CONTROLLED% == "1"

* <DEFINE THE CONTROL PARAMETER VALUES WHEN SCRIPT IS NOT RUN IN CONTROLLED
* STATE>

$Else.controlled

* <CODE SPECIFIC TO WHEN SCRIPT RUNS IN CONTROLLED STATE>

$EndIf.controlled



* Include Model-wide Compile-time Variables
********************************************************************************
$INCLUDE conf/inc.gms
;


* Script Logic
********************************************************************************
* <CODE THAT IMPLEMENTS THE SCRIPT'S PURPOSE>
```

Control parameters add flexibility to runnable scripts. They are $SET statements that control the behaviour of the script. As shown in Box 4, they are defined in a dedicated section at the beginning of the script. The modeller can define their values inside the

GMS script, but in cases in which the script is called externally (e.g. a $call or an execute statement), if the –CONTROLLED=1 parameter is passed, they will be ignored and the external control parameter values will be used instead. An example with extended comments is provided in Box 5.

Box 5. Example of control parameters

```
* <MORE LINES ABOVE>

* If you call from another gams script with the following statement, the next $IfThen will
* be ignored
* $call <script name> -CONTROLLED=1

* If running in controlled state then the GUI overrides these options
$IfThenI.controlled NOT %CONTROLLED% == "1"

* Defining what baseline data are used
$SET BASELINE_DATA "CAPRIbaselineIA_1230_2020"

* Defining which NUTS 2 region(s) to run
$SET NUTS2 "EL140000"

* This is enforced only in the case you run the script through GAMS IDE
OPTION profile=1;

$Else.controlled

$EndIf.controlled

* <MORE LINES BELOW>

$BATINCLUDE versions/main/projects/baselines/inc/load_data.gms
"%DatabaseDirectory%/v.main/p.Baselines/%BASELINE_DATA%/input/nuts2/model_data_%NUTS2%.gdx"

