Co-funded by
EUROPEAN COMMISSION
EUROSTAT

Grant Agreement 08441.2017.008-2018.0217 Norway

Final Report

Date:	2020.10.30
Version:	0.2
Authors:	Ole Rognstad, Geir Inge Gundersen, Tatsiana Pekarskaya, Leiv Solheim and Stein Olav Kolle

Table of Contents

Executive summary

This document gives an overview of results achieved from the project "Modernisation, improvement and integration of statistics related to Integrated Farm Statistics and nutrient budgets". The final total duration of the project was 26 months.

Introduction

The general objective of this project was to modernise, improve and integrate agricultural statistical systems at Statistics Norway related to Integrated Farm Statistics (IFS) and Gross Nutrient Budgets (GNB). The work was divided into the following objectives:

1. IFS
   1.1. Improve data processing and validation system
   1.2. Improve sample design and estimation methods
2. Identify manure management characteristics relevant to IFS and GNB
3. Provide data on crop from grazing relevant to GNB
4. Participate in Nordic seminar on agricultural statistics

The following chapters will summarise the results achieved with regard to the objectives.

Project status

1 IFS

1.1 Improve data processing and validation system

1.1.1 A-ordningen

1.1.1.1 Introduction

From 2015 onwards, a new joint reporting system called A-ordningen gathers the reporting from the employers to the Register of Employers and Employees produced by the Norwegian Labour and Welfare Administration and the Register of Wage Sums (at end of the year) reported to the Tax Administration and Statistics Norway.

A-ordningen is co-ordinated digital collection of data on employment, income and tax deductions to the Norwegian Labour and Welfare Administration, the Tax Administration and Statistics Norway. The new joint collection provides Statistics Norway with information instead of multiple sources used before. However, some other administrative registers are used to assure the data quality.

One goal of this grant is to investigate whether questions in surveys regarding labour input on agricultural holdings could be replaced, completely or partially, by information from A-ordningen in IFS 2020 and future IFSs in 2023 and 2026. A possible outcome could be less resources used on editing in Statistics Norway, less respondent burden and increased quality of the statistics.

1.1.1.2 Statistics on labour input in agriculture are important

In Norway, statistics on labour input in agriculture are important in the yearly negotiations between the two farmers' unions and the government about prices, subsidies etc. relating to the agricultural policy. Due to this, Statistics Norway conducts surveys about labour input every second or third year.

FSS has been a part of this and IFS will be in the future. Information from the census every tenth year is considered as a benchmark and thus very important.

Norway collects information on labour input in working hours and not in percentage bands of an annual working unit (AWU). Particularly holders are frequently working more than an annual working unit, which in Norway is stipulated to 1 845 working hours in agriculture.

In 2018, Norway conducted a sample survey on labour input in agriculture (SSA 2018). Except for working time off the holding for family members, the questions were equal to those used by Norway in the former FSSs and are equal to IFS 2020. Data from the survey in 2018 have been used in the examination of A-ordningen.

1.1.1.3 Working time data in A-ordningen

All employers having employees, are paying wage, pension or other payments relating to employment must report to A-ordningen. Wage per person per year not exceeding NOK 1 000 shall not be reported.

Holder and spouse of a sole holder holding shall not report their working hours and income from self-employment to A-ordningen. The same relates to partners in general partnerships. Neither data for persons that are *not* directly employed by the agricultural holding are available for the hiring holding, e.g. self-employed persons, workers hired from another farm or staffing company.

A-ordningen covers the following categories of labour force relating to an agricultural holding:

1) Manager if he/she is employed by the holding
2) Cohabitant employed by the holding (infrequent existence)
3) Family members employed by the holding
4) Non-family labour force regularly working on the holding and employed by the holding
5) Non-family labour force employed on a non-regular basis and employed by the holding
6) Holder and spouse/cohabitant on a sole holder holding; working time as employee off the holding

Working hours for labour force employed by the agricultural holding are not distributed by type work, i.e. farm work and other gainful activities related to the holding. Furthermore, experience indicates that family members to some extent work on the holding without being employed.

For each person fulfilling the conditions for being included in A-ordningen, the following information shall be reported each month:

- Identification (national identity number or identification number for foreign people working in Norway)
- Wage
- Agreed working time

There are two options for reporting working time to A-ordningen:

- Monthly salary; agreed employment fraction and hours per week relevant for 100 % employment
- Hourly wage; hours actually worked and related payment

For workers with monthly salary, this means that employers do not report directly the hours worked, but indirectly by the employment fraction and the number of working hours in 100 % employment. Wage reported is paid salary for last month, while working hours in the latest month are based on working hours for 100 % employment in one week and relevant employment fraction in the same week.

It seems easier to accurately derive real working time of employees with hourly wage from data reported to A-ordningen. However, there are some difficulties: the paid salary may include additional payment for extra hours, while the amount of extra working time is difficult to identify.

Difficulties also occur when monthly salaries are reported only once a year, e.g. at end of the growing season or end of the year, and the reference period for statistics is different from the reporting period to A-ordningen.

For employees employed by a holding, A-ordningen should mainly cover the same working time as filled in by the respondent in surveys conducted by Statistics Norway, if the reporting period is equal. The further analysis indicates to what extent working time reported to A-ordningen could replace information reported in questionnaires to Statistics Norway.

1.1.1.4 Analysis of correlation between working time from A-ordningen and the Sample Survey of Agriculture 2018

The SSA 2018 covered farm work and work in OGA related to the holding and work as employee and self-employed off the holding for holder and spouse/cohabitant in the calendar year 2017. The survey has been compared with A-ordningen in 2017.

Number of holdings in SSA 2018 linking A-ordningen 2017

About 50 per cent of the agricultural holdings did not report any information at all to A-ordningen. The number of holdings in A-ordningen having relevant information about working time and wage is even less. Holdings reporting data only on employees' pensions, vacation money, share payments or other payments, which are not directly relevant to working time in 2017, were excluded.

Table 1.1 SSA 2018; total number of agricultural holdings and number of agricultural holdings in A-ordningen.

SSA 2018	Holdings in SSA 2018 with link to A-ordningen	
	Total in A-ordningen	With relevant information
7 935	3 978	3 047
100.0 %	50.1 %	38.4 %

To find possible differences between holdings reporting to A-ordningen and holdings not reporting to A-ordningen, the total holdings were classified by different characteristics, e.g. holding type, type of farming, type of employee, standard output and total working hours reported in SSA 2018. It is not surprising that the frequency of holdings reporting to A-ordningen increased by total working hours reported in SSA 2018.

Total annual wage and working hours

As earlier mentioned, the actual working time is not reported for those with monthly salary. However, approximately working time could be derived from agreed hours, paid salary and other available information in A-ordningen. In SSA 2018, and other questionnaire-based surveys about labour force, Statistics Norway did not ask for identification of employees. Thus, comparison SSA – A-ordningen at individual level is impossible. Instead, aggregated working hours for all employees per holding are used when comparing the two sources.

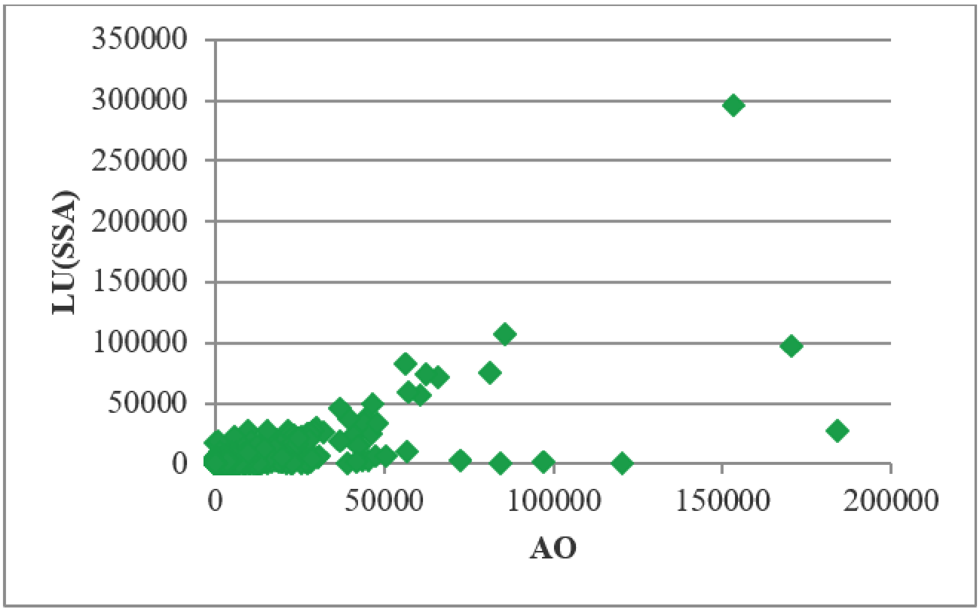
Table 1.2 Data used to produce total working hours for 2017 from A-ordningen and SSA 2018

<table>
<tr><th colspan="2">A-ordningen</th><th>SSA 2018</th></tr>
<tr><td>1.</td><td>Identify enterprises in A-ordningen which link to agricultural holdings in SSA 2018</td><td rowspan="7">Add working hours for: cohabitant, family members, non-family labour force regularly working on the holding and employed by the holding, non-family labour force employed on a non-regular basis and employed by the holding</td></tr>
<tr><td>2.</td><td>Remove holdings not having paid salary relevant to working hours (Those not specified type of work and those not having code of occupation relevant to agriculture.)</td></tr>
<tr><td>3.</td><td>Remove negative salary in January.</td></tr>
<tr><td>4.</td><td>If wage per hour is too high or too low according to working hours, then apply Nearest Neighbour imputation (NNI) based on salary and code of occupation*</td></tr>
<tr><td>5.</td><td>If wage is reported but employment fraction = 0 % in one month, then set employment fraction = 100. Working hours are to be edited.</td></tr>
<tr><td>6.</td><td>Calculating working days per month per employee: remove national holidays, sick days, leave of absence, lay-off, changes in payment and vacation for those who work more than one year. Consider date of start and date of stop of working.</td></tr>
<tr><td>7.</td><td>Calculate monthly working time per employee. If wage is considered to be too high or too low, then do editing by using NNI based on salary and code of occupation.</td></tr>
</table>

*While applying NNI here, all employees were split into groups by code of occupation and to define the nearest neighbour was used payed salary for Euclidian distance calculations in objective function.

A scatter plot (figure 1.1) shows total working hours reported in SSA 2018 (LU) and total working hours in A-ordningen (AO) for the same sub-group of holdings. The plot shows over-coverage for A-ordningen. The difference might point that there are holdings reporting much fewer working hours in SSA compared with working hours in A-ordningen.