* <MORE LINES BELOW>
```

3.3.3.2. *Using the graphical user interface*

IFM-CAP has also been linked to the GGIG framework, written in Java (Britz, 2011). This offers a GUI that has functionalities that are organised into work steps and tasks. Each version has its own GUI, and it can be started by running '<Code>/versions/<VERSION NAME>/gui.bat'.

As can be seen in Figure 26, the actions that a user can carry out have a hierarchy. There are work steps that contain tasks. Each of the tasks has various options and makes a runnable GAMS script run. In addition, within each task the user can compile (and test if the program compiles without errors), start (execute) or stop the GAMS program.

Figure 26. IFM-CAP GUI parts

Table 22 provides an example of the tasks that are available to the user for the main version of the model.

Table 22. GUI work steps and tasks for the main version of the model

Work step	Task	Description
Administration tasks	01 Create Directories	Create the directories required in the %DatabaseDirectory%. See 'versions/main/others/create_directories.gms'.
	02 Clean Tmp	Clean temporary files in the %DatabaseDirectory%. See 'versions/main/others/clean_data.gms'.
Baseline tasks	01 Compile Single GDX	Create a single GDX file that contains all model data for all farms.
	02 BSL Compile NUTS2 GDX	Create GDX files for several NUTS 2 regions.
	03 BSL Runner NUTS2	Run the baseline model for several NUTS 2 regions.
	04 BSL GrassOver Calc	Calculate the grassland overshooting for several NUTS 2 regions.
	05 BSL Runner GrassOver NUTS2	Run the baseline model considering the grassland overshoot for several NUTS 2 regions.
<Project> scenario tasks	01 PSC Compile NUTS2 GDX	Create GDX files for several NUTS 2 regions and one scenario.
	02 LP Runner NUTS2	Run the legal proposal model for several NUTS 2 regions and one scenario.
	03 LP GrassOver Calc	Calculate the grassland overshooting for several NUTS 2 regions and one scenario.
	04 LP Runner GrassOver NUTS2	Run the legal proposal model considering the grassland overshoot for several NUTS 2

		regions and one scenario.
	05 LP Check Run Status	Report for each of steps 01–04 how many NUTS 2 regions and how many scenarios have not been run.
	06 LP Reporting	Create a report that compiles the solutions from many NUTS 2 regions and many scenarios.

The contents of the GUI are controlled through the '<Code>/versions/<VERSION NAME>/ggig/ifmcap_default.xml' file. More information on how to edit this file in order to add elements in the GUI is found in the GGIG documentation (Britz, 2010). For every task, the file '<Code>/versions/main/ggig/ggig_controller.gms' is called. This file acts as an intermediate between the GUI and the individual runnable scripts.

3.3.3.3.Using batch runs

The GGIG binaries provide a batch run feature that is very useful for running all steps (prepare data, run model, check grassland and run model with grassland overshoot) for the whole of the EU. It is the best option to follow if debugging of the model has been carried out.

Batch runs are project specific and are stored inside '/versions/<VERSION NAME>/<PROJECT NAME>/batch_runs'.

The batch runs can be either initiated through the GUI or run from the command line with a Java command.

3.4. Naming conventions

Title	Rule	Examples
GAMS commands	Use upper case for GAMS commands and GAMS-reserved words (e.g. 'SUM' and 'AND').	
Variables and parameters	Start parameter names with 'p_', continuous variables with 'v_' and binary (or integer) variables with 'b_'. *Exception to the above:* All variables after the model solution are initially stored as parameters with a prefix 'rv_' indicating that they represent the resulting values of variables. Such parameters are later used in the reporting process and transformed in the reporting parameter p_Res. Use camel case names* for naming variables and parameters. If a parameter represents a specific attribute of another parameter or modelling concept, use an underscore (e.g. maximum or minimum values). Try to use names of adequate length when defining parameters and variables, striking a balance between the number of characters and self-explanatory names. When some concepts modelled have longer names or are difficult to describe, use the shortest names possible that provide a clear understanding of each variable or parameter. This is in line with the GAMS good coding practices** for using longer names and additionally enhances the readability of the code.	p_DSub_TVal Decoupled subsidies total value p_DSub_UVal Decoupled subsidies unit value
Sets	Use upper case.	ACT Set of all production activities DPAY Set of all decoupled payments
Subsets	Subset names begin with a prefix referring to the superset. Use upper case.	CROP_ACER(CROP) Set of all cereal activities, which is a subset of set CROP DPAY_PIL1(DPAY) Subset of all pillar 1 decoupled

		payments, which is a subset of set DPAY
Aliases	Name any set alias with the name of the set and a number from 1–9. For example, for set ACT, if one alias is needed, name it ACT1; if another alias is needed, name it ACT2, etc.	
Equation naming	Use upper case. Equations start with E_XXX. Variables that are the direct result of an equation calculation can have the same name as the equation.	

NB: *Camel case is the practice of writing phrases such that each word or abbreviation of a phrase begins with a capital letter, with no intervening spaces or punctuation, for example HelloWorld and UpperCamelCased. **The good coding practices are available on the GAMS website (https://www.gams.com/latest/docs/UG_GoodPractices.html).

References

Acs, S., Hanley, N., Dallimer, M., Gaston, K.J., Robertson, P., Wilson P. and Armsworth P.R. (2010), 'The effect of decoupling on marginal agricultural systems: implications for farm incomes, land use and upland ecology', *Land Use Policy* 27: 550–563.

Alford, A., Griffith, G. and Cacho, O. (2004), *A northern tablelands whole-farm linear program for economic evaluation of new technologies at the farm-level*, Economic Research Report No 13, NSW Agriculture, Armidale.

Antle, J.M. (1983), 'Incorporating risk in production analysis', *American Journal of Agricultural Economics* 65: 1099–1106.

Arata, L., Peerlings, J. and Sckokai, P. (2013), 'Manure market as a solution for the nitrates directive in Italy', *New Medit* 12: 22–33.

Arata, L., Donati, M., Sckokai, P. and Arfini, F. (2017), `Incorporating risk in a positive mathematical programming framework: a dual approach', *Australian Journal of Agricultural and Resource Economics* 61: 265–284.

Arribas, I., Louhichi, K., Perni, A., Vila, J. and Gomez y Paloma, S. (2017), 'Modelling farmers' behaviour toward risk in a large scale positive mathematical programming (PMP) model', in Tsounis, N. and Vlachvei, A. (eds), *Advances in Applied Economic Research, Proceedings of the 2016 International Conference on Applied Economics (ICOAE),* Springer, Cham, pp. 625–643.

Baldoni, E. and Ciaian, P. (2021) *The capitalisation of CAP subsidies into land rents and land values in the EU*, EUR 30736 EN, Publications Office of the European Union, Luxembourg, ISBN 978-92-76-38670-4, doi:10.2760/404465, JRC125220.

Bakker, R.R.C. (2013), *Availability of lignocellulosic feedstocks for lactic acid production*, Report No 1391, Food and Biobased Research, Wageningen University.

Bielza, M., J.M. Terres, A. Leip, M.L. Paracchini, L. Nisini, A. Wania, V. Angileri, G. Carmona (2015). Agri-environmental indicators for the IFM-CAP model. JRC Technical Reports JRC97254.

Bielza, M. (2017). Agri-environmental indicators for the IFM-CAP model: summary and factsheets. Working document, iMAP9 project. European Commission.

Bielza, M., A. Perni, A. Leip, J.-M. Terres, K. Louhichi, P. Ciaian (2021). Quality check of IFM CAP agri environmental indicators. Comparison with other sources. Internal JRC report.

Blanco-Fonseca, M. (2010), *Literature review of methodologies to generate baselines for agriculture and land use*, CAPRI-RD Deliverable D4.1, available online: http://www.ilr.uni-bonn.de/agpo/rsrch/capri-rd/docs/d4.1.pdf

Brink, L., & McCarl, B. (1978). The tradeoff between expected return and risk among cornbelt farmers. *American Journal of Agricultural Economics*, *60*(2), 259-263.

Britz, W. (2011), *The graphical user interface for CAPRI*, University of Bonn, Bonn.

Britz, W. and Arata, L. (2019), `Econometric mathematical programming: an application to the estimation of costs and risk preferences at farm level', *Agricultural Economics* 50: 191–206.

Britz, W., Ciaian, P., Gocht, A., Kanellopoulos, A., Kremmydas, D., Müller, M., Petsakos, A. and Reidsma, P. (2021), `A design for a generic and modular bio-economic farm model', *Agricultural Systems* 191(C): 1-14.

Britz, W. and Witzke, P. (2014), *CAPRI model documentation 2012*, available online: http://www.capri-model.org/docs/capri_documentation.pdf

Buysse, J., Van Huylenbroeck, G. and Lauwers, L. (2007), 'Normative, positive and econometric mathematical programming as tools for incorporation of multifunctionality in agricultural policy modelling', *Agriculture Ecosystems & Environment* 120: 70–81.

Cabrera, V.E. (2012), 'A simple formulation and solution to the replacement problem: a practical tool to assess the economic cow value, the value of a new pregnancy, and the cost of pregnancy loss', *Journal of Dairy Science* 95: 4683–4698.