*Figure 1.1 Scatter plot of total working hours in A-ordningen 2017 (AO) compared with total working hours in SSA (LU)**

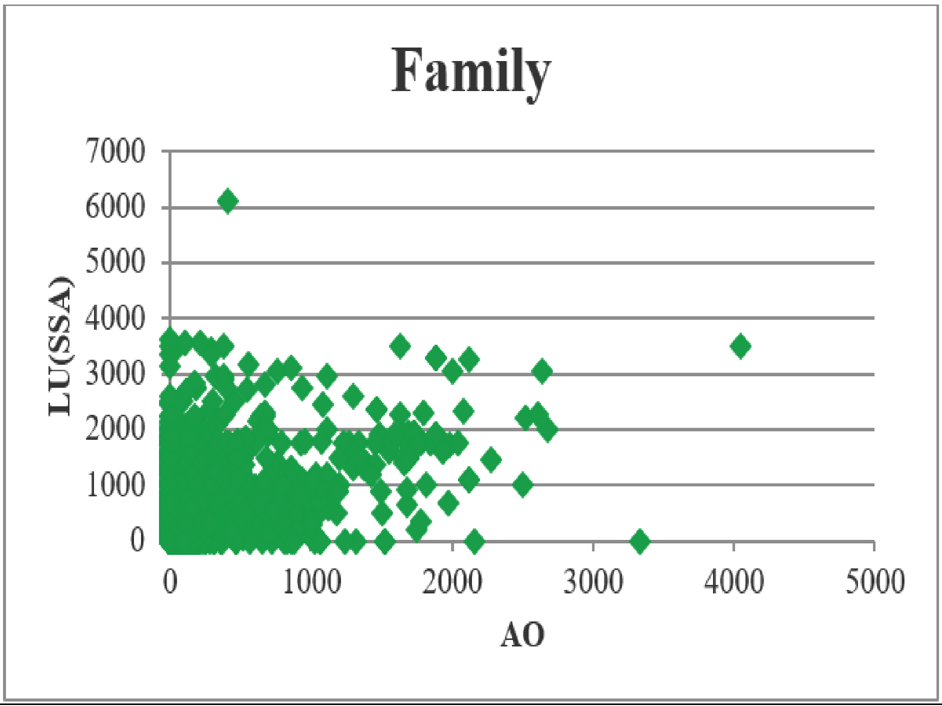


	N	Sum	Mean	Median	1st Qu.	3rd Qu	StdDev	Min	Max
AO	2821	7388810	2619.22	478	170	1887	8878.72	0	183894
LU(SSA)	2821	5848045	2073.04	650	200	1820	7793.18	0	295290
Person Correlation Coefficients		0.6984462							

*Excluding employees with irrelevant occupation.

Holdings in SSA 2018 with little work in SSA compared with wage reported in A-ordningen were removed. The difference could be caused by incorrect editing in SSA, e.g. non-response in SSA or irrelevant professions included from A-ordningen. The latter may occur when a holder operates OGA not relating to the agricultural holding.

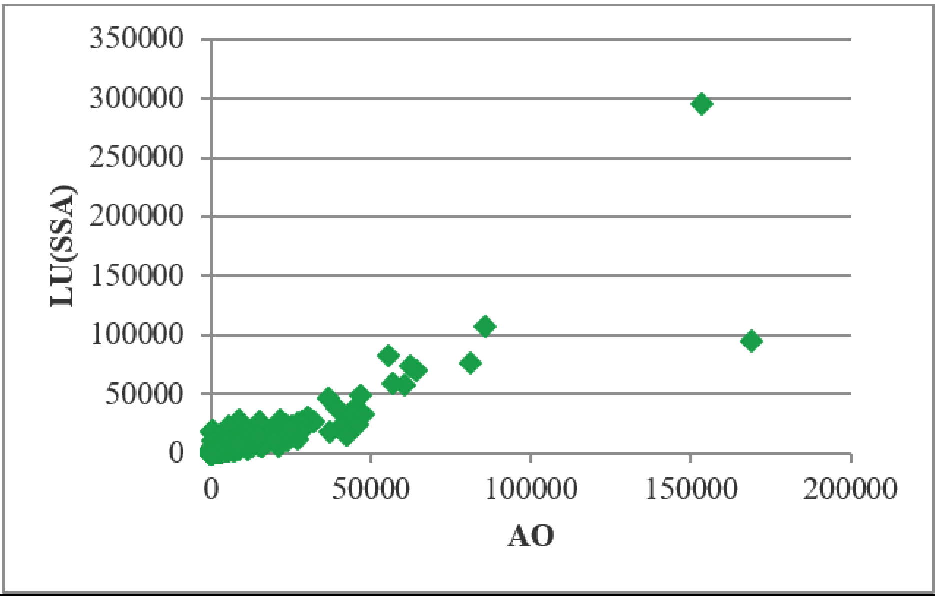
Figure 1.2 Scatter plot of total working hours for family members in A-ordningen 2017 (AO) compared with total working hours in SSA (LU)



	N	Sum	Mean	Median	1st Qu.	3rd Qu	StdDev	Min	Max
AO	2363	431941	182.79	33	0	200.5	357.44	0	4049
LU(SSA)	2363	984994	416.84	182	0	500	620.01	0	6100
Person Correlation Coefficients	0.4413072								

Figure 1.2 shows that working hours for family members in SSA 2018 is not corresponding well with the working hours in A-ordningen. A conclusion is that working time for family members on agricultural holdings still has to be collected via questionnaire.

Figure 1.3 Scatter plot of total working hours for non-family labour force employed by the holding (regular and non-regular), in A-ordningen 2017 (AO) compared with working hours in SSA (LU)



	N	Sum	Mean	Median	1st Qu.	3rd Qu	StdDev	Min	Max	W_Sum*
AO	2237	4569535	2042.71	206	0	1484	7558.29	0	169236	8226395
LU(SSA)	2237	4488617	2006.53	300	0	1700	8671.1	0	295290	8472951
Person Correlation Coefficients			0,8889629							

*Weighted sum

Working hours for employed working force other than cohabitant and family members indicate that one should separate regular and non-regular labour force. However, this division could not be done on data from A-ordningen. Instead, data from both A-ordningen and SSA 2018 were used, together with Random Forest, Support Vector Machine and Naïve Bayer's models for classification of regular and non-regular employees. The best classification models reached only about 60 % precision.

In addition, total working hours in A-ordningen could significantly deviate from working hours in SSA 2018. Some agricultural holdings with most working hours in SSA 2018, had only about half of the working hours in A-ordningen.

Total working time per year for the population U
Before concluding that data from SSA is better than data in A-ordningen, one should consider errors in SSA. Thus, it is definitely worth checking the sources at macro level. An estimate for working hours carried out by employees who are not family members was calculated:

$$\hat{t}_{andre,SSA} = \sum_{i \in s} T^{SSA}_{andre,i} w_i$$

If we replace data from SSA with data from A-ordningen, we could calculate total working hours for the same category of employees:

$$\hat{t}_{andre,SSA-A-ordn} = \sum_{i \in s} \left(T^{SSA}_{andre,i} I(i) + T^{A-ordn}_{andre,i} (1 - I(i)) \right) w_i,$$

$$where\ I(i) = \begin{cases} 1, if\ unit\ i\ not\ in\ A - ordn \\ 0, ellers \end{cases}$$

*Table 1.3 Estimated working hours in 2017 for non-family labour force employed by the holding (regular and non-regular), based on SSA 2018 and a combination of SSA 2018 and A-ordningen**

	Male	Female	Total
$\hat{t}_{andre,\mathrm{SSA}}$	9 121 208	3 316 156	12 437 364
$\hat{t}_{andre,\mathrm{SSA}-A-ordn}$	11 540 450	4 127 363	15 667 812

*Weighted by weights of SSA 2018.

Estimate of the difference between SSA and SSA and A-ordningen:

$$\frac{\hat{t}_{andre,\mathrm{SSA}} - \hat{t}_{andre,\mathrm{SSA}-A-ordn}}{\hat{t}_{andre,\mathrm{SSA}}} * 100\% = 25.97\ \%$$

Classifying the agricultural holdings by type of farming (table 4) indicates a large difference for some farm types. This indicates that SSA should not be replaced by A-ordningen.

*Table 1.4 Estimated working hours in 2017 by type of farming for non-family labour force employed by the holding (regular and non-regular), based on SSA 2018 and a combination of SSA 2018 and A-ordningen**

Type of farming	$\hat{t}_{andre,\mathrm{SSA}}$	$\hat{t}_{andre,\mathrm{SSA}-A-ordn}$	Difference
1 Cereals and oil-seeds	449 793	613 152	-163 359
2 Other field crops	420 180	500 596	-80 417
3 Horticulture and permanent crops	4 286 092	6 435 539	-2 149 448
4 Cattle – dairying	2 623 825	2 771 662	-147 837
5 Cattle – rearing and fattening	579 931	665 908	-85 977
6 Cattle – mixed	551 666	612 397	-60 730
7 Sheep	756 747	662 473	94 274
8 Various grazing livestock	397 863	487 053	-89 190
9 Granivores	1 050 668	1 233 293	-182 625
10 Mixed crop production	388 807	565 863	-177 056
11 Mixed livestock production	323 630	379 699	-56 069
12 Mixed crop and livestock production	608 162	740 177	-132 015

*Weighted by weights of SSA 2018.

Total working hours off the holding for holder and spouse/cohabitant

SSA 2018 and A-ordningen were examined for reported working hours off the holding for holder and spouse/cohabitant.

Table 1.5 Number of spouses/cohabitants who are employed off the holding in 2017, not weighted

	Yes, in A-ordningen	No, in A-ordningen	Total
Yes, in SSA 2018	1 156	203	1 359
No, in SSA 2018	1 212	1 067	2 279
Total	2 368	1 270	3 638

The percentages of yes-yes, no-no and yes-no were respectively 31.8 %, 29.3 % and 38.9 %.

There are quite many holdings with “yes” in A-ordningen and “no” in SSA 2018. This indicates that the respondents in SSA 2018 did not manage or ignore to answer this question. The latter could be caused by respondents considering work off the holding not to be relevant in an agricultural survey. In this case, A-ordningen could be used to increase the quality of the questionnaire-based survey.

Table 1.6 Number of holders who are employed off the holding in 2017, not weighted

	Yes, in A-ordningen	No, in A-ordningen	Total
Yes, in SSA 2018	1 286	563	1 849
No, in SSA 2018	578	1 211	1 789
Total	1 864	1 774	3 638

The percentages of yes-yes, no-no and yes-no were respectively 35.3 %, 33.3 % and 31.3 %.

For both spouses/cohabitants and holders, there is one weak point in the comparison of working time off the holding in SSA 2018 and A-ordningen. Working time reported in SSA 2018 includes both working time as employee and working time as self-employed person, while A-ordningen includes only working time as employee.

1.1.1.5 Conclusion and recommendation

The output of the analysis of the data sources SSA 2018 and A-ordningen indicates that questions relating to working hours in SSA 2018 and in similar future surveys (IFS) could not be replaced by information from A-ordningen. Nevertheless, information from A-ordningen will be useful in the editing of questionnaire-based census/survey data on labour input.

There is an option to present calculated working hours or paid wage per employee to the respondent (employer) when he/she is filling in the questionnaire. Despite calculated working hours/wages are not correct for every individual in A-ordningen, for a part of the population is there coherence at macro level. At the same time, it is important to enable the respondent to edit suggested working hours because of possible under-coverage in A-ordningen or calculated hours may differ from the real. In any case, the suggested data based on A-ordningen should ease the respondent burden.

Figure 1.4 Example of question with recommended data

Give working hours for each employee on the agricultural holding divided by category of work and category of labour force:

				In agriculture	In forestry	In other OGA	Family-member	Regular employee	Non-regular employee
Name 1	Date of birth	Gender	Number*	|_____|	|_____|	|_____|									
Name 1	Date of birth	Gender	Number*	|_____|	|_____|	|_____|									

If some employees are not included in A-ordningen; for each individual add gender, working hours by category and category of labour force:

	Gender	In agriculture	In forestry	In other OGA	Family-member	Regular employee	Non-regular employee
Individual 1	|______|	|_____|	|_____|	|_____|									
……							

*Number can be provided salary or total working hours for each employee in A-ordningen.

By implementing suggested data in the questionnaire and after finish of the IFS 2020, Statistics Norway could test machine learning (ML) to classify employed labour force (regular and non-regular) and adjust predicted working hours closer to actual working hours. This option exists because both the calculated and the actual data per employee is known. As basis for machine learning, data from A-ordningen, other relevant registers and former surveys could be used. When machine learning is implemented, less holdings could be surveyed. Then could be asked; 1) holdings included in A-ordningen to update the ML model and 2) some holdings not included in A-ordningen but included in the population which are representative for this part of population. At last RENI (Restricted Nearest Neighbour Imputation) could be used for the remaining part of the population. Then, the population will consist of three parts: A-ordningen, survey and predicted data. Data from the annual tax return could be used for holdings not included in A-ordningen.

1.1.1.6 Use of a-ordningen in IFS2020

Based on test of A-ordningen for 2019, we can expect that approximately 25 per cent of the holdings have reliable information of their employees in 2020. Therefore, the data collection for IFS 2020 will be done in two phases. The main bulk of the holdings, i.e. holdings without information on employees in A-ordningen, will receive a questionnaire without any pre-fill of employees in the last week of September. The remaining holdings, i.e. holdings with information on employees in A-ordningen, will receive the questionnaire by the end of October. This is caused by the fact that information on working hours will refer to the period 1 October 2019 – 30 September 2020. The data from A-ordningen for September 2020 are not available before 20 October.

Based on the examination, the use A-ordningen in IFS 2020 will be as follows for those holdings with relevant information:

- A suggestion of working hours for each employed person on the agricultural holding, which is either the reported number of working hours to A-ordningen if available, or an estimated number of working hours based on salary and employment code reported to A-ordningen if working hours are not reported:
    - The suggested working hours may be corrected or confirmed by the respondent and should be divided on the categories; a) agriculture, b) forestry, and c) other OGA
    - Category of labour force for each person should be recorded; a) family member, b) regular employee or c) non-regular employee.

    - If one or more employees are missing in the list, there is an option to add these to the list.
    - Holdings having up to 30 employees are pre-printed in the electronical questionnaire for IFS 2020. Due to technical limitation in the electronical questionnaire, holdings having more than 30 employees will receive an attached Excel-sheet with a list containing all employees.
- Information from A-ordningen will be used in the edit and imputation process.
    - This information will be used to control and validate the recorded number of working hours as employee off the holding for holder and spouse/cohabitant on sole holder holdings, working hours on the holding executed by cohabitant, family members, regular employees and non-regular employees. Furthermore, working hours on and off the holding executed by employed manager. In addition, there will be validation between the suggested working hours and the corrected working hours for each employee

The questionnaire has been tested by several respondents in general. And later, the suggested working hours for employees were tested by 10 respondents specific for this purpose. The conclusion from the latter was that the suggested working hours was reliable, but some of the respondents made corrections in working hours for some of the employees.

Due to this step of modernisation of the agricultural statistics, we expect that the respondent burden will be reduced for these holdings, as well as increased quality in reported information on labour force. We also expect a rise in quality of the labour force data because A-ordningen should be used in the editing of these variables.

1.1.2 Automated editing and imputation (E&I)

1.1.2.1 Background

Statistics Norway took the first step to develop an application for automated editing and imputation in FSS 2010. The idea was to reduce manual corrections and establish an application which could be used not only for the agricultural census 2010, but for future agricultural surveys. However, the developed application was used only for the census. This was caused by lack of flexibility in the SAS data program and the existence of human resources which have long experience in manual editing routines.

In agricultural surveys after 2010 data were checked and edited by means of a general data editing software application called DYNAREV, developed in Statistics Norway. Due to missing or incorrect data items, follow-up contacts to the respondents by phone or e-mail were necessary. Manual imputation was done, based on other information in the questionnaire, information from other sources and average values for similar holdings in the municipality or county.

After entry, the data were subject to comprehensive computerized controls consisting of completeness checks, valid value checks (included minimum and maximum values), range checks, relational checks and arithmetic checks.

The staff at Statistics Norway had online access to the Business register at Statistics Norway, the latest version of the applications for governmental production subsidies and the latest version of tax return data at Statistics Norway. Conformity checks at holding level between the census and applications for governmental production subsidies were important.

Last checks were done on aggregated results. Aggregates of various characteristics were compared with results from previous surveys.

While working with modernisation of agriculture statistics, a need and wish in getting automated E&I (E&I 2020) has appeared again and should be used not only for the census in 2020 but also for further surveys. In addition, new data sources have appeared, which can be used in the E&I process. That is why there is a need for modernisation of E&I 2010.

To improve understanding should be made a remark: further will be used editing and imputation terminology from https://ec.europa.eu/eurostat/documents/64157/4373903/05-Handbook-on-data-quality-assessment-methods-and-tools.pdf/c8bbb146-4d59-4a69-b7c4-218c43952214:

- "*Data editing* is the application of checks that identify missing, invalid or inconsistent entries or that point to data records that are potentially in error"
- "*Imputation* is the process used to assign replacement values for missing, invalid or inconsistent data that have failed edits"

1.1.2.2 Prepare a revised and extended method

Types of errors

All errors which are edited and imputed during the E&I process will be divided into four groups:

1. *Illegal values*: Check if the values collected are legal. (Answers correspond to information from registers, values are laying within reasonable limits). If not – impute legal values.
2. *Low-probability values*: Checks on collected values that are already verified whether they are probable. (It is probable that an agricultural holding can spend much more working hours than other farms with approximately the same agricultural area and the same type of farming).
3. *Partly missing data* (item non-response): An agricultural holding has not answered one or more questions.
4. *Totally missing data* (unit non-response): The holding did not respond at all. Then one must impute answers for all questions.

Most of the variables in the agricultural surveys are numerical. For all numerical variables are applied automated editing and automated imputation (if a holding is not sent to manual review). Among categorical variables, those containing gender information and 1/0 values were validated and corrected in automated E&I.

1.1.2.3 E&I process

The E&I process consists of two Generic Statistical Business Process Model (GSBPM) sub-processes *5.3 'Review and validate'* and *5.4 'Edit and impute'*. Improvements in automated E&I 2010 are made in both sub-processes and the new automated E&I will be further referred to as E&I 2020.

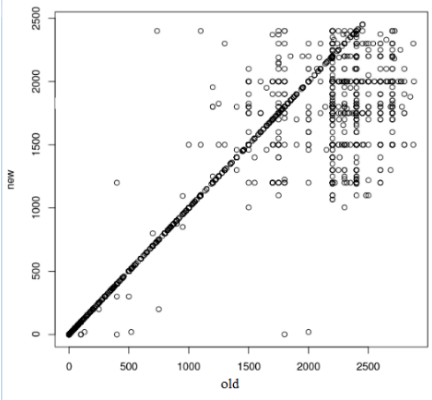
E&I in FSS 2010 could be split in 3 processes: 1) Editing and imputation of illegal values in DYNAREV, 2) Editing and imputation of low-probability and partly missing data in SAS and 3) Imputation of data for totally missing units in SAS.

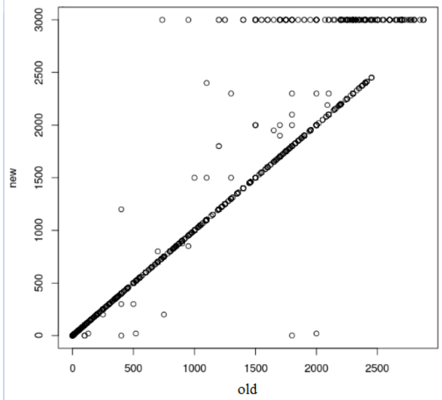
Methods which were used in part 2 of *E&I 2010* for "Review and validation" were validation based on expert opinion, maximum valid value, coherence with other sources (register data), coherence within respondent answers, outlier detection by combining two correlated respondent answers. In "Edit and Impute" sub-process, data were imputed based on frequency distribution, expert opinion, maximum valid value and nearest neighbour (NNI). For totally missing values, there was used restricted nearest neighbour (RENI).

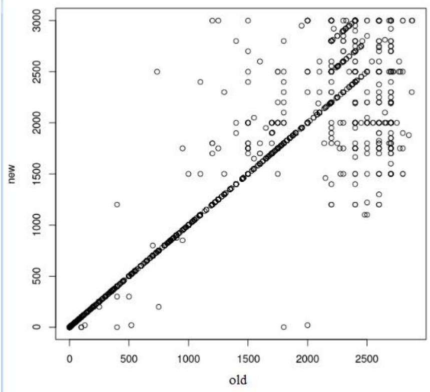
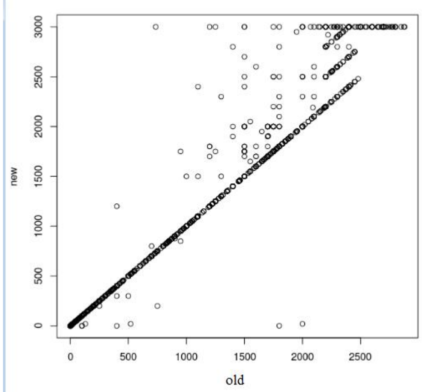
In *E&I 2020*, the same three processes exist and is added a new part with *control* of potential unit non-response. This part is aiming to catch those respondents who were not willing to provide reliable answers: answering either too big values which are not possible in reality, giving only 0/1 answers or giving answers with duplicated digits. The respondents who did not go through the validation will be sent to manual check. As a result, these holdings 1) Shall get the questionnaire in return, if it is confirmed to be a unit non-response and the holding has a significant impact on the statistical result; 2) Will be manually corrected or 3) Will be marked as unit non-response and will be imputed in automated E&I part.

Some methods used in E&I 2010 will be reused in E&I 2020. However, part 2, defined above, is developed in R and includes modifications compared to the methods used in 2010. Some validation rules instead of controlling against a max valid "hard" value, were modified to controlling a quantile-dependent number (Median + (3^{rd} Q – 1^{st} Q) * k), which is flexible. In "Edit and Impute" for more imputations were applied NNI, instead of imputing a max valid value. Quantile validation method was compared with Random forest validation and together with NNI correction provided better results. (See an example in figure 1.5).

*Figure 1.5 Comparing Random Forest and Quantile validation methods together with two imputation methods (NNI and Max valid value imputation). Example is provided for working time of farm holder in agriculture**

<table>
<tr><td></td><td>NNI</td><td>Impute max valid value</td></tr>
<tr><td>RF (Random Forest)</td><td>

old</td><td>
</td></tr>
<tr><td>Quantile limit

k was chosen to best suit to max value in LU18. For HJHV the value was 3005, k = 1.</td><td>
</td><td>
</td></tr>
</table>

*In the table Old – manual editing for Agriculture survey 2018, New – editing of the same data by using automated E&I 2020.

For the same example results on macro level can be found in table 1.7. Both imputation methods for the case when there were used quantiles providing better results than Random Forest and one of them should be used for E&I. Ignoring the fact that imputing quantile max value has lower mean(err^2) than the rest of performance measures, all other performance measures (among which is mean error, which is more important for bias) are better for NNI case and thus is preferable and will be used in E&I.

Table 1.7 Example from figure 1.5. Comparing different validation and imputation methods in E&I 2020 (New) and data without any changes (Raw) with manual corrections made in 2018 (Old).

	New				Raw	Old
	RF, NNI	RF, max value	Quantile, NNI	Quantile, max value		
Mean(err)	-348.8608	557.7858	185.9189	349.0095	529.9031	
Mean(err^2)	1162.88	1455.032	1041.348	980.753	1690.716	
Max	28373	42559.5	36884.9	42559.5	53608.2	34047.6
Delta	-0.05845905	0.09346888	0.03115467	0.05848398	0.0887965	

*In the table $\boldsymbol{err} = \boldsymbol{w_i} * (\boldsymbol{y_i} - \boldsymbol{y_{i,old}})$, $\boldsymbol{max} = \boldsymbol{max}(\boldsymbol{y_i} * \boldsymbol{w_i})$, $\boldsymbol{delta} = (\boldsymbol{sum}(\boldsymbol{y_i} * \boldsymbol{w_i}) - \boldsymbol{sum}(\boldsymbol{y_{i,old}} * \boldsymbol{w_i}))/\boldsymbol{sum}(\boldsymbol{y_{i,old}} * \boldsymbol{w_i})$. Where $\boldsymbol{y_i}$ – time values for raw data or new data, $\boldsymbol{y_{i,old}}$ – time values for old, $\boldsymbol{w_i}$ – weights, which were used in 2018.