Chavas, J.P. (2000), 'On information and market dynamics: the case of the U.S. beef market', *Journal of Economic Dynamics and Control* 24: 833–853.

Chavas, J.P., Pope, R.D. and Kao, R.S. (1983), 'An analysis of the role of futures prices, cash prices and government programs in acreage response', *Western Journal of Agricultural Economics* 65: 27–33.

Cimino, Orlando, Roberto Henke, and Francesco Vanni. "The effects of CAP greening on specialised arable farms in Italy." *New Medit* 14.2 (2015): 22-31.

Cortignani, R. and Severini, S. (2012), 'Modeling farmer participation to a revenue insurance scheme by means of positive mathematical programming', *Agricultural Economics – Czech*, 58: 324–331.

Cortignani, R., Severini, S. and Dono, G. (2017), 'Complying with greening practices in the new CAP direct payments: an application on Italian specialized arable farms', *Land Use Policy* 61: 265–275.

Coyle, B.T. (1999), 'Risk aversion and yield uncertainty in duality models of production: a mean–variance approach', *American Journal of Agricultural Economics* 81: 553–567.

Crosson, P., O'Kiely, P., O'Mara, F.P. and Wallace, M. (2006), 'The development of a mathematical model to investigate Irish beef production systems', *Agricultural Systems* 89: 349–370.

De Cara, S. and Jayet, P.A. (2000), 'Emissions of greenhouse gases from agriculture: the heterogeneity of abatement costs in France', *European Review of Agricultural Economics* 27: 281–303.

De Cara, S., Houze, M. and Jayet, P.A. (2005), 'Methane and nitrous oxide emissions from agriculture in the EU: a spatial assessment of sources and abatement costs', *Environmental and Resource Economics* 32: 551–583.

de la Guardia, F. H., & Sturdy, J. (2019). Best Practices for Transparent, Reproducible, and Ethical Research.

DG AGRI (2013), *Overview of CAP reform 2014–2020*, Agricultural Policy Perspectives Brief No 5, Unit for Agricultural Policy Analysis and Perspectives, DG AGRI, available online: http://ec.europa.eu/agriculture/policy-perspectives/policy-briefs/05_en.pdf

DG AGRI (2014), EU Sugar Market Observatory, Balance sheet, Committee for the Common Organisation of Agricultural Markets, Working Document, available online: http://ec.europa.eu/agriculture/sugar/balance-sheets/balance-sheet_en.pdf

DG AGRI (2015a), *Voluntary coupled support – sectors mostly supported*. *Notification of decisions taken by Member States by 1 August 2014*, available online http://ec.europa.eu/agriculture/direct-support/direct-payments/docs/voluntary-coupled-support-note_en.pdf

DG AGRI (2015b), *Voluntary coupled support – other sectors supported. Notification of decisions taken by Member States by 1 August 2014*, available online: http://ec.europa.eu/agriculture/direct-support/direct-payments/docs/voluntary-coupled-support-note-2_en.pdf

Dijkstra, E. W. (1982). On the role of scientific thought. In *Selected writings on computing: a personal perspective* (pp. 60-66). Springer, New York, NY.

Ducros, D., Kephaliacos, C. and Ridier, A. (2005), 'Mixed crop livestock farming incorporating agroforestry orchards facing the new CAP', Paper presented at the 99th seminar of the EAAE (European Association of Agricultural Economists), Copenhagen, Denmark, 24–27 August 2005.

Edwards, R.A.H., Šúri, M., Huld, T.A. and Dallemand, J.F. (2005), 'GIS-based assessment of cereal straw energy resource in the European Union', in *Proceedings of the 14th European Biomass Conference and Exhibition*, Biomass for Energy, Industry and Climate Protection, Paris, France, 17–21 October 2005.

Errington, A. (1989), 'Estimating enterprise input–output coefficients from regional farm data', *Journal of Agricultural Economics 40*: 52–56.

EU (2013), Regulation (EU) No 1307/2013 establishing rules for direct payments to farmers under support schemes within the framework of the common agricultural policy and repealing Council Regulation (EC) No 637/2008 and Council Regulation (EC) No 73/2009, OJ L 347/608, 20.12.2013, p. 608–670.

EU (2014a), Regulation (EU) No 639/2014 of 11 March supplementing Regulation (EU) No 1307/213 of the European Parliament and of the Council establishing rules for direct payments to farmers under support schemes within the framework of the common agricultural policy and amending Annex X to that regulation. Article 65(1)(c) on 'Greening monitoring output indicators'.

EU (2014b), Regulation (EU) No 640/2014 of 11 March 2014 supplementing Regulation (EU) No 1306/2013 of the European Parliament and of the Council with regard to the integrated administration and control system and conditions for refusal or withdrawal of payments and administrative penalties applicable to direct payments, rural development support and cross compliance, OJ L 181/48, 20.6.2014, p. 48–73.

European Commission (2013), Report from the Commission to the European Parliament and the Council 'Evolution of the sugar imports in the European Union from LDC and ACP countries', COM(2013) 323 final, European Commission, available online: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2013:0323:FIN:en:PDF

European Commission (2015), *Direct payments post 2014 decisions taken by Member States by 1 August 2014 – state of play on 07.05.2015*, European Commission, DG Agriculture and Rural Development.

European Commission (2016), *Review of greening after one year*, Commission Staff Working Document SWD(2016) 218, European Commission, Brussels.

European Commission (2017), *The small farmers scheme*, Directorate-General for Agriculture and Rural Development, European Commission, Brussels.

European Commission (2018). "Proposal for a Regulation of the European Parliament and of the Council establishing rules on support for strategic plans to be drawn up by Member States under the Common agricultural policy (CAP Strategic Plans) and financed by the European Agricultural Guarantee Fund (EAGF) and by the European Agricultural Fund for Rural Development (EAFRD) and repealing Regulation (EU) No 1305/2013 of the European Parliament and of the Council and Regulation (EU) No 1307/2013 of the European Parliament and of the Council." COM(2018) 392 final, Brussels.

European Commission (2020a). "Analysis of links between CAP Reform and Green Deal." SWD(2020) 93 final, Brussels.

European Commission (2020b). "EU agricultural outlook for markets, income and environment, 2020-2030." DG Agriculture and Rural Development, Brussels.

FADN. (2015). Farm Accountancy Data Network, Public Database, Directorate-General for Agriculture and Rural Development, European Commission website: ec.europa.eu/agriculture/rica/. Accessed 15 January 2015.

de Frahan, B. H., Buysse, J., Polomé, P., Fernagut, B., Harmignie, O., Lauwers, L., ... & Van Meensel, J. (2007). Positive mathematical programming for agricultural and environmental policy analysis: review and practice. *Handbook of operations research in natural resources*, 129-154.

Freund, R. (1956), 'The introduction of risk into a programming model', *Econometrica* 24: 253–263.

Garnache, C. and Mérel, P. R. (2015), `What Can Acreage Allocations Say about Crop Supply Elasticities? A Convex Programming Approach to Supply Response Disaggregation', *Journal of Agricultural Economics* 66: 236–256.

Gartner, J.A. (1982), 'Replacement policy in dairy herds on farms where heifers compete with cows for grassland – Part 3: A revised hypothesis', *Agricultural Systems* 8: 249.

GfE (2006), *Empfehlungen zur Energie – und Nährstoffversorgung von Schweinen*, Gesellschaft für Ernährungsphysiologie/Ausschuss für Bedarfsnormen.

Gocht, A., Ciaian, P., Bielza, M., Terres, J. M., Röder, N., Himics, M. and Salputra, G. (2017), 'EU-wide economic and environmental impacts of CAP greening with high spatial and farm-type detail', *Journal of Agricultural Economics* 68: 651–681.

Golan, A., Judge, G., and Miller D. (1996), *Maximum Entropy Econometrics*. Chichester UK: Wiley, 1996.

Havlík, P., Jacquet, F., Boisson, J.M., Hejduk, S. and Veselý P. (2006), 'Mathematical programming models for agri-environmental policy analysis: a case study from the White Carpathians', *Agricultural Economics – Czech* 52: 51–66.

Hazell, P.B. and Norton, R.D. (1986), '*Mathematical programming for economic analysis in agriculture'*, New York: Macmillan Publishing Co.

Heckelei, T. (2002), *Calibration and estimation of programming models for agricultural supply analysis*, Habilitation Thesis, Agricultural Faculty, Bonn University, Bonn.

Heckelei, T. and Britz, W. (2000), 'Positive mathematical programming with multiple data points: a cross-sectional estimation procedure', *Cahiers d'Economie et Sociologie Rurales* 57: 28–50.

Heckelei, T. and Britz, W. (2005), 'Models based on positive mathematical programming: state of the art and further extensions', Plenary paper presented at the 89th EAAE Seminar, 3–5 February 2005, Parma.

Heckelei, T. and Wolff, H. (2003), 'Estimation of constrained optimisation models for agricultural supply analysis based on generalised maximum entropy', *European Review of Agricultural Economics* 30: 27–50.

Heckelei, T., Britz, W. and Zhang, Y. (2012), 'Positive mathematical programming approaches – recent developments in literature and applied modelling', *Bio-based and Applied Economics* 1: 109–124.

Heckelei, T., Mittelhammer, R. and Jansson T. (2008), *A Bayesian alternative to generalized cross entropy solutions for underdetermined econometric models*, Agricultural and Resource Economics, Discussion Paper, Institute for Food and Resource Economics, Bonn.