In part 2, outliers in variables: number of temporary workers and their working time in E&I 2020 are defined by combining these two variables. Also, there were included edits (difference of logarithms) controlling mistypes in digits (17500 instead of 1750).

One more important modification of part 2: use of an additional data source, A-ordningen. By using this new data source in the survey, we are already aiming to get better results. However, there still might be room for editing. A-ordningen is used for controlling working time off the holding as employee for cohabitant/spouse and farm holder and working time on the holding for family members and non-family workers. Edits are checking if working time answered in survey (t) had a significant difference with salary reported in AO (L^{AO}). If this takes place, then corrections are made correspondent with information in A-ordningen (optimal adjustment or NNI):

$$if L^{AO} \neq 0 \; \& \frac{t}{L^{AO}} \leq 0.0015 \Rightarrow t = t^{AO}$$

NNI, RENI and optimal adjustment [1] will be added to R programs that will be available for the agriculture division and other divisions in Statistics Norway. The data program was developed in the way which gives one a possibility to easily modify it for use in coming surveys.

1.1.2.4 Test extended method against survey 2018

In table 1.8 is provided comparison of raw data, data after work of E&I 2020 and corrections made in survey 2018, which are the numbers used while publishing statistics. To get numbers on population levels were used the same weights which were used while publishing statistics. In each field there is provided 5 numbers: a. before E&I (raw), b. after E&I2020 (new), c. after E&I in survey 2018(old) and differences:

$$Delta\ NR = \frac{new \div raw}{old} * 100\%$$

$$Delta\ NO = \frac{new \div old}{old} * 100\%$$

Similarly, one can find comparison for number of temporary workers (table 1.9).

Table 1.8 Comparison of raw data, data after work of E&I 2020 and corrections made in survey 2018. Variable: working time.

	Farm holder	Spouse/Cohabitant	Family members	Permanent workers	Temporary workers	Partners
Agriculture	Raw: 44443299.50 Old: 40825539.91 New: 41026557.78 NR: -7.69 NO: 0.49	Raw: 11023715.09 Old: 10932244.88 New: 10867680.36 NR: -1.42 NO: -0.59	Raw: 9148167.86 Old: 8795777.87 New: 8827915.27 NR: -3.50 NO: 0.37	Raw: 6381984.49 Old: 6079643.45 New: 5950738.53 NR: -6.76 NO: -2.12	Raw: 5464857.13 Old: 5280280.62 New: 5332936.67 NR: -2.44 NO: 0.997	Raw: 7770.79 New: 7433.54 Old: 7769.71 NR: -0.014 NO:4.327
Forestry	Raw: 1009457.64 Old: 1000094.43 New: 1001277.30 NR: -0.81 NO: 0.12	Raw: 142668.78 Old: 151451.14 New: 139713.14 NR: -2.07 NO: -7.75	Raw: 299172.57 Old: 298882.53 New: 306080.45 NR: 2.31 NO: 2.41	Raw: 61595.29 Old: 61580.40 New: 59534.28 NR: -3.34 NO: -3.32	Raw: 78405.92 Old: 79684.34 New: 80478.00 NR: 2.64 NO: 0.996	Raw: 2395131.62 New: 2354458.73 Old: 2337807.57 NR: -2.452 NO: -0.712
OGA	Raw: 3130555.78 Old: 3019476.00 New: 3132242.22 NR: 0.05 NO: 3.73	Raw: 934264.88 Old: 956031.39 New: 923891.47 NR: -1.11 NO: -3.36	Raw: 1220601.76 Old: 1091833.80 New: 1190577.97 NR: -2.46 NO: 9.04	Raw: 574702.16 Old: 551570.10 New: 567028.21 NR: -1.34 NO: 2.80	Raw: 200305.87 Old: 199984.00 New: 203998.16 NR: 1.84 NO: 2.01	Raw: 181825.01 New: 178422.34 Old: 177382.31 RN: -2.505 NO: -0.586

Table 1.9 Comparison of raw data, data after work of E&I 2020 and corrections made in survey 2018. Variable: Number of temporary workers

Number of temporary workers, men	Number of temporary workers, women
Raw: 262725.9029 Old: 22469.6034 New: 21731.6364 NR: -91.7283997656403 NO: -3.28429027812747	Raw: 73335.0815 Old: 9988.3877 New: 9791.0524 NR: -86.6488831815098 NO: -1.97564718077572

From tables 1.9 and 1.10 one can notice that E&I 2020 (application just of part 2) is quite close to corrections made in agriculture survey 2018: NO value for all values is less than 5 % excluding group OGA and family, which has less than 10% difference. Taking into consideration that total in old data could have difference from real number, we cannot say that the difference is big.

Comparison for editing of variables of OGA has shown a big difference (around 80 %) only for the variables 'OGA not given in the specific list; specify (i.e. a free text field to fill in the activity)' and 'Does not have OGA'. It was caused by need of manual correction of the categorical values specified in 'OGA not given in the specific list; specify'. By applying the manual correction there will be no difference.

1.1.2.5 Assess quality characteristics and efficiency related to resource use and quality

Data programming in R developed to be used in both Agriculture Census and sample surveys, expects to reduce manual correction. It is expected that time for preparation *before* E&I will be increased, while it is expected that it will significantly reduce working hours used *on* the editing process. According to the expert opinion; in an average year it is used 1 860 hours for the FSS survey process, of which in average 50 % is the editing process. It is expected to be reduced to the half.

Since there are several quality characteristics which are defined by Eurostat, we will refer to each of them. Here will be compared E&I which were made in 2018 and the new offered editing.

Relevance

The output of published statistics will not be influenced by change of E&I method. I. e. the characteristics will not be improved using new method.

Accuracy and reliability

Measurement error is mainly occurring because of respondents. As a rule, a respondent tends to provide higher values than real. To avoid measurement errors, the questionnaire uses the most common formulations and integrated controls. Included in the modernization, for some respondents there is included information from an administrative source in the questionnaire. Further, all responses go to validation. If values did not go through validation, they are sent to correction or in some infrequent cases back to respondent.

In GSBPM 'Edit and Impute' stage, it is advised to look at un-weighted and weighted imputation rates to "measure the relative amount of imputed values and the relative influence on the final estimates from the imputation procedures" correspondingly. Further are provided these rates for edits (unit non-response part was not included).

Table 1.10 Unweighted rate, statistics for 20 main variables

Old (E&I 2018)					
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.01327	0.05306	0.92200	2.14447	2.28841	19.12974
New (E&I 2020)					
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.0398	0.1924	1.1542	2.7375	3.1673	15.6010

Table 1.11 Weighted rate, statistics for 20 main variables

Old (E&I 2018)					
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.000	0.131	4.464	7.428	10.013	30.731
New (E&I2020)					
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.02243	3.01274	5.49107	7.33324	10.50974	19.17403

In tables, one can see that the new process tends to define more errors, while average number of caught errors and relative influence are close. At the same moment old editing could have a higher influence on results.

Further were compared percent of different values after old and new E&I from total number of units (Table 1.12). Here we are not aiming to say that one E&I is worse than another but identifying if there are variables with a big difference in editing process. As one can see from table 1.12 in most of the cases outputs got the same numbers, except one variable with 23 % in difference. This variable (working time of farm holder in agriculture) also got max weighted and unweighted rate in table 1.10 and 1.11. Since relative influence on result value of the variable after old E&I was higher and both old and new processes achieved close values (see table 1.8) it is concluded that new E&I performed better.

Table 1.12 Percent of different values after old and new E&I, statistics for 20 main variables

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.05306	0.19568	1.10772	3.20509	3.16397	22.71160

Accuracy: Lack of accuracy is a metric including variance and bias coming from sampling and non-sampling errors. As far as sampling part is concerned it is remaining the same in both old and new E&I. Non-sampling error is in interest here, which are difficult to assess. To calculate the bias there is a need to know real values, which are not available.

To conclude; if two total values for variable x in a defined domain of interest (county, type of farm activity) h are close to each other there was conducted Hidiroglou & Berthelot method with C = 4, U = 0.1, A = 0.05. Some total values were classified as potential outlier according Hidiroglou & Berthelot method. Then these values were further analysed and was concluded if they are real outliers. If this was taking a place, there were applied steps to test and improve new E&I. However, close to equal values in old and new E&I were not achieved for all variables in domains. Since old E&I still can have difference from real numbers, it was concluded that values did not have a big difference.

At the same time instead of calculations a bias in each old (t1) and new (t2) resulting sets, we can look at difference t2 – t1 = z in each domain of interest to say about how accuracy has changed. For z value was calculated total and standard deviation in each domain. Those domains which have a high standard deviation, are leading to higher MSE(z), hence lower accuracy of differences. Follow the logic that the bias of the difference is equal to difference in the biases:
E((t2 – t1) – (T2 – T1)) = E(T1 – t1) – E(T2 – t2) = [Since in our case T1 = T2] = E(t2-t1)
Further we had a look at bias ratio of difference:

$$\gamma = |E(z)|/SD(z)$$

which while having small values is showing that bias is small relative to the standard error in MSE(z). In https://www.jstor.org/stable/pdf/2281815.pdf?refreqid=excelsior%3Ae430dbf1d25a10d0e349ae61bb46982e it is advised to use as rule of thumb, saying that $\gamma \leq 0.2$ gives that bias does not have a significant impact on accuracy. If both SD and γ were low, then z has a high accuracy meaning that old and new methods introduced the same results.

One can find an example for county domains in table 1.13. Provided variable working time of farm holder in agriculture. In table 1.13 we see that γ was lower 0.2 in all columns, while the highest SD was in counties, 5, 11, 50.

Reliability: In both methods it is possible to find how many values have been corrected. However, when there is involved manual correction, it is difficult to find documented defined rules on why and which value was corrected, while automated correction means usage of defined rules. That is why reliability should be increased with usage of new E&I.

In the new E&I, reasons for edits and imputations will be documented. Thus, metadata is better than in old one. It is easier to explain, control and modify the new E&I. Thus, if it will be detected a problem it easy to find what should be improved.

Timeliness and punctuality

Timelines: The new E&I method will provide possibility to run E&I of data faster. Thus, the goal is to publish statistics earlier.

Punctuality: By e.g., using data from A-ordningen, the quality of statistics could improve by change of E&I method.

Accessibility and clarity

These characteristics concern more the way the final output is presented rather than the processed which were obtained to get the statistics. At the same moment, when a statistics user will refer to documentation of obtained processes, there is more clarity in automated methods rather than in manual correction.

Coherence and comparability

Coherence: Both new and old E&I involve corrections on the information available from registers. However, methodologically justified correction contributes a better coherence across domains than manual since the first is less human dependent. Additionally, there was included a new data source in the new E&I. Hence, coherence of surveyed data with the source was increased with the new E&I.

Comparability: Again, old E&I involves more manual corrections than new E&I. That is why use of the latest will improve comparability over time and geographically, by making it is easier to control value error.

Table 1.13 Total and standard deviation for difference(z) between results in old(t1) and new(t2) E&I by county. Variable: Working time of farm holder in agriculture.