Himics, H., Artavia, M. and Hélaine, S. (2014), *Calibrating the CAPRI and ESIM models to the mid-term commodity market outlook of the European Commission*, JRC Technical Reports No JRC72882, European Commission, Joint Research Centre, Institute for Prospective Technological Studies, available online: https://ec.europa.eu/jrc/sites/default/files/JRC72882_Calibration_report_CAPRI_ESIM.pdf

Himics, M., Ciaian, P., Van Doorslaer, B. and Salputra, G. (2013), *Management guidelines for the CAPRI baseline*, CAPRI-RD Deliverable D4.8, available online: http://www.ilr.uni-bonn.de/agpo/rsrch/capri-rd/docs/d4.8.pdf

Howitt, R.E. (1995), 'Positive mathematical programming', *American Journal of Agricultural Economics* 77: 329–342.

IPCC (2006), *IPCC Guidelines on National Greenhouse Gas Emission Inventories*, IPCC.

Jansson, T. and Heckelei, T. (2011), 'Estimating a primal model of regional crop supply in the European Union', *Journal of Agricultural Economics* 62: 137–152.

Jansson, T., Heckelei, T., Gocht, A., Basnet, S.K., Zhang, Y. and Neuenfeldt, S. (2014), 'Analysing impacts of changing price variability with estimated farm risk-programming models', Selected paper, 14th European Association of Agricultural Economists Congress, pp. 26–29.

Jarrige, R. (ed.) (1988), *Alimentation des bovins*, *ovins et caprins*, INRA, Paris.

Jarrige, R. (ed.) (1989), *Ruminant nutrition. Recommended allowances and feeding tables*, INRA, Paris.

Jaynes, E. T. (1957a), 'Information theory and statistical methods, I'. *Physics Reviews* 106:620-630.

Jaynes, E. T. (1957b), 'Information theory and statistical methods, II'. *Physics Reviews* 108:171-190.

Jones, P. and Salter, A. (2013), 'Modelling the economics of farm-based anaerobic digestion in a UK whole-farm context', *Energy Policy* 62: 215–225.

Kanellopoulos, A., Berentsen, P., Heckelei, T., Van Ittersum, M., & Lansink, A. O. (2010). Assessing the forecasting performance of a generic bio-economic farm model calibrated with two different PMP variants. *Journal of Agricultural Economics*, *61*(2), 274-294.

Kalantari, A.S., Cabrera, V.E. and Solis D. (2014), 'A comparison analysis of two alternative dairy cattle replacement strategies: optimization versus simulation models', *Economía Agraria Volumen* 18, 12–24.

Kenyon, D.E. (2001), 'Producer ability to forecast harvest corn and soybean prices', *Review of Agricultural Economics* 23: 151–162.

Kleinhanss, W. (2011), *Implementation, validation and results of the cost of production model using the EU FADN*, FACEPA Deliverable No. 3.2 (FP7/2007-2013).

Koots, K.R. and Gibson, J.P. (1998), 'Economic values for beef production traits from a herd level biophysical model', *Canadian Journal of Animal Science* 78: 29–45.

Kristensen, A.R. (1992), 'Optimal replacement in the dairy herd: a multi-component system', *Agricultural Systems* 39: 1.

Lelyon, B., Chatellier, V. and Daniel, K. (2010), 'Impact of decoupling and price variation on dairy farmers' strategy: overview of theoretical and real effects', paper presented at the OCED conference 'Evaluation of CAP reform at disaggregated level', Paris, France, 10–11 March 2010.

Lelyon, B., Chatellier, V. and Daniel, K. (2011), 'Decoupling and prices: determinant of dairy farmers' choices?', *Review of Agricultural and Environmental Studies* 92: 47–68.

Lence, H.S., and Miller, D.J. (1998a), 'Recovering Output-Specific Inputs from Aggregate Input Data: A Generalized Cross-Entropy Approach', *American Journal of Agricultural Economics*, 80: 852-867.

Lence, H.S., and Miller, D.J. (1998b), 'Estimation of Multi-Output Production Functions with Incomplete Data: A Generalized Maximum Entropy Approach', *European Review of Agricultural Economics*, 25: 188-209.

Léon, Y., Peeters, L., Quinqu, M. and Surry, Y. (1999), 'The use of maximum entropy to estimate input–output coefficients from regional farm accounting data', *Journal of Agricultural Economics* 50: 425–439.

LfL (2014), *Gruber Tabelle zur Fütterung in der Rindermast*, Bayerische Landesanstalt für Landwirtschaft, available online: www.lfl.bayern.de/mam/cms07/publikationen/daten/informationen/p_31941.pdf

Louhichi, K., Ciaian, P., Espinosa, M., Colen, L., Perni, A. and Gomez y Paloma, S. (2018a), 'Economic Impacts of CAP greening: an application of an EU-wide individual farm model for CAP analysis (IFM-CAP)', *European Review of Agricultural Economics* 45(2): 205–238.

Louhichi, K., Espinosa, M., Ciaian, P., Perni, A., Vosough Ahmadi, B., Colen, L. and Gomez y Paloma, S. (2018b), *The EU-Wide Individual Farm Model for Common Agricultural Policy Analysis (IFM-CAP v.1): Economic Impacts of CAP Greening*. JRC Technical Reports, EUR 28829 EN, European Commission, Joint Research Centre, available online: https://publications.jrc.ec.europa.eu/repository/handle/JRC108693

Louhichi, K., Ciaian, P., Espinosa, M., Colen, L., Perni, A. and Gomez y Paloma, S. (2017), 'Does the crop diversification measure impact EU farmers' decisions? An assessment using an individual farm model for CAP analysis (IFM-CAP)', *Land Use Policy* 66: 250–264.

Louhichi, K., Ciaian, P., Espinosa, M., Colen, L., Perni, A. and Gomez y Paloma, S. (2015), *An EU-wide individual farm model for common agricultural policy analysis (IFM-CAP). First application to crop diversification policy*, JRC Science and Policy Reports Series, No JRC92574, European Commission, Joint Research Centre, available online: http://publications.jrc.ec.europa.eu/repository/handle/JRC92574

Louhichi, K., Jacquet, F., and Butault, J.P. (2012), 'Estimating input allocation from heterogeneous data sources: A comparison of alternative estimation approaches', *Agricultural Economics Review,* Vol 13, No. 2.

Lusoli, W. (Ed.). (2020). *Reproducibility of Scientific Results in the EU: Scoping Report*. Publications Office of the European Union.

Mahy, L., Dupeux, B., Van Huylenbroeck, G. and Buysse J. (2015), 'Simulating farm level response to crop diversification policy', *Land Use Policy* 45: 36–42.

Markowitz, H. (1952), 'Portfolio selection', *The Journal of Finance* 7: 77–91.

M'barek, R., Barreiro-Hurle, J., Boulanger, P., Caivano, A., Ciaian, P., Dudu, H., Goded, M.E., Fellmann, T., Ferrari, E. and Paloma, S.G.Y. (2017). *Scenar 2030-Pathways for the European Agriculture and Food Sector beyond 2020*. EUR 28797 EN, Publications Office of the European Union, Luxembourg.

McCarl, B. A., & Spreen, T. H. (1997). Applied mathematical programming using algebraic systems. *Cambridge, MA*.

Mérel, P. and Bucaram, S. (2010), 'Exact calibration of programming models of agricultural supply against exogenous sets of supply elasticities', *European Review of Agricultural Economics* 37: 395–418.

Mérel, P., Simon, L.K. and Yi, F. (2011), `A fully calibrated generalized constant-elasticity of substitution programming model of agricultural supply', *American Journal of Agricultural Economics* 93(4): 936–948.

Mindmore, P. (1990), 'Estimating input-output coefficients from Regional Farm Data: A comment', *Journal of Agricultural Economics,* 41: 108-111.

Moxey, A. and Tiffin, R. (1994), 'Estimating linear production coefficients from Farm Business Survey Data: A note', *Journal of Agricultural Economics*, 45(3):381-385.

Nasuelli, N., Palladino, G., Setti, M., Zanasi, C. and Zucchi, G. (1997), *FEED MODULE: requirement functions and restriction factors*, CAPRI Working Paper 97–12, available online: http://www.ilr.uni-bonn.de/agpo/publ/workpap/pap97-12.pdf

Nerlove, M. (1958), 'Adaptive expectations and cobweb phenomena', *The Quarterly Journal of Economics* 72: 227–240.

Nerlove, M. and Bessler, D.A. (2001). 'Expectations, information and dynamics', in Gardner, B.L. and Rausser, G.C. (eds) *Handbook of Agricultural Economics, Vol 1A*, Amsterdam, The Netherlands, pp. 155–206.

Nicholson, C.F., Lee, D.R., Boisvert, R.N., Blake, R.W. and Urbina, C.I. (1994), 'An optimization model of the dual-purpose cattle production system in the humid lowlands of Venezuela', *Agricultural Systems* 46: 311–334.

Nielsen, B.K., Kristensen, A.R. and Thamsborg, S.M. (2005), 'Optimal decisions in organic steer production – a model including winter feeding level, grazing strategy and slaughtering policy', *Livestock Production Science* 88: 239–250.

NRC (1994), *Nutrient Requirements of Poultry: Ninth Revised Edition*, Subcommittee on Poultry Nutrition, Committee on Animal Nutrition, National Research Council. National Academy Press. Washington, D.C.

Paris, Q. and Howitt, R.E (1998), 'An Analysis of Ill-Posed Production Problems Using Maximum Entropy', *American Journal of Agricultural Economics*, Vol. 80, No. 1, pp. 124-138.

Paris, Q. and Arfini, F. (2000), 'Frontier cost functions, self-selection, price risk, PMP, and Agenda 2000', *Rivista di Economia Agraria 55*: 211–242.

Paris, Q. and Caputo, M.R. (2001), 'Sensitivity of the GME Estimates to Support Bounds', Working Paper No. 01-008*.* Department of Agricultural and Resource Economics, University of California*,* 18p.

Paris, Q. (2011), *Economic foundation of symmetric programming*, Cambridge University Press, New York, pp. 1–550.

Parnas, D. L. (1972). On the criteria to be used in decomposing systems into modules. In *Pioneers and Their Contributions to Software Engineering* (pp. 479-498). Springer, Berlin, Heidelberg.

Petsakos, A. and Rozakis S. (2015), 'Calibration of agricultural risk programming models', *European Journal of Operational Research* 242: 536–545.

Pratt, J.W. (1964), 'Risk aversion in the small and in the large', *Econometrica* 32: 122–136.

Kitzes, J., Turek, D., & Deniz, F. (2017). The Practice of Reproducible Research.

Ramsden, S., Gibbons, J. and Wilson, P. (1999), 'Impacts of changing relative prices on farm level dairy production in the UK', *Agricultural Systems* 62: 201–215.

Ray, S.C. (1985), 'Methods for estimating the input coefficients for linear programming models', *American Journal of Agricultural Economics,* 67: 660-665.

Russell, A. L. (2012). Modularity: An interdisciplinary history of an ordering concept. *Information & Culture*, *47*(3), 257-287.

Scarlat N., Martinov, M. and Dallemand, J.F. (2010), 'Assessment of the availability of agricultural crop residues in the European Union: Potential and limitations for bioenergy use', *Waste Management* 30: 1889–1897.

Sckokai, P. and Moro, D. (2006), 'Modelling the reforms of the common agricultural policy for arable crops under uncertainty', *American Journal of Agricultural Economics* 88: 43–56.

Shideed, K.H. and White, F.C. (1989), ‘Alternative forms of price expectations in supply analysis for U.S. corn and soybean acreages’, *Western Journal of Agricultural Economics* 14: 281–292

Solazzo, R. and Pierangeli, F. (2016), ‘How does greening affect farm behaviour? Trade-off between commitments and sanctions in the Northern Italy’, *Agricultural Systems* 149: 88–98.

Solazzo, R., Donati, M., Arfini, F. and Petriccione, G. (2014), ‘A PMP model for the impact assessment of the Common Agricultural Policy reform 2014–2020 on the Italian tomato sector’, *New Medit* 2: 9–19.

Stevens, W. P., Myers, G. J., & Constantine, L. L. (1974). Structured design. *IBM Systems Journal*, *13*(2), 115-139

Thorne, P.J., Hengsdijk, H., Janssen, S., Louhichi, K., van Keulen, H. and Thornton, P.K. (2009), *Modelling livestock component in FSSIM*, SEAMLESS report No 35, SEAMLESS Integrated Project, EU 6th Framework Programme, Contract No 010036-2, available online: www.SEAMLESS-IP.org.

van der Hoek, A., & Lopez, N. (2011, March). A design perspective on modularity. In *Proceedings of the tenth international conference on Aspect-oriented software development* (pp. 265-280).

Veysset, P., Bébin, D. and Lherm, M. (2005), ‘Adaptation to Agenda 2000 (CAP reform) and optimisation of the farming system of French suckler cattle farms in the Charolais area: a model-based study’, *Agricultural Systems* 83: 179–202.

Visagie, S.E. and Ghebretsadik, A.H. (2005), ‘Modelling risk in farm planning’, *Agrekon* 44: 561–585.

Vosough-Ahmadi, B., Shrestha, S., Thomson, S.G., Barnes, A.P. and Stott, A.W. (2015), ‘Impacts of greening measures and flat rate regional payments of the Common Agricultural Policy on Scottish beef and sheep farms’, *Journal of Agricultural Science* 153: 676–688.

Westbury, D. B., Park, J. R., Mauchline, A. L., Crane, R. T., & Mortimer, S. R. (2011). Assessing the environmental performance of English arable and livestock holdings using data from the Farm Accountancy Data Network (FADN). *Journal of Environmental Management*, *92*(3), 902-909.

Wolf, W. (1995), *SPEL system: Methodological documentation (Rev. 1)*, Basics, Vol. 1, BS, SFSS, Office for Official Publications of the European Communities, Luxembourg.

Zellner, A. (1971), *Introduction to Bayesian inference in econometrics*, Wiley, New York.

Annex A: Literature review of modelling livestock activities

Table A-1. Reviewed papers that used a normative approach

Authors	Year	Type of model	Model purpose	Farm speciality	Scale	Data used	Livestock activities	Purchase and sale	Calibration	Intergenerational dependences	
										Base	Simulation
Jones and Salter	2013	Normative, static annual LP models with embedded, anaerobic digestion (AD) enterprise	Economic evaluation of farm-based AD	Arable, dairy, AD enterprise	Representative arable farm in east England, a larger than average dairy farm in south-east England	Defra Farm Business Survey data, literature, farm management handbooks	Adult cow, replacement heifers, calves that are sold at 3 weeks	Selling beef crosses (calves) at age of 3 weeks		Adult cows endogenous, herd size constrained 0.5–2 LSU/ha; Calves and heifers as shares of adult	Same
Lelyon et al.	2011	Normative LP, static annual with four seasons per year	Analysis of response to decoupling and the price variation	Grass-based farm, semi-intensive, milk plus cereals and, milk plus young bulls	Farm level, plains regions, France	The annual survey of the Institute de l'Elevage (2008) with more than 600 dairy producers	Dairy cows, heifers, calves and young bulls	Selling female and male calves, buying male calves for fattening, selling bulls (fattened male calves)	Technical coefficients were modified (2005 basis)	Adults cows endogenous; calves as a share of adults cows; bought-in male calves endogenous	Same
Acs et al.	2010	Normative, static LP	Analysis of impacts of CAP decoupling on a range of farm types	Regional, farm level, farm types in marginal hill area of Peak District, UK	Dairy, beef, breeding sheep, forage	Survey of 44 farms that identified six representative farm types in 2006/2007	Suckler cows, dairy cows, calves, heifers	Purchase heifers, selling young beef calves, selling fattened calves, selling young dairy calves, selling young fat dairy calves	Results compared with surveyed farms data	Dynamics determined exogenously (replacement rate) and by selling activities endogenous)	Same
Crosson et al	2006	Normative static LP for beef farming systems	Adaptation to variations in prices, technical development, participation in an agri-environmental	Beef animal and forage production	Irish beef production systems	Grass production data from experiments Teagasc, for the period 2001–2004	Suckler beef cow (young and adult), replacement heifer, calf, yearlings, and finishing	Sale activities for weaners and store animals at various ages of fattening; only replacement heifers are purchased	Based on expert judgement and based on financial and technical criteria	Adult cows endogenous; calves and heifers based on exogenous factors such replacement rate, feed	Same

Authors	Year	Type of model	Model purpose	Farm speciality	Scale	Data used	Livestock activities	Purchase and sale	Calibration	Intergenerational dependences	
										Base	Simulation
			scheme				activities			requirements	
Havlík et al.	2006	Normative, static LP	Environmental analysis of organic suckler cow farms	Specialised suckler cow production, crop	Protected Landscape Area White Carpathians, Czech Republic	FADN CZ 2002	Suckler cow, weaners, heifers, bulls at different ages	Replacement heifers from own breeding but they can be sold, calves can be sold	Results compared with 2002 survey	Dynamic determined exogenously	Same
Visagie and Ghebretsadik	2005	Normative static LP	Modelling risk in farm planning	Crop, adult dairy cattle, young cattle, sheep for wool	Farm level Swartland, South Africa	One farm data	Adult cow, Adult sheep	Buying/selling adult/young cows and sheep		Adults cows endogenous; young cattle as a share of adults cows	Same
Ducros et al.	2005	Normative, static LP	Analysis of impact of policies such as stocking density and nitrogen balance on environmental and economic performance	Breeding dairy cattle, forage and apples	Farm level, mixed crop-livestock-orchard farming Normandy, France	Literature	Dairy cows, fattening calves and heifers	Sales are considered for all livestock categories	Test the coherence of technical coefficients used in the model with data from surveyed farm.	Dynamics determined exogenously (replacement rate) and by selling activities (endogenous)	Same
Veysset et al.	2005	Normative, static LP with two seasons of summer and winter	Analysis of economic adaptation of two farm types to Agenda 2000 CAP reform	A mixed crop-livestock farm and a livestock farm	Farm-level Charolais suckler cattle, Northern from Massif Central, France	Data of 20 years from 90 Charolais suckler farms from three regions	Suckler cows, male and female calves, heifers	Male calves sold as store and fattened, female calves sold as store and fattened; 33-month heifers could be bought in	Based on expert judgement and based on four observed activities	Adult cows endogenous; calves determined by share (exogenous)	same
Ramsden et al.	1999	Normative, static annual	To evaluate the impact of changes in milk to milk-quota-leasing price ratios, nitrogen	Dairy cow, beef cow, forage	Farm level (only one farm modelled), dairy sector, UK	Literature	Dairy cows, heifers and calves. Cows have five milk production levels. Male animals from	Heifers bought in as 2-year olds, female calves can be sold	Results compared with actual farm data for England and Wales based on a survey	Adult cows endogenous, calves and heifers based on exogenous data such as	Same

Authors	Year	Type of model	Model purpose	Farm speciality	Scale	Data used	Livestock activities	Purchase and sale	Calibration	Intergenerational dependences	
										Base	Simulation
			fertiliser and concentrate				dairy go to beef			replacement rate	