County	Z	SD * N_h	γ
1 Østfold	-47893.82	645358.01	0.07
2 Akershus	-31168.22	532730.26	0.06
3 Oslo	0	0	0
4 Hedmark	37672.49	1019333.22	0.04
5 Oppland	2550.4	1456136.9	0.0
6 Buskerud	-52013.05	589535.49	0.09
7 Vestfold	-56855.45	370935.86	0.15
8 Telemark	-53691.55	441627.14	0.12
9 Aust-Agder	-5004.79	157287.66	0.03
10 Vest-Agder	-27212.63	406911.14	0.07
11 Rogaland	155643.25	1434326.69	0.11
12 Hordaland	30274.81	767474.18	0.04
14 Sogn og Fjordane	12209.91	817425.89	0.01
15 Møre og Romsdal	-28724.51	761926.69	0.04
18 Nordland	-50695.20	533870.54	0.09
19 Troms	-45859.21	280590.59	0.16
20 Finnmark	20812.27	113864.87	0.18
50 Trøndelag	-61062.58	2024076.98	0.03

*In table Z = sum(t1) – sum(t2) in domain h, N_h – size of domain, SD – standard deviation of z, γ – bias ratio. For all were used weights.

1.1.2.6 Conclusion

The GSBPM processes 5.3 ‘Review and validate’ and 5.4 ‘Edit and impute’ are really important and time-consuming processes in all statistical surveys. These processes may have a significant impact on the results of a survey. In Statistics Norway, these processes have traditionally been a combination of computer-based validations and controls where it is up to the persons working with a survey to decide whether a given number should be edited or not, as well as what number to replace. In that way, edits could be subjective.

Modernisations of the editing process by more use of automated editing and imputation will have at least two major advantages. First, the time-consuming editing will be far more efficient with less use of human resources. The effect on this process will also lead to better timeliness and punctuality of the results. Second, the use of more edit and impute by a given set of computer-based rules will lead to a process which is more coherent and punctual since it is better documented and less dependent on human resources.

The automated edits and imputations have been accurately tested on the Sample Survey of Agriculture 2018 module on *Labour force & and other gainful activities* with an acceptable outcome. Further, the same principles have been used on the Fertiliser Survey 2018. The automated edit and imputation routine will be used in the IFS 2020, on both modules *Animal housing and manure management* and *Labour force and other gainful activities*.

1.2 Improve sample design and estimation methods

The modernization of agricultural statistics shall result in both more efficient compiled statistics and an improved quality of the figures disseminated. In this chapter we are looking on 6 different aspects of processing the agricultural statistics based on a sample from the whole population of agricultural holdings.

1.2.1 Stratification of agricultural holdings by size, localization and type of farming

In agricultural statistics, the agricultural holding (farm) is the basic unit. The agricultural holdings in Norway vary considerably relating to size, localization and type of farming. Variation between groups formed by these three characterizations is often much higher than variation among farms inside each group. We call these groups for strata when designing a sample plan to select a sample of agricultural holdings to an agricultural survey. Before we can select a sample, we divide the population of agricultural holdings into strata characterized by these three characteristics:

- *3 size groups*: The agricultural holdings are classified either as small, medium or large holding by a size measure constructed by combining size of the agricultural area and size of the livestock. This common size-measure is based on *calculated manure units (GDE)*, see annex 1 for number of animals per calculated manure unit. Furthermore, one calculated manure unit is equal to 0.4 hectare of arable land[1].
- *10 counties*: We distinguish between which county the agricultural holdings are located in. As from 2020 Norway consists of 11 counties, see annex 2. Related to agriculture, Oslo is merged with Viken county.
- *10 types of farming*: The characteristic of type of farming is based on the EU typology, see annex 3.

If we combine these characteristics; 3 size groups, 10 counties and 10 types of farming, we partition the entire population into 300 strata. Statistics Norway, now planning the IFS 2020, has estimated the number of agricultural holdings to be about 39 000. If we partition the entire population into 300 strata, the average number of units in each stratum will be 130.

1.2.2 Sample design

We assume from previous chapter that the agricultural population consisting of 39 000 farms is partitioned into 300 strata. We propose that the sample plan of the agricultural surveys after IFS 2020 and onwards to the next census in 2030 is based on the following:

- Firstly, we remove a number of agricultural holdings that are 1) general partnership with dairy cows, 2) holdings operating specialist horticultural production in greenhouse and 3) holdings operating livestock husbandry without utilised agricultural area.
- For the remaining population of agricultural holdings, a sample plan is drawn up based on the 300 strata, subdivision is described above.
- The total size of the sample is set equal to 6 000 agricultural holdings.

The sample design for the remaining population (bullet point 2 above) is based on the following three steps:

- Firstly, in each stratum is allocated 2 farms, i. e. 200 farms are removed from each of the three size groups.

---

[1]According to Regulation concerning organic fertilisers, holdings must have enough available utilised agricultural area for spreading of manure. The basic rule says that the minimum area has to be 0.4 hectare of arable land per calculated manure unit.

- Then, each of the three size groups are allocated proportional by the product of the average value of the farms' calculated manure units defined above and the number of farms minus 200, see the formula below.
- Finally, the number of agricultural holdings determined in the previous point for each size group are distributed to the 100 combinations of counties and types of farming proportional to the population size minus 2. At last, the 2 farms firstly removed are added to the size in each of the 100 strata for the small, medium and large size groups, respectively.

We outline the three bullet points in mathematical formulas:

The notation of total population, total sample of agricultural holdings and number of calculated manure units (x) for agricultural holdings:

$$(1)\quad N = 39\,000 \quad , \quad n = 6\,000 \quad and \quad x_i \, , i = 1,2,3,...,N$$

The three special groups, general partnerships with dairy cows (Gen_partner), specialist horticultural production in greenhouse (Greenhouse) and specialist livestock husbandry without utilised agricultural area (Without_areal) are deducted from sample (n) and the population (N):

$$(2)\quad \begin{cases} n^{Draw} = n - n_{Gen\_partner} - n_{Greenhouse} - n_{Without\_areal} \\ N^{Draw} = N - N_{Gen\_partner} - N_{Greenhouse} - N_{Without\_areal} \end{cases}$$

After subtracting 200 farms from each size group, the number in the sample is calculated:

$$(3)\quad n_S^{Draw} - 200 = \frac{(N_S^{Draw} - 200)\cdot \overline{x}_S^{GDE}}{\sum_{S^*=1}^{3} (N_{S^*}^{Draw} - 200)\cdot \overline{x}_{S^*}^{GDE}} \cdot (n^{Draw} - 600) \quad , S = 1\,(small), 2\,(middle), 3\,(l\arg e)$$

The sample is distributed proportionally to the population number in each combination within the three size groups:

$$(4)\quad n_{S,C,F}^{Draw} - 2 = \frac{(N_{S,C,F}^{Draw} - 2)\cdot (n_S^{Draw} - 200)}{\sum_{C^*=1}^{10} \sum_{F^*=1}^{10} (N_{S,C^*,F^*}^{Draw} - 2)} \quad , S = 1,2,3\,; C = 1,2,3,...,10 \; og \; F = 1,2,3,...,10$$

Then we have calculated for each stratum the number of farms to be drawn ($n_{S,C,F}^{Draw}$) for each combination of size (S), county (C) and type of farming (F). The sum of the sample size for each stratum then gives the total sample size (n^{Draw}) for main part of the sample that we can insert in the first row of formula (2) to find the sample size of the total sample.

The calculation of the number in the sample is prepared with the aim of minimizing the variance of the estimate of the total population. The reason for adding the two farms in each stratum is to improve the quality of the totals for each county and each type of farming and any other totals relevant to calculate. By keeping 10 per cent of the sample out of the distribution of agricultural holdings into the strata, the minimization of the uncertainty of the total is very little weakened compared to the gain achieved by improving the quality of the calculation of the totals for counties and types of farming.

1.2.3 Selection of samples

The sample plan in the previous chapter specifies how many farms shall be selected in each stratum. This means, the major selection of the sample beside the three special types (Gen_partner, Greenhouse and Without_areal) is conducted as simple random sampling in each of the three hundred strata (stratified simple random sampling).

We ensure that the sample is representative with respect to other factors as well, namely landscape (lowland areas or other areas) and standard output:

- If an agricultural holding is in lowland areas or other areas respectively. This is due to variation in length of growing season and effectiveness of farming caused by the landscape, e.g. soil quality, slope, parcel size, distance between parcels etc.
- The holding's standard output

We do this by sorting the population within each stratum by these two factors. In addition, we reduce the sample size by 2 in each stratum before we draw the sample. The purpose of this approach is to compare the sample after drawing with the entire population and draw the last two farms within each stratum so that the distribution of farms in the whole sample is improved regarding the three characteristics' distribution in the population.

The plan is to develop a process for drawing the sample to the yearly agricultural sample surveys, so it can be implemented in a general system for conducting all survey samples in Statistics Norway for business statistics.

1.2.4 Editing and imputation of data

Editing and imputation is handled in chapter 1.1.2.

1.2.5 Estimation and calibration

Based on the sample design prepared in chapter 2.2.2, weights can be calculated equal to the inverse probability of selecting a farm in each stratum:

$$(5)\quad \begin{cases} w_{i,S,C,F} = \dfrac{N_{S,C,F}^{Draw}}{n_{S,C,F}^{Draw}}, i \in s_{S,C,F} \quad , S = 1,2,3, C = 1,2,3,\dots,10 \; og \; F = 1,2,3,\dots,10 \\ w_{i,G} = \dfrac{N_G}{n_G}, i \in s_G \quad , G = Gen\_partner, Greenhouset, Without\_areal \end{cases}$$

Here are $s_{S,C,F}$ and s_G the samples in respectively strata (S,C,F) and (G = General partnership with dairy cows, specialist horticulture in greenhouse and livestock husbandry without utilised agricultural area, respectively). This means, if we sum up agricultural holdings in a stratum with these weights, we will estimate the total for a statistical variable Y:

$$(6)\quad \begin{cases} \hat{T}_{Y,S,C,F} = \sum_{i \in s_{S,C,F}} w_{i,S,C,F} \cdot Y_{i,S,C,F} \; and \; T_{Y,S,C,F} = \sum_{i \in U_{S,C,F}} Y_{i,S,C,F} \quad , S = 1,2,3, C = 1,2,3,\dots,10 \; and \; F = 1,2,3,\dots,10 \\ \hat{T}_{Y,G} = \sum_{i \in s_G} w_{i,G} \cdot Y_{i,G} \; and \; T_{Y,G} = \sum_{i \in U_G} Y_{i,G} \quad , G = Gen\_partner, Greenhouset, Without\_areal \end{cases}$$

Here are $U_{S,C,F}$ *and* U_G the population for respectively strata (S,C,F) and (G = General partnership with dairy cows, specialist horticulture in greenhouse and livestock husbandry without utilised agricultural area, respectively).

The statistics that is going to be compiled should preferably coincide with other conditions as well. This can be achieved by adjusting the weights using methods from regression analysis used in the Fertiliser Survey 2018 (Reports 2020/9) and is called calibration. Reports 2020/9 describes how these calibrated weights are calculated. For example, we can impose certain conditions so that the weights are coinciding with marginal totals for farm types and counties for data derived from the production subsidy system.

1.2.6 Coefficient of variation

In the new Statistics Act, official statistics are defined by meeting certain quality requirements. This is explicitly defined in paragraph 5 of the act:

- Official statistics should be developed, compiled and disseminated in a professional independent, impartial, objective, reliable and cost-effective manner.
- Development, compilation and dissemination of official statistics shall be based on uniform standards and harmonized methods. The statistics shall be relevant, accurate, timely, accessible, clear, comparable and coherent.