Nicholson et al.	1994	Normative, multi-period LP	Analyse alterative nutritional management strategies	Farm level, representative lowlands of western region	Dairy, beef cattle, forage mixed milk-meat cattle farms, Venezuela	Data from 22 farms surveyed in the study region	Three cow status, one calves, two age groups heifers, three age groups steers	No purchase of animals modelled, but all animal categories can be sold	Validation by construct		Same

Source: Louhichi et al. (2018b)

Table A-2. Reviewed papers that used a positive approach

Authors	Year	Type of model	Model purpose	Farm speciality	Scale	Data used	Livestock Activities	Purchase and sale	Calibration	Intergenerational dependences		Observed behaviour at simulation
										Base	Simulation	
Gill et al.	2015	PMP	Assessment of policy and price changes on hog sector	Crops, beef, breeding sow and growing pigs	Provincial, Canada	Regional data	Sows and growing pigs	Sows culled/move next cycle. Growers Slaughtered/exported as live animals or replace culled sows	Quadratic cost function (Howitt, 1995)	Sows exogenous; growers based on sows and farrowing cycles/year ; PMP applied to growers.	Same; PMP terms removed	Ratio of growers to sows, replacement rates, market hogs per sow, birth rates and death are exogenous
Jitea et al.	2015	PMP	*Ex-ante* analysis of 2014 CAP reform, land abandonment	Crop and livestock	On region in north-western Romania (NUTS2)	Independent survey (207 farms)	Dairy (m/f, age), beef, sheep, goat, pig	Selling meat and milk. They don't mention any purchase activity	quadratic cost function (Howitt ,1995)	Yes, exogenous parameters, such as fertility rate and replacement rate	Same	
Fragoso et al.	2011	PMP	Assessment of the effects of CAP on farm income, land, labour and	Forestry, beef cattle, sheep, extensive swine	Regional, Alentejo, Portugal	Regional data from the Official Network of Agricultural Account data	Beef (breeding, calves), sheep, swine	No purchase, but sale	Quadratic cost function (Howitt, 1995)	No	No	

Authors	Year	Type of model	Model purpose	Farm speciality	Scale	Data used	Livestock Activities	Purchase and sale	Calibration	Intergenerational dependences		Observed behaviour at simulation
										Base	Simulation	
			capital			(RICA)						
Thorne et al.	2009	PMP	*Ex-ante* policy analysis	Crop and livestock	EU	FADN	Dairy, suckler, beef, sheep, goat	Only dressed animals (i.e. breeding adults)	quadratic cost function (Howitt 1995)	Yes, based on shares, static	Same	
Judez et al.	2001	PMP	*Ex-ante* analysis of agenda 2000	Crops and Beef and Veal	Regional farm types in Spain	Regional FADN	Suckler cows & young male	Selling 1< young male cattle	quadratic cost function (Howitt 1995)	Yes; share of young per cow	same	an increase of suckler cows and a decrease of young males

Source: Louhichi et al. (2018b)

Annex B: Animal feed requirement functions in IFM-CAP

This annex presents the functions used in IFM-CAP to determine the nutrient requirement by animal category. These requirement functions are based on CAPRI (Nasuelli et al., 1997; IPCC, 2006; Britz and Witzke, 2014), LfL (2014); GfE (2006) and NRC (1994).

1. Dairy cows (DCOW) and suckler cows (SCOW)

The dry matter requirements for cows include minimum dry matter for lactation per year (*DRMN*) and maximum dry matter (*DRMX*):

$$DRMN_i = 305(0.0185LW_i + 0.305MC_i) + 60(0.0185LW_i)$$

$$DRMX_i = DRMN_i * 1.2$$

The energy requirements of cows are expressed as net energy for each of the main biological functions of the animal including lactation, maintenance, activity, pregnancy and growth: net energy lactation (*NEL*), net energy maintenance (*NEM*), net energy activity (*NEA*), net energy pregnancy (*NEP*) and net energy for growth (*NEG*).

$$NEL_i = (0.4 \cdot MC_i + 1.47)MPD_i LP_i$$

$$NEM_i = 0.17(0.386LW_i^{0.75}) \cdot 365$$

$$NEA_i = 0.17(0.386LW_i^{0.75})365 \text{ if there is grassland on-farm; otherwise } NEA_i = 0$$

$$NEP_{DCOW} = (LW^{0.75} \cdot 0.386) \cdot 0.10 \cdot 365$$

$$NEP_{SCOW} = (LW^{0.75} \cdot 0.386) \cdot 0.\frac{10 * CALV_{SC}}{1000}$$

The total net energy (*ENNE*) is the sum of the above energy components:

$$ENNE_i = NEL_i + NEM_i + NEA_i + NEP_i$$

The crude protein requirement (*CRPR*) is calculated as follows:

$$CRPR_{DCOW} = \frac{14 \cdot MC_{DCOW} + 28}{1000} MPD_{DCOW} LP_{DCOW} + \frac{117 + 0.6 \cdot LW_{DCOW}}{1000} 365 + 1300 \cdot 42$$

$$CRPR_{SCOW} = \frac{14MC_{DCOW} + 28}{1000} MPD_{DCOW} LP_{DCOW} + \frac{1.27 + LW_{DCOW} + 127.3}{1000} \cdot 365 + 1300 \cdot 42$$

The requirements for fibre digestibility (*FIDI*), maximum fibre (*FICO*) and fibre long (*FILG*) are:

$$FIDI_i = DRMN_i \left(\frac{MC_i}{100} - 0.29 + \frac{43.92}{100}\right)$$

$$FICO_i = 0.7 \cdot (0.14 \cdot LW_i^{0.75}) \cdot 365$$

$$FILG_i = \frac{FIDI_i}{3}$$

where subscripts *DCOW* and *SCOW* stand for dairy cow and suckler cow, respectively, *i* = *DCOW*, *SCOW*; $CALC_{SG}$ is calves per cow; and *MC* is adjusted milk production per day corrected by fat milk content (*MF*). *MF* is extracted from Eurostat. *MC* depends on milk production per day (*MPD*), which it is derived from FADN; *COMI* and *COMF* are, respectively, milk production for feeding and milk production (not for feeding) for suckler/dairy cows derived from FADN (in kg per day). The raising period (*PD*) is 365 days, of which the duration of lactation (*LP*) is assumed to be 305 days for dairy cows and 125 days for suckler cows. The mean live weight (*LW*) is calculated by dividing the selling value of cows available from FADN by the meat price obtained from Eurostat.

$$MC_i = MPD_i(0.4 + 0.15 \cdot MF_i)$$

$$MPD_i = \frac{(COMI_{DCOW/SCOW} + COMF_{DCOW})}{125}$$

Lower and upper bounds of nutrient requirements for dairy and suckler cows are obtained by varying the fat milk content (*MF*) and the mean live weight (*LW*) by three standard deviations around their mean values. The lower and upper bounds represent the interval within which the actual animal requirements most probably lie.

2. Fattening and raising of calves (CAMF/CAFF; CAMR/CAFR)

The nutrient requirements for fattening of male (*CAMF*) and female (*CAFF*) calves and raising of male (*CAMR*) and female (*CAFR*) calves are taken from LfL (2014) and are reported in Table B-1 and Table B-2. The requirement tables deliver the requirement on a daily basis from birth to about 800 days. The table provides average values as well as minimum (*min*) and maximum (*max*) values for daily weight increase (*DAILY*) (kg/day), animal weight in a given day (*XALW*) and nutrient requirements (*ENNE*, *ENMR*, *CRPR*, *DRMA*, *DRMN*, *DRMX*). The minimum and maximum values are used to derive the lower and upper bounds of animal requirements.

The annual requirements are calculated from Table B-1 and Table B-2 as follows:

$$X_i = \frac{\sum_{DAY=startD}^{DAY=endD} X_{i,DAY}}{endD - startD} \qquad \text{for } i = CAMF, CAFF, CAMR, CAFR$$

where *DAY* stands for day for *DAY*0 to *DAY*805; *startD* is start day of the fattening/raising process and *endD* is end day of the fattening/raising process; and *X* = *ENNE*, *ENMR*, *CRPR*, *DRMA*, *DRMN*, *DRMX.*

The equations above calculate annual nutrient requirements for fattening and raising calves. All animal categories in IFM-CAP represent average number of head available on-farm in a year. This implies 365 production days for all animal categories including fattening of calves. The total requirements are calculated by multiplying the average nutrient per day by 365 days.

The mean value of the start day (*startD*) and end day (*endD*) are defined based on the FADN definition. The start day for fattening (raising) calves is assumed to be day zero, whereas the end day is set to 60 (180) days (Table B-3). However, the actual start and end day of calve activities of farms may depart significantly from the mean values. To account for this uncertainty we consider lower and upper bounds for these two parameters as defined in Table B-3.

The annual requirements for fibre are defined as follows:

$$FICT_i = DRMX_i \qquad \text{for } i = CAMF, CAFF, CAMR, CAFR$$

The lower and upper bounds of animal requirements for fattening and raising of calves are obtained by using minimum and maximum values of nutrient requirements as reported in Table B-1 and Table B-2 as well as by varying the start and end day as reported in Table B-3. Note that this variation implicitly implies that the daily weight increase (*DAILY*) (kg/day) and animal weight in a given day (*XALW*) (kg/animal) also change as reported in Table B-1 and Table B-2 following the duration of the fattening and the raising process defined by the minimum and maximum values of start and end days in Table B-3.

3. Adult cattle fattening (BULF, HEIF) and heifers for breeding (HEIR)

Similarly to calf activities, the nutrient requirements for male and female adult cattle fattening (*BULF, HEIF)* and heifers for breeding (*HEIR)* are taken from LfL (2014) and are reported in Table B-1 and Table B-2. The annual requirements are calculated from those tables as follows:

$$X_i = 365 \cdot \frac{\sum_{DAY=startD}^{DAY=endD} X_{i,DAY}}{endD - startD} \qquad \text{for } i = BULF, HEIF, HEIR;\ X = ENNE, ENMR, CRPR, DRMA, DRMN, DRMX$$

The annual requirements for fibre are defined as follows:

$$FICT_i = DRMX_i \qquad \text{for } i = BULF, HEIF, HEIR$$

To obtain the end day (*endD*) of the adult cattle, first the mean live weight (*LW*) is calculated by dividing the selling value of adult cattle (*PRIC*) available from FADN by the meat price extracted from Eurostat. Second, to obtain the end day of the fattening process, we use the animal weight in a given day (*XALW*) from Table B-1 (for *BULF*) and Table B-2 (for *HEIF* and *HEIR*) to derive the period (days) needed to reach the derived mean live weight *LW*. The start day *startD* of the fattening process of *BULF*, *HEIF* and *HEIR* is defined based on the FADN definition and is reported in Table B-3.

To account for the uncertainty in the data, we vary the mean live weight by three standard deviations, which results in variation of the end day *endD* based on the corresponding information on *XALW* available from Table B-1 and Table B-2. The start day is varied as reported in Table B-3. The nutrient requirements are varied using the minimum and maximum values as reported in Table B-1 and Table B-2, while the start day is varied as reported in Table B-3.

4. Sows (SOWS)

The nutrient requirements for sows (*SOWS*) are taken from GfE (2006) and are reported in Table B-4. The table reports the daily nutrient needs over the whole year (365 days). The table assumes that the requirements of a medium breeding performance are independent of the number of piglets, as there is no reliable relation between number of piglets and milk yield described in GfE (2006).

The annual requirements are calculated from Table B-4 as follows:

$$X_{SOWS} = \sum_{DAY=0}^{DAY=364} X_{DAY} \qquad \text{for } X = ENNE, ENMP, CRPR$$

The minimum (*DRMN*) and maximum (*DRMX*) requirements of dry matter are calculated as follows:

$$DRMN_{SOWS} = \frac{ENMP_{SOWS}}{14.82}$$

$$DRMX_{SOWS} = \frac{ENMP_{SOWS}}{13.47}$$

The minimum and maximum values reported in Table B-4 are used to derive the lower and upper bounds of animal requirements. This is to account for the uncertainty in the underlying data (e.g. number of piglets, milk yield).

5. Fattening of pigs (PIGF)

Nutrients for fattening pigs are calculated by summing up the nutrient requirements over the growth period of pigs from the start day until the end day of the fattening process. The nutrient requirements are taken from GfE (2006) and are reported in Table B-5. The annual requirements are calculated from this table as follows:

$$X_{PIGF} = 365 \cdot \frac{\sum_{DAY=startD}^{DAY=endD} X_{DAY}}{endD - startD} \qquad \text{for } X = ENNE, ENMP, CRPR$$

The minimum (*DRMN*) and maximum (*DRMX*) requirements of dry matter are calculated as follows:

$$DRMN_{PIGF} = \frac{ENMP_{PIGF}}{(13.4 \cdot 0.88 \cdot 0.588)}$$

$$DRMX_{PIGF} = \frac{ENMP_{PIGF}}{(12.6 \cdot 0.88 \cdot 0.588)}$$

As for adult cattle, to obtain the end day (*endD*) of the pig-fattening process, first the mean live weight (*LW*) is calculated by dividing the selling value of pigs available from FADN by the meat price extracted from Eurostat. Then, to obtain the end day of the fattening process, we use the animal weight in a given day (*XALW*) from Table B-5 to derive the period (days) needed to reach the derived mean live weight *LW*. The start day *startD* of the fattening process is defined based on the FADN definition and is reported in in Table B-3.

To account for the uncertainty in the data, we vary the mean live weight by three standard deviations, which results in variation of the end day based on the corresponding information on *XALW* available from Table B-5. The nutrient requirements are varied using the minimum and maximum values as reported in Table B-5, while the start day is varied as reported in Table B-3. Note that the main source of variation of requirements (around minimum and maximum values) is daily live weight gains of pigs. The growth rate of pigs can strongly vary across MS and across farms within a MS. Moreover, the relative ratios of different nutrient requirements vary across different growth stages of pigs. The minimum and maximum values of requirements reported in Table B-5take into consideration both these sources of variation and are available from GfE (2006).

6. Laying hens (HENS)

$$ENMC_{HENS} = 365 \cdot (0.46 \cdot LW_{HENS} + 0.57 \cdot EGGY_{HENS}) \cdot 1000 \cdot k_{ENMC}$$

$$DRMA_{HENS} = \frac{ENMC_{HENS}}{12}$$

$$DRMN_{HENS} = \frac{ENMC_{HENS}}{15}$$

$$DRMX_{HENS} = \frac{ENMC_{HENS}}{8}$$

$$CRPR_{HENS} = 0.14 \cdot \frac{ENMC_{HENS}}{11.1}$$

$$LISI_{HENS} = 0.0095 \cdot EGGS_{HENS} + 1.9 \cdot 60$$

where

$$EGGY_{HENS} = \frac{EGGS/57}{365}$$

ENMC is metabolisable energy for chicken*;* *EGGY* is number of eggs per laying hen per day with the assumption of average egg weight of 57 g and 365 production days; *EGGS* is egg production (in kg per 1 000 heads); LW_{HENS} is mean live weight assumed 1.62 kg, k_{ENMC} is unit conversion factor for energy requirements (Table B-8).

7. Poultry (POUF)

The nutrient requirements for poultry are taken from NRC (1994) and are reported in Table B-6. The requirement tables are for broilers and provide nutrient requirement on a daily basis from birth until the end day of the production process. The annual requirements are calculated from this table as follows:

$$X_{POUF} = 365 \cdot \frac{\sum_{DAY=startD}^{DAY=endD} X_{DAY}}{endD - startD} \quad \text{for } X = ENNE, ENMC, CRPR, DRMA$$

The mean value of the start day (*startD*) and end day (*endD*) of the production process are defined based on FADN definition. The start day is assumed zero, whereas the end day is set to 40 days (Table B-3). To account for the uncertainty, we consider lower and upper bounds of the start and end day as defined in Table B-3.

The minimum (*DRMN*) and maximum (*DRMX*) requirements of dry matter are calculated as follows:

$$DRMN_{POUF} = \frac{ENMC_{POUF}}{(13.4 \cdot 0.88 \cdot 0.717)}$$

$$DRMX_{POUF} = \frac{ENMC_{POUF}}{(12.6 \cdot 0.88 \cdot 0.717)}$$

To account for the uncertainty in the data, the lower and upper bounds of poultry nutrient requirements are obtained by using minimum and maximum values of nutrient requirements as reported in Table B-6 as well as by varying the start and end day as reported in Table B-3. The main sources of uncertainty in deriving the poultry requirements are the duration of production process, the type of poultry (e.g. broiler, turkey) and daily growth rate.

8. Ewes and goats for milk (SHGM)

First, nutrient requirements are calculated for ewes (*EWES*) and goats (*GOAT*) separately, second, the nutrient requirements for the combined sheep and goat activity (*SHGM*) are obtained as the weighted average over *EWES* and *GOAT*.

8.1 Nutrient requirements for EWES and GOAT

$$NEM_i = 0.217 \cdot LW_i^{0.75} 0.10 \cdot 7PD_i$$

$NEA_i = 0.0107\, LW_i PD_i$ if there is grassland on-farm; otherwise $NEA_i = 0.009 LW_i PD_i$

$$NEL_i = 4.6 \cdot MPD_i \cdot 170$$

$$ENNE_i = NEM_i + NEA_i + NEL_i$$

$$CRPR_{EWES} = 135\,(0.026 + 0.0014 LW_{EWES}) + 170(0.0634 + 0.0012 LW_{EWES} + 0.0895 MPD_{EWES}) + 1.35 \cdot (2.22 \cdot LW_{EWES} - 19.88) \cdot 60 \frac{1}{1000}$$

$$CRPR_{GOAT} = 305 \cdot (12.66 + 0.8 \cdot LW_{GOAT}) + 61 \cdot MPD_{GOAT} 170 + \frac{60 \cdot (1.425 \cdot LW_{GOAT} + 14.666)}{1000}$$

$$\begin{aligned} DRMN_{EWES} = {} & 135 \cdot (0.36 + 0.023 \cdot LW_{EWES}) + 170 \cdot (1.112 + 0.0187 \cdot LW_{EWES} + 0.279 \cdot MPD_{EWES} \\ & + 60(0.0268 \cdot LW_{EWES} - 0.24) \end{aligned}$$

$$DRMN_{GOAT} = 305 \cdot (0.55 + 0.013 \cdot LW_{GOAT}) + 0.3 \cdot MPD_{GOAT} 170 + 60 \cdot (0.0122 \cdot LW_{GOAT} + 0.5316)$$

where

$$MPD_i = \frac{SGMI_i + SGMF_i}{170}$$

i = EWES, GOAT; *MPD* is sheep/goat milk production per day. It is assumed that there are 170 milk production days, 135 days maintenance only and 60 days of final mating; *SGMI* is milk production per sheep/goat; *SGMF* and SGMI are milk production for feeding and milk production (not for feeding) for sheep and goats, respectively, derived from FADN (in kg per day); $PD_i = 365$; $LW_{EWES} = 55$; $LW_{GOAT} = 60$.

8.2 Nutrient requirements for sheep and goat activity (SHGM)

$$REQ_{SHGM} = sh_{EWES} REQ_{EWES} + sh_{GOAT} REQ_{GOAT}$$

$$DRMX_{SHGM} = 1.5 \cdot DRMN_{SHGM}$$

$$FISM_{SHGM} = 120 \cdot \frac{LW_{SHGM}^{0.75}}{1000} \cdot 365$$

where

$$LW_{SHGM} = sh_{EWES} LW_{EWES} + sh_{GOAT} LW_{GOAT}$$

REQ = *ENNE*, *CRPR*, *DRMN*; *FISM* is fibre for sheep and goats; sh_{EWES} and sh_{GOAT} are shares of ewes and goats in the total herd size, respectively, derived from FADN.

The lower and upper bounds of requirements are obtained by varying the average milk production per day and the mean live weight by 30 % around their mean values.

9. Sheep and goats fattening (SHGF)

$$ENNE_{SHGF} = (0.1596 \cdot LW_{SHGF} + 0.0303 \cdot DAILY_{SHGF} - 0.56)(1 - 0.2) FD_{SHGF} k_{ENMR}$$

$$CRPR_{SHGF} = \frac{(21.778 + 0.33 \cdot LW_{SHGF}) + 0.258 \cdot DAILY_{SHGF} 1.35 \cdot FD_{SHGF} \cdot 1000}{1000}$$

$$DRMN_{SHGF} = (0.038286 \cdot LW_{SHGF} + 0.06381) FD_{SHGF}$$

$$DRMX_{SHGF} = 1.5 \cdot DRMN_{SHGF}$$

$$FISF_{SHGF} = 0.075 \cdot LW_{SHGF}^{0.75} FD_{SHGF}$$

where

$$FD_{SHGF} = Min\left[320; Max\left(45; \frac{SGMT_{SHGF}/CW_{SHGF}}{DAILY_{SHGF}}\right)\right]$$

$$LW = \frac{\frac{Max[8; Min(25; SGMT_{SHGF})]}{CW_{SHGF}}}{2}$$

SGMT is meat production per animal; *CW* = 0.6; *DAILY* = 0.250 kg; k_{ENMR} is conversion factor for metabolisable energy ruminants (*ENMR*) (Table B-7 and Table B-8).

The lower and upper bounds of requirements are obtained by varying the duration of the fattening period and the mean live weight up to 80 % around their mean values.

Table B-1. Nutrient requirement table for male cattle fattening and raising (CAMF, CAMR, BULF)