Among other factors, there is a strong demand on the statistics divisions to work out coefficients of variation for all statistics where it makes sense. Thus, for agricultural surveys after 2020 coefficients of variation shall be calculated for all published figures.

The Division for methods shall prepare a tool that all statistics divisions can use to calculate coefficients of variation for all sample-based statistics published from 2021 on Statistics Norway's website.

1.2.7 Preliminary figures for the Agricultural Census 2020

The Division for housing, property, spatial and agricultural statistics in co-operation with the Division for Methods are planning to prepare and disseminate preliminary figures for the IFS 2020. A sub-group of agricultural holdings is drawn and will be prioritized in the editing process. The sub-group will be drawn based on the sample design described in chapter 2.2.2. Experience from the agricultural census will be used to adjust both the sample design and the various work processes during the compilation of the statistics for preliminary figures such that the procedure for the future agricultural surveys will be improved compared to the standing now.

2 Identify manure management characteristics relevant to IFS and nutrient budgets

2.1 Introduction

As described in this action, objective 2, Statistics Norway carried out a sample survey on utilisation of mineral fertiliser and manure with reference year 2018 (Fertiliser Survey 2018). This survey could serve as a pilot for the IFS 2020 module Animal housing and manure management.

When finalising the preparation of the questionnaire for the Fertiliser Survey 2018, the final list of characteristics for the animal housing and manure management module in IFS was not yet decided. Comparison between the Fertiliser Survey 2018 and the final animal housing and manure management module indicates mainly the same topics and characteristics. Differences are:

- The Fertiliser Survey 2018 did not provide the average number of places by of type of livestock and type of housing.
- The Fertiliser Survey 2018 provided detailed information about types of fertilisers used on various crops.
- The Fertiliser Survey included a question on equipment for further processing of manure at the farm, see next chapter.

2.2 Manure treatment facilities

The Fertiliser Survey 2018 included one question relating to equipment used by the agricultural holding for processing of manure.

Mark for equipment used by the holding for treatment of livestock manure in 2018:

- ☐ Separating solid and liquid fractions
- ☐ Drying the manure
- ☐ Biogas treatment
- ☐ Composting
- ☐ None of the treatments used in 2018

The Fertiliser Survey 2018 stated that except for composting, treatment of livestock manure at agricultural holding level is infrequent. 2 990 holdings, about 10 % of the holdings with own manure, composted manure, while 220 holdings were drying the manure. The latter number is based on few answers. Both separating solid and liquid fractions and biogas treatment were too infrequent for this sample survey.

The Norwegian Agriculture Agency is managing a scheme covering grants to holdings delivering manure to biogas plants. In 2018, 68 900 tonnes of manure were covered by this scheme, of which 62 000 tonnes were delivered to one plant. A total of 35 holdings delivered manure to 6 biogas plants of which 4 were connected to agricultural holdings.

To conclude; except for composting, treatment of livestock manure is infrequent in Norway, although there is a subsidy scheme for delivering manure to biogas plants.

2.3 Imputation of partial responses in the Fertiliser Survey 2018

2.3.1 Background

When starting on this action in 2018, one objective was to look into possibilities for transmitting methods from action 1.1. Improve data processing and validation system for IFS, to the animal housing and manure management module. The method was tested on the Fertiliser Survey 2018.

2.3.2 Introduction

While reviewing the responses to the fertiliser survey of 2018, we found cases where the respondents had replied that they did not have livestock, while they for the same period had applied subsidies for livestock according to register data. We identified 3 cases, cf. table 1, where the respondents may have returned protest answers.

Table 2.1 Description of the cases and their respective numbers

Case	Number of protests	Description of cases
Case 1:	43	Responded that they had livestock on pasture but answered "NO" when asked if they had livestock. They had applied and been granted subsidies for livestock according to register data.
Case 2:	85	Responded that they had livestock but answered "NO" when asked if they had livestock on pasture. The had applied and been granted subsidies for livestock on pasture according to register data.
Case 3:	300	A combination of case 1 and 2; answered "NO" on both questions regarding if they had livestock and livestock on pasture, while for the same period they had applied and been granted subsidies for both schemes.

The survey was sent to a subsample of the population consisting of 5 260 respondents, out of these 3 469 had applied and been granted subsidies for livestock husbandry. From these we had 428 probable protest responses.

We investigated these cases by sending e-mail to 20 respondents chosen via a stratified random sample of the protest respondents and asking if the information provided in the survey were correct. Out of these, 15 responded that they in fact had livestock in 2018, and that the response had been incorrect.

Since the response clearly indicated that there were mistakes while filling in the questionnaire, we decided to send an e-mail to the remaining respondents, asking them to provide supplementary information to their survey responses. This resulted in obtaining responses from an additional 150 individual respondents. Leaving us with 278 partial responses.

To solve the issue of partial responses we decided to implement nearest neighbour methodology to complete the dataset.

2.3.3 Nearest neighbour imputation – method and validation

For development of an imputation method the work was divided into the following steps:

1. Understanding the variables, which are missing
2. Identification of auxiliary variables
3. Development of imputation method
4. Validation of the method

2.3.3.1 Understanding the variables, which are missing

In the data there were defined holdings with missing values. All of them can be divided into 3 cases:

Case 1. Holdings, which have missing or incorrect values in at least one of the variables:
DYR_GJTYP and FOR_GJTYP, where DYR € {MKU, AKU, AFE, GRIS, SAU, GEIT, HEST} and GJTYP € {KJBLOT, KUMBLOT, KJFAST, BAKK, PLATE, LAND, ITALLE, UTALLE};
DEL_GJTYP, where GJTYP € {KJBLOT, KUMBLOT, KJFAST, PLATE, LAND, ITALLE}; DYR OPEN DYRHALM DYRFLIS DYRTORV

Case 2. Holdings, which have missing or incorrect values in at least one of the variables:
DYR_BEITEINF, where DYR € {MKU, AKU, AFE, SAU, GEIT, HEST} and BEITEINF € {DYRINN, UKEINN, DYRUTM, UKEUTM, DYRLUFT, UKELUFT};
DYRINN DYRUTM DYRLUFT;

Case 3.Holdings, which are combination of cases 1 and 2.

2.3.3.2 Identification of auxiliary variables

For identification of auxiliary variables, were observed only holdings without any missing values (non-missing dataset). For non-missing dataset, there were calculated mathematical statistics of missing variables and provided correlation analyses of missing with potentially auxiliary variables.

For correlation of continues variables were used Person distribution, for categorical or discrete – Cramers' V, for combination of continuous and categorical - Kruskal-Wallis H Test and the pairwise comparison.

Based on expert opinion, there were defined the next potential auxiliary variables:

- For Case 1 PRO_DYR, where DYR € {MKU, AKU, AFE, GRIS, SAU, GEIT, HEST}, AUX_SONER, AUX_DRIFTSFORM, AUX_HUSDYR;
- For Case 2 HJ_BIU_DYR, HJ_BU_DYR, where DYR € {MKU, MKU_AKU, AFE, SAU, GEIT, HEST}, IDRIFT_REST, PRO_FULL, PRO_INNM, PRO_OVER, AUX_SONER, AUX_DRIFTSFORM, AUX_HUSDYR, AUX_BEITE

Auxiliary variables analysis for Case 1 has shown:

- For each DYR, where DYR € {MKU, AKU, AFE, GRIS, SAU, GEIT, HEST}, variable PRO_DYR defines existence of at least one variable DYR_GJTYP, where GJTYP € {KJBLOT, KUMBLOT, KJFAST, BAKK, PLATE, LAND, ITALLE, UTALLE}. PRO_DYR variable will be used as auxiliary.
- AUX_DRIFTSFORM has at least medium correlation with 20% of DYR_GJTYP variables, 87,5% of DEL_GJTYP and 66,7% of FOR_GJTYP. AUX_DRIFTSFORM will be used as auxiliary.
- AUX_SONER did not show correlation with DYR_GJTYP variables. The variable has shown not more than a medium correlation only with 37,5% of DEL_GJTYP variables. However, AUX_SONER provides correlation with 66,7% of FOR_GJTYP variables. The usage of the variable can improve the quality DEL_GJTYP variables but decrease the quality of other. The usage of the variable will be tested.
- AUX_HUSDYR did not show any satisfying results for DEL_GJTYP or FOR_GJTYP. AUX_HUSDYR will not be used as auxiliary.

- Existence of correlation between some auxiliary variables was checked, too. Thus, correlation between AUX_DRIFTSFORM and AUX_SONER was not defined, meaning that both variables can be used as auxiliary.

Correlation analyses for Case 2 has shown:

- HJ_BIU_DYR and HJ_BU_DYR have impact on DYR_BEITEINFO variables for approximately a half of them. Among those, for which correlation was not defined are variables of DYRLUFT, UKELUFT type. Variables of UKEINN, UKEUTM type does not have a high correlation, too. The variables will be used as auxiliary.
- To find out variables, which can be correlated with UKEINN and UKEUTM, variables IDRIFT_REST, PRO_FULL, PRO_INNM, PRO_OVER, AUX_HUSDYR were analysed. They did not show correlation. None of the variables will be used.
- However, AUX_SONER, AUX_BEITE, DRIFTSFORM have impact on UKEINN and UKEUTM. Following that the variables will be used.

To get rid of a significant influence of 0 values for variables, while calculating correlation, for each PRO_DYR is zero DYR_GJTYP not considered.

2.3.3.3 Development of imputation method

To save internal correlation between variables of one holding, for the imputation was used Multivariate Nearest Neighbour imputation.

Since auxiliary variables are of different types (quantitative, qualitative) and in different scaling (hundreds, tens, units), the standardization of quantitative auxiliary variables using “Min-Max scaling” was provided.

$$Xnew = (X - Xmin)/(Xmax - Xmin)$$

To satisfy strict requirements for imputation and to increase the speed of the process, were developed two additional parameters:

- AUX_PRO_DYR, which shows, for which types of DYR at least one value of DYR_GJTYP should be imputed.
- AUX_PRO_BEITE, which shows, for which types of DYR variables DYR_BEITEINF should be imputed.

These parameters, based on data from registers, show restrictions on input variables. They are developed as a combination of binary values, showing if the unit should have the defined variable.

The first parameter defines strict requirements for Case 1, the second – for Case 2, their combination – for Case 3. These variables define a pattern of animals’ existence for each holding. Since each holding can have only one pattern, the donors for holdings with missing variables were searched among the groups which had the same values of AUX_PRO_DYR or/and AUX_PRO_BEITE.

A distance between two units was calculated as sum of:

a. $\left(X_{aux_i} - X_{imp_i}\right)^2$, where i – quantitative;

b. $\begin{cases} 0, & if\ (X_{aux_i} = X_{imp_i}) \\ 1, & otherwise \end{cases}$, where i – qualitative, categorical without logical order;

c. $\left|X_{aux_i} - X_{imp_i}\right|$, where i – qualitative, categorical with logical order.

To exclude cases, when outliers are imputed, adjustment by mean values or probabilities for binary variables is applied. While choosing a donor for imputation, in the model is provided a calculation of mean value without an imputation of a variable and with an imputed value. If the difference of two values is more for $alpha$ percent of a mean without imputation, then the donor will not be imputed. If among 5 nearest donors it is not possible to find one, which with its imputation does not give a high change of mean, then none of donors is imputed and an expert imputation is required.

2.3.3.4 Validation of the method

For model validation the non-missing dataset was split into train and test subsets in proportion 80/20. The process of imputation ran for 5 different splits. The train subset was used as a set of potential donors. The test subset was considered as set with missing values, for which imputation was required.

Alpha was chosen to be equal to 30%.

After application of the method, there were 3 results, which were compared:

1. Test subset before imputation, correct data;
2. Test subset after imputation of data, following the rules of missing data according Case 1/Case 2;
3. Test subset after imputation of data, following the rules of missing data according Case 3.

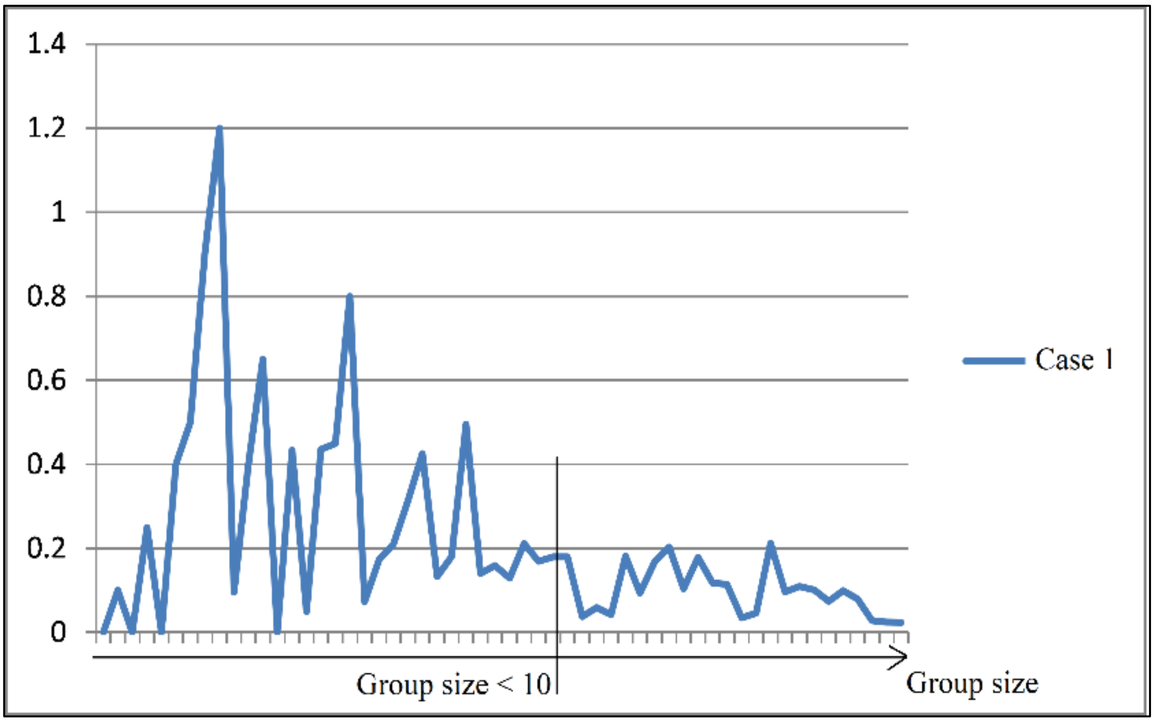
For imputation we are interested in high precision at macro level, whereas at micro can be found inaccuracies. Thus, even if the accuracy of data at micro level is not high, the difference of statistics in test subset before and after imputation is more important and shows real productivity of the model.

Results has shown for Case 1:

- At micro level imputation of at least 90% of 56 variables DYR_GJTYP for each unit was correct in 85% cases in average. What provides satisfying results at macro level as well. Important to notice, at macro level (Figure 1) the relative difference of some variables before and after imputation, is very high for those groups, which are of small size. Among the group size of which is normal/big, the relative difference of mean is varying from 12 to 30%.
- The most problematic at micro level was DEL_GJTYP variables, which provided only 23% accuracy. Applying AUX_SONE has helped a bit to improve the results of imputation for these variables in some cases, however in another provided worse results. By these reasons was decided not to use AUX_SONER as an auxiliary.
  At macro level imputation of DEL_GJTYP provides the difference in average for maximum 26% of mean (variable DEL_PLATE). The rest of the variables of DEL_GJTYP type has the difference less than 13%.
- For FOR_GJTYP results at macro level are even better and maximum is 19% difference (FOR_PLATE). The second worst - FOR_LAND with 17% of difference. The rest has the difference in mean of less than 13%.
- Imputation for DYR, OPEN, DYRHALM, DYRFLIS, DYRTORV gave the difference in accuracy for maximum 21%.

- The results of imputation for the whole dataset will be even better, because the smaller number of variables need to be imputed and number of examples in the groups is more.

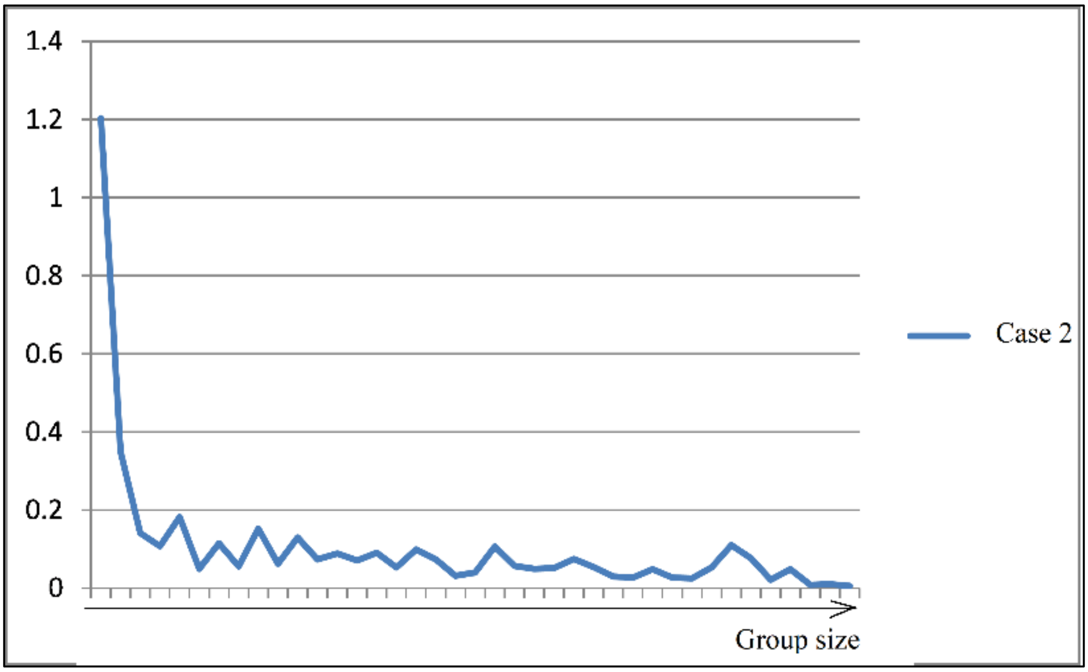
Figure 2.1 Case 1. Relative difference of means by group size increase



<u>*Results has shown for Case 2:*</u>

- For 2,7% of the holdings it was not possible to find any donor, because the same pattern of AUX_PRO_BE was not found.
- At micro level DYR_BEITEINFO in average only for 30 percent was given more than for 90 percent correct result. However, on macro level the results occurred to be of satisfying quality. Similarly, for Case 1, the most problematic groups were those, which are of a small size (GEIT_DYRLUFT, GEIT_UKELUFT. GEIT_DYRUTM). For the rest, a maximum relative difference of mean is 13%.
- Imputation of flags DYRINN, DYRUTM, DYRLUFT is of high accuracy (less than 1% difference).

Figure 2.2 Case 2. Relative difference of means by group size increase



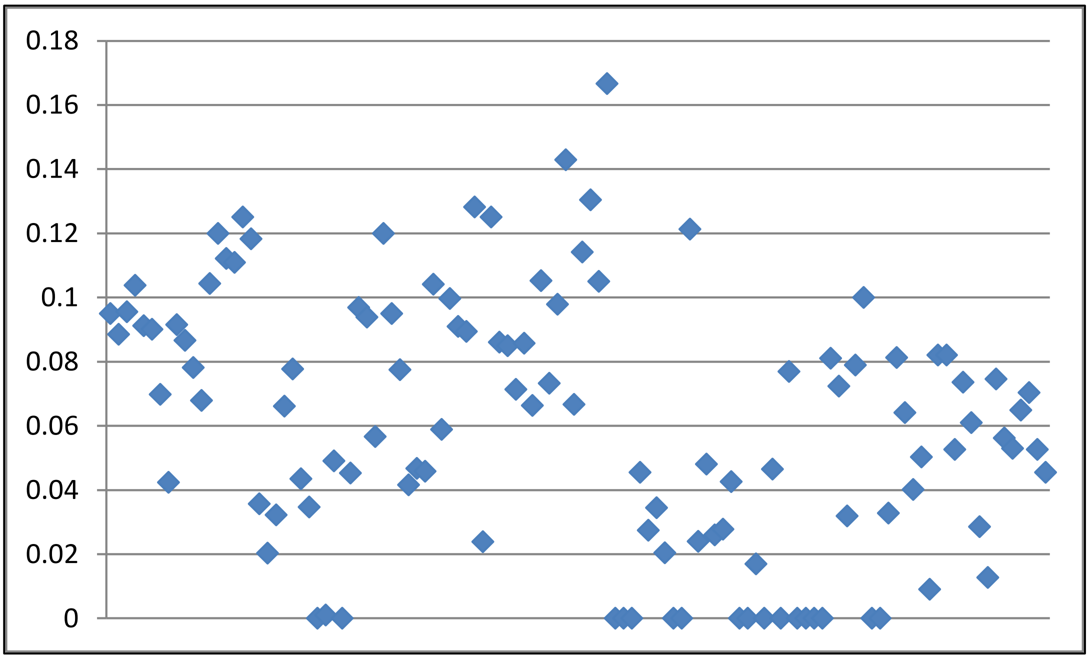
<u>*Results has shown for Case 3:*</u>

- For 7,7% of the holdings it was not possible to find any donor. It is occurring because of smaller possibility to find a donor following 2 patterns.
- In comparison with Case 1 and Case 2 (Figure 1, Figure 2), Case 3 gives higher relative differences.

Independent application of Case 1 and Case 2 did not manage to find donors for a smaller number of holdings than in Case 3. Moreover, relative mean's difference was worse in Case 3. Thus, because of less possibilities for donor finding and worse accuracy results of Case 3, in comparison with Case 1 and Case 2, it was decided to use for imputation following independently Case 1 and Case 2. For holdings, which have missing variables following both Case 1 and Case 2, the imputation was provided independently, too.

An imputation for the whole dataset increased values of means for maximum 16.67%.

Figure 2.3 Relative difference of means after imputation for the whole dataset



2.3.4 Comment and discussion

Implementing the imputation method to the dataset and reviewing it found that the method provided satisfactory results when doing controls at micro level. However, going forward it will be necessary to do a rigorous sensitivity analysis on the dataset at the macro level to ensure the method does not provide us with any unforeseen consequences in the variables imputed. The method will be used when editing the *Animal housing and manure management module* in IFS 2020.

3 Data on crop from grazing relevant to nutrient budgets

3.1 Background

Norway has signed an ESS-agreement regarding nutrient budgets. An objective in this grant has been to provide data on crop from grazing in Norway.

As it was found difficult to find suitable factors linking ruminants to nutrient uptake, the task was divided into three parts:

1. Using an average ruminant time on pasture for each animal species based on a comprehensive work done in connection to the Fertiliser Survey 2018.
2. Using estimates for energy uptake from pasture, measured in FEm (Feed unit), for each ruminant species.
3. Using suitable factors for nitrogen, phosphorus and potassium for pasture, linked via FEm.

3.2 Description of the data

Data and sources used in the calculations are shown in tables below.

Table 3.1 Livestock in Norway 2018 and 2019

	2018	2019
Dairy cow	219 016	210 351
Beef cow	98 017	99 659
Other cattle	564 958	562 945
Sheep 1 year and over	956 389	898 310
Goats, total	65 826	65 343
Horses	26 592	26 427

Source: Statbank, ssb.no

Table 3.2 Average number of weeks and days ruminants spent on inland pasture.2018

	Weeks	Days
Dairy cow	10,93	76,51
Beef cow	13,2	92,4
Other cattle	8,37	58,59
Sheep 1 year and over	11,85	82,95
Goats, total	8,58	60,06
Horses	13,55	94,85

Source: Fertiliser Survey 2018 («Bruk av gjødselressurser i jordbruket 2018»), table A79.

Table 3.3 Grazed inland pasture, divided into categories. 2018

	Area 2018	Percentage of total
Meadows on arable land	1 167 930	52 %
Surface-cultivated meadows	86 890	4 %
Other infield pastureland	952 310	43 %
Other forage crops	29 620	1 %
Total	2 236 750	100 %

Source: Fertiliser Survey 2018 («Bruk av gjødselressurser i jordbruket 2018»), table A80.

Part of the questionnaire of the Fertiliser Survey 2018 relating to grazing time and grazed areas is attached in annex 4.

3.3 Factors and calculations

3.3.1 Factors for forage uptake

The forage uptake by ruminants from pastureland was found in Hegrenes, A. & Asheim, L.J. «Verdi av fôr frå utmarksbeite og sysselsetting i beitebaserte næringar», NILF Notat 2006-15.

Table 3.4 Energy requirements per day. FEm

Animal species	FEm from pasture per day
Dairy cow	7,2
Beef cow	7,2
Other cattle	4,8
Sheep 1 year and over	2,68
Goats, total	1,2
Horses	7

Source: Hegrenes and Asheim (2006)

The base assumption for this method is that animals grazing on pastureland would maintain or increase their weight during their time on pasture. These factors do not include the eventuality of additional fodder given to the animals and must be taken as rough estimates.

3.3.2 Factors for nutrient content

The factors for nutrient content are based on the NorFor Feed table (http://feedstuffs.norfor.info/) for "Grass, mixed meadow". Data used were:

Table 3.5 Nutrient content of different maturity in «Grass, mixed meadow»

Group-Code		FEm/kg DM	NEL20 (MJ/kg DM)	DM g/kg	Crude protein g/kg DM	Phosphorus g/kg DM	Potassium g/kg DM
006-0511	Grass, mixed meadow. Early maturity	1,01	6,98	150	193	3,4	24,8
006-0512	Grass, mixed meadow. Medium maturity	0,96	6,5	170	160	3	25,4
006-0513	Grass, mixed meadow. Late maturity	0,89	6,17	190	148	2,8	22

Source: Norfor

To calculate the amount of nitrogen, we assumed that *crude protein* in the fodder contains 16% nitrogen (Harstad and Vangen 2015).

3.3.3 Deduced factors

3.3.3.1 Nutrient factors per FEm

To find grams N per FEm, with the assumption that 16% of crude protein (RP) is nitrogen, the following formula was used:

$$N_{g/FEm} = \frac{0{,}16 \cdot RP_{g/Kg/ts}}{FEm_{Kg/ts}}$$

Example using "Grass, mixed meadow. Medium maturity":

$$N_{g/FEm} = \frac{0{,}16 \cdot 160}{0{,}96} = 26{,}67_{g/FEm}$$

For phosphorous and potassium, the assumption of a percentage of crude protein was not used, because they were reported directly in the feed table.

Table 3.6 Calculated factors for grams N, P and K per FEm for different maturities of "Grass, mixed meadow"

		g N/FEm	g P/FEm	g K/FEm
006-0511	Grass, mixed meadow. Early maturity	30,57	3,37	25,05
006-0512	Grass, mixed meadow. Medium maturity	26,67	3,13	26,46
006-0513	Grass, mixed meadow. Late maturity	26,61	3,15	24,72

Example: Calculating nutrient content:
To calculate from grams to kilograms, and kilograms to tonnes, the calculations were multiplied with 0,001:

$$Kg\ N_{Dyr} = FEm_{dyr} \cdot Dager_{dyr} \cdot N_{g/FEm} \cdot 0{,}001$$

Example: Calculating uptake of Kg N for a single dairy cow grazing on "Grass, mixed meadow. Medium maturity":

$$Kg\ N_{mku} = 7{,}2\ \cdot 76{,}51\ \cdot 26{,}67\ \cdot 0{,}001 = 14{,}69$$

Example: Calculating total tonnes N for all dairy cows grazing on "Grass, mixed meadow. Medium maturity" in 2018:

$$TonnN_{dyr} = KgN_{dyr} \cdot Antall_{dyr} \cdot 0{,}001$$

Substituting the numbers into the formula above gives:

$$Tonn\ N_{mku\_all} = 14{,}69\ \cdot\ 219\ 016\ \cdot\ 0{,}001\ =\ 3\ 217{,}4$$

3.4 Results

3.4.1 Factors for nutrient uptake per ruminants

Factors for calculating total N, P and K for a single animal.

Table 3.7 Nutrient uptake per animal from «Grass, mixed meadow. Early maturity». Kg per animal

	Kg per animal		
	N	P	K
Dairy cow	16,84	1,85	13,53
Beef cow	20,34	2,24	16,34
Other cattle	8,60	0,95	6,91
Sheep 1 year and over	6,80	0,75	5,46
Goats, total	2,20	0,24	1,77
Horses	20,30	2,24	16,30

Table 3.8 Nutrient uptake per animal from «Grass, mixed meadow. Medium maturity». Kg per animal

	Kg per animal		
	N	P	K
Dairy cow	14,69	1,72	14,58
Beef cow	17,74	2,08	17,60
Other cattle	7,50	0,88	7,44
Sheep 1 year and over	5,93	0,69	5,88
Goats, total	1,92	0,23	1,91
Horses	17,71	2,07	17,57

Table 3.9 Nutrient uptake per animal from «Grass, mixed meadow. Late maturity». Kg per animal

	Kg per animal		
	N	P	K
Dairy cow	14,66	1,73	13,62
Beef cow	17,70	2,09	16,45
Other cattle	7,48	0,88	6,95
Sheep 1 year and over	5,91	0,70	5,50
Goats, total	1,92	0,23	1,78
Horses	17,67	2,09	16,41

3.4.2 Total nutrient uptake by ruminants from pastureland

Table 3.10 Total nutrient uptake using "Grass, mixed meadow. Early maturity". Tonnes

	2018			2019		
	N	P	K	N	P	K
Dairy cow	3 689	406	2 962	3 543	390	2 845
Beef cow	1 994	220	1 601	2 027	223	1 628
Other cattle	4 858	535	3 901	4 840	533	3 887
Sheep 1 year and over	6 500	716	5 221	6 106	672	4 904
Goats, total	145	16	116	144	16	116
Horses	540	59	434	536	59	431
I alt	17 726	1 952	14 236	17 197	1 893	13 811

Table 3.11 Total nutrient uptake using "Grass, mixed meadow. Medium maturity". Tonnes

	2018			2019		
	N	P	K	N	P	K
Dairy cow	3 217	377	3 192	3 090	362	3 066
Beef cow	1 739	204	1 725	1 768	207	1 754
Other cattle	4 237	497	4 204	4 222	495	4 189
Sheep 1 year and over	5 670	664	5 625	5 325	624	5 284
Goats, total	127	15	126	126	15	125
Horses	471	55	467	468	55	464
I alt	15 460	1 812	15 339	14 999	1 758	14 882

Table 3.12 Total nutrient uptake using "Grass, mixed meadow. Late maturity". Tonnes

	2018			2019		
	N	P	K	N	P	K
Dairy cow	3 210	380	2 982	3 083	365	2 864
Beef cow	1 735	205	1 612	1 764	209	1 639
Other cattle	4 227	500	3 927	4 212	498	3 913
Sheep 1 year and over	5 657	669	5 256	5 313	628	4 936
Goats, total	126	15	117	125	15	116
Horses	470	56	436	467	55	434
I alt	15 425	1 824	14 331	14 965	1 770	13 903

3.4.3 Total nutrient uptake divided by type of pastureland

The percentage distributions of area used by ruminants are held constant for all years for the purpose of these calculations, as seen in table 2.3.

Table 3.13 Total nutrient uptake divided by type of pastureland, given "Grass, mixed meadow. Early maturity". Tonnes

	2018			2019		
	N	*P*	*K*	*N*	*P*	*K*
Meadows on arable land	9 255	1 019	7 433	8 979	989	7 211
Surface-cultivated meadows	689	76	553	668	74	536
Other infield pastureland	7 547	831	6 061	7 322	806	5 880
Other forage crops	235	26	189	228	25	183
Total	17 726	1 952	14 236	17 197	1 893	13 811

Table 3.14 Total nutrient uptake divided by type of pastureland, given "Grass, mixed meadow. Medium maturity". Tonnes

	2018			**2019**		
	N	**P**	**K**	**N**	**P**	**K**
Meadows on arable land	8 073	946	8 009	7 832	918	7 771
Surface-cultivated meadows	601	70	596	583	68	578
Other infield pastureland	6 582	771	6 531	6 386	748	6 336
Other forage crops	205	24	203	199	23	197
Total	15 460	1 812	15 339	14 999	1 758	14 882

Table 3.15 Total nutrient uptake divided by type of pastureland, given "Grass, mixed meadow. Late maturity". Tonnes

	2018			2019		
	N	P	K	N	P	K
Meadows on arable land	8 054	952	7 483	7 814	924	7 260
Surface-cultivated meadows	599	71	557	581	69	540
Other infield pastureland	6 567	777	6 102	6 371	754	5 919
Other forage crops	204	24	190	198	23	184
Total	15 425	1 824	14 331	14 965	1 770	13 903

3.5 Discussion

The results are based on available resources and information; no special outtakes were taken from the fertiliser survey 2018. This was done so that the results could be reproduced for comparable years.

The numbers presented here reflect the mean values chosen as factors and are fundamentally uncertain regarding how mean values for energy requirements, foraging times, and nutrient composition have been used here.

Better estimates could be obtained by refining the questionnaire to ask specifically about the different types of inland pasture, as opposed to using a summarized approach for inland pasture.

When choosing which factors to report, the recommendation would be to use "Grass, mixed meadow, medium maturity". We can argue that the farmer wishes to maximize the utility of the forage, and thus does not want to allow ruminants on pasture at early maturity since it will stunt the growth of the pasture. On the contrary, at late maturity pasture might be better harvested and made into silage/grass bales.

By choosing “Grass, mixed meadow, medium maturity”, we can argue that the farmer puts their ruminants on pastures where there is enough biomass for the animals to maintain their weight and welfare, and if it is not viable for grazing, the farmer moves them elsewhere.

Based on these results, Norway will be able to deliver uptake of nitrogen and phosphorus from grazing to the nutrient budget run by Eurostat.

3.6 Definitions

Inland pasture	Inland pasture includes meadows on arable land, surface-cultivated meadows, other infield pastureland and grazed forage crops
FEm	Feed unit lactation
DM	Dry matter
NEL20	Net energy lactation 20 kg DM

4 Nordic seminar on agricultural statistics

The Nordic meeting in agricultural statistics organised by the Swedish Board of Agriculture and Statistics Sweden took place in Stockholm on 5 – 6 September 2018. Three people from Statistics Norway, Division for housing, property, spatial and agricultural statistics, participated