		DAY 0	DAY 1	DAY 2	DAY 3	...	DAY 803	DAY 804	DAY 805
DAY	Average		1	2	3		803	804	805
	Min		1	2	3		803	804	805
	Max		1	2	3		803	804	805
DAILY	Average	800	800	800	800		1412	1412	1412
	Min	690	690	690	690		1290	1290	1290
	Max	800	800	800	800		1506	1506	1506
XALW	Average	80	80.8	81.6	82.4		1177.937	1179.349	1180.761
	Min	80	80.69	81.38	82.07		813.052	814.342	815.632
	Max	80	80.8	81.6	82.4		1259.177	1260.683	1262.189
ENNE	Average	10.659	10.659	10.659	10.659		75.4908	75.4908	75.4908
	Min	9.405	9.405	9.405	9.405		70.9137	70.9137	70.9137
	Max	11.286	11.286	11.286	11.286		78.375	78.375	78.375
ENMR	Average	17	17	17	17		120.4	120.4	120.4
	Min	15	15	15	15		113.1	113.1	113.1
	Max	18	18	18	18		125	125	125
CRPR	Average	0.239	0.239	0.239	0.239		1.213	1.213	1.213
	Min	0.213	0.213	0.213	0.213		1.213	1.213	1.213
	Max	0.265	0.265	0.265	0.265		1.32	1.32	1.32
DRMA	Average	1.05	1.05	1.05	1.05		10.03333	10.03333	10.03333
	Min	0.95	0.95	0.95	0.95		9.466667	9.466667	9.466667
	Max	1.15	1.15	1.15	1.15		10.33333	10.33333	10.33333
DRMX	Average	1.2	1.05	1.05	1.05		10.03333	10.03333	10.03333
	Min	1.1	0.95	0.95	0.95		9.466667	9.466667	9.466667
	Max	1.3	1.15	1.15	1.15		10.33333	10.33333	10.33333
DRMN	Average	0.9	1.05	1.05	1.05		10.03333	10.03333	10.03333
	Min	0.8	0.95	0.95	0.95		9.466667	9.466667	9.466667
	Max	1	1.15	1.15	1.15		10.33333	10.33333	10.33333

Source: LfL (2014).

Table B-2. Nutrient requirement table for female cattle fattening and raising (CAFF, CAFR, HEIF, HEIR)

		DAY 0	DAY 1	DAY 2	DAY 3	…	DAY 803	DAY 804	DAY 805
DAY	Average		1	2	3		803	804	805
	Min		1	2	3		803	804	805
	Max		1	2	3		803	804	805
DAILY	Average	690	690	690	690		825	825	825
	Min	690	690	690	690		825	825	825
	Max	690	690	690	690		825	825	825
XALW	Average	80	80.69	81.38	82.07		805.8	806.625	807.45
	Min	80	80.69	81.38	82.07		799.77	800.595	801.42
	Max	80	80.69	81.38	82.07		810.71	811.535	812.36
ENNE	Average	9.405	9.405	9.405	9.405		62.7	62.7	62.7
	Min	9.405	9.405	9.405	9.405		59.565	59.565	59.565
	Max	9.405	9.405	9.405	9.405		65.835	65.835	65.835
ENMR	Average	15	15	15	15		100	100	100
	Min	15	15	15	15		95	95	95
	Max	15	15	15	15		105	105	105
CRPR	Average	0.213	0.213	0.213	0.213		1.149	1.149	1.149
	Min	0.213	0.213	0.213	0.213		1.092	1.092	1.092
	Max	0.213	0.213	0.213	0.213		1.205	1.205	1.205
DRMA	Average	0.95	0.95	0.95	0.95		10.5	10.5	10.5
	Min	0.95	0.95	0.95	0.95		10.5	10.5	10.5
	Max	0.95	0.95	0.95	0.95		10.5	10.5	10.5
DRMX	Average	1.1	0.95	0.95	0.95		10.5	10.5	10.5
	Min	1.1	0.95	0.95	0.95		10.5	10.5	10.5
	Max	1.1	0.95	0.95	0.95		10.5	10.5	10.5
DRMN	Average	0.8	0.95	0.95	0.95		10.5	10.5	10.5
	Min	0.8	0.95	0.95	0.95		10.5	10.5	10.5
	Max	0.8	0.95	0.95	0.95		10.5	10.5	10.5

Source: LfL (2014).

Table B-3. The start and the end day of the fattening/raising process of animal activities

	Start day (*startD*)			End day (*endD*)		
	Average	Min	Max	Average	Min	Max
CAFF	0	0	60	180	60	240
CAMF	0	0	60	180	60	240
CAFR	0	0	180	365	180	912.5
CAMR	0	0	180	365	180	912.5
HEIR	365	180	730	Calc.	s.d.	s.d.
HEIF	180	60	360	Calc.	s.d.	s.d.
BULF	272.5	120	545	Calc.	s.d.	s.d.
PIGF	0	0	17	Calc.	s.d.	s.d.
POUF	0	0	10	40	30	62

Notes: Calc.: calculated based on the mean live weight derived from FADN and Eurostat and corresponding values of *endD* from Table B-1 and Table B-2; s.d.: calculated based on the standard deviation of the *endD*.

Source: derived based on FADN definitions

Table B-4. Nutrient requirement table for sows (SOWS)

		DAY 0	DAY 1	DAY 2	DAY 3	…	DAY 362	DAY 363	DAY 364
DAY	Average		1	2	3		362	363	364
	Min		1	2	3		362	363	364
	Max		1	2	3		362	363	364
ENNE	Average	24.5196	24.5196	24.5196	24.5196		24.5196	24.5196	24.5196
	Min	23.0496	23.0496	23.0496	23.0496		23.0496	23.0496	23.0496
	Max	25.9896	25.9896	25.9896	25.9896		25.9896	25.9896	25.9896
ENMP	Average	41.7	41.7	41.7	41.7		41.7	41.7	41.7
	Min	39.2	39.2	39.2	39.2		39.2	39.2	39.2
	Max	44.2	44.2	44.2	44.2		44.2	44.2	44.2
CRPR	Average	0.33	0.33	0.33	0.33		0.33	0.33	0.33
	Min	0.3	0.3	0.3	0.3		0.3	0.3	0.3
	Max	0.36	0.36	0.36	0.36		0.36	0.36	0.36

Source: GfE (2006).

Table B-5. Nutrient requirement table for fattening of pigs (PIGF)

		DAY 0	DAY 1	DAY 2	DAY 3	…	DAY 175	DAY 176	DAY 177
DAY	Average		1	2	3		175	176	177
	Min		1	2	3		175	176	177
	Max		1	2	3		175	176	177
DAILY	Average	600	600	600	600		700	700	700
	Min	600	600	600	600		700	700	700
	Max	700	700	700	700		800	800	800
XALW	Average	20	20.6	21.2	21.8		150.2	150.9	151.6
	Min	20	20.6	21.2	21.8		150.2	150.9	151.6
	Max	20	20.7	21.4	22.1		166.9	167.7	168.5
ENNE	Average		7.644	7.644	7.644		21.168	21.168	21.168
	Min		7.644	7.644	7.644		21.168	21.168	21.168
	Max		8.82	8.82	8.82		22.932	22.932	22.932
ENMP	Average		13	13	13		36	36	36
	Min		13	13	13		36	36	36
	Max		15	15	15		39	39	39
CRPR	Average		0.202353	0.202353	0.202353		0.225882	0.225882	0.225882
	Min		0.202353	0.202353	0.202353		0.225882	0.225882	0.225882
	Max		0.235294	0.235294	0.235294		0.254118	0.254118	0.254118

Source: GfE (2006).

Table B-6. Nutrient requirement table for poultry (POUF)

		DAY 0	DAY 1	DAY 2	DAY 3	…	DAY 60	DAY 61	DAY 62
DAY	Average		1	2	3		60	61	62
	Min		1	2	3		60	61	62
	Max		1	2	3		60	61	62
DAILY	Average	21.14	21.14	21.14	21.14		57.86	57.86	57.86
	Min	21.14	21.14	21.14	21.14		57.86	57.86	57.86
	Max	21.14	21.14	21.14	21.14		57.86	57.86	57.86
XALW	Average		0.02	0.04	0.06		3.02	3.08	3.14
	Min		0.02	0.04	0.06		3.02	3.08	3.14
	Max		0.02	0.04	0.06		3.02	3.08	3.14
ENNE	Average	0.18	0.18	0.18	0.18		1.94	1.94	1.94
	Min	0.18	0.18	0.18	0.18		1.94	1.94	1.94
	Max	0.18	0.18	0.18	0.18		1.94	1.94	1.94
ENMC	Average	0.25	0.25	0.25	0.25		2.70	2.70	2.70
	Min	0.25	0.25	0.25	0.25		2.70	2.70	2.70
	Max	0.25	0.25	0.25	0.25		2.70	2.70	2.70
CRPR	Average	0.00	0.00	0.00	0.00		0.04	0.04	0.04
	Min	0.00	0.00	0.00	0.00		0.04	0.04	0.04
	Max	0.00	0.00	0.00	0.00		0.04	0.04	0.04
DRMA	Average	0.02	0.02	0.02	0.02		0.18	0.18	0.18
	Min	0.02	0.02	0.02	0.02		0.18	0.18	0.18
	Max	0.02	0.02	0.02	0.02		0.18	0.18	0.18

Source: NRC (1994).

Table B-7. Carcass share, live start weight and coefficient of energy for growth

	Carcass to live weight (*CW*) ***Coeff. 0-1***
SHGF	0.60
HENS	0.80

Source: CAPRI.

Table B-8. Conversion factors for energy requirements (KENMR, KENMC, KENMH, KENMP)

ENMR	**ENMC**	**ENMH**	**ENMP**
0.627	0.717	0.631	0.588

Source: CAPRI.

Annex C. Agri-environmental indicators

Table C-1 reports the state of implementation of the environmental indicators in the current version of the IFM-CAP model.

Table C-1. Agri-environmental indicators in the IFM-CAP model

Domain/dimension	Indicator	Sub-indicator	Status
Public policy	Agri-environmental commitments	Agri-environmental payments per hectare	Operational
Market signals and production systems	Intensification/extensification	Input expenditure	Operational
	Intensification/extensification	Low, medium and high input expenditure	Not implemented
Climate change and air	Greenhouse gas emissions	Methane and nitrous oxide	Implemented but not yet operational
	Ammonia emissions	Ammonia	Implemented but not yet operational
Water	Nutrient management	Nitrogen budget	Implemented but not operational
		Phosphorus budget	Implemented but not operational
	Nutrient management	Fertiliser consumption	Operational
		Fertiliser expenditure	Operational
	Pesticide risk	Expenditure on plant protection products	Operational
Soil	Soil erosion by water	Soil loss equation	Not yet implemented. Need for farm spatial allocation
	Soil erosion by water	Crop system and support	Operational

		practices factor	
	Soil organic matter	Soil organic matter	To be designed
	Soil organic matter	Share of permanent grassland	Operational
Biodiversity and landscape	Crop richness	Crop richness from functional crop groups	Operational
	Crop diversity	Crop diversity from functional crop groups	Operational
	Diversity of land uses		Cannot be implemented until shifts between land uses are included in the model
	Extensiveness	Extensiveness in arable land	Cannot be implemented until farm yields are endogenised
		Extensiveness in grassland	Implementation forthcoming
	Extensive permanent grasslands	Share of extensive permanent grasslands	Operational
	Environmental compensation zones (ECZs)	Share of ECZs in the UAA	Cannot be implemented until shifts between land uses are included in the model

List of abbreviations

CAP	common agricultural policy
CAPRI	common agricultural policy regionalised impact
CARA	constant absolute risk aversion
CPU	central processing unit
EFA	ecological focus area
FADN	farm accountancy data network
FOC	first-order condition
GAMS	general algebraic modelling system
GDX	GAMS Data eXchange
GGIG	GAMS graphical user interface generator
GUI	graphical user interface
HPD	highest posterior density
IFM-CAP	individual farm model for common agricultural policy
IQR	interquartile range
LP	linear programming
MS	Member State
NUTS	Nomenclature of Territorial Units for Statistics
PMP	positive mathematical programming
RCR	residue-to-crop ratio
RRA	relative risk aversion
UAA	utilised agricultural area
VCS	voluntary coupled support

List of figures

List of tables

KJ-NA-30949-EN-N

The European Commission’s science and knowledge service

Joint Research Centre

JRC Mission

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

EU Science Hub
ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub - Joint Research Centre

EU Science, Research and Innovation

EU Science Hub

Publications Office
of the European Union

doi:10.2760/248136

ISBN 978-92-76-46336-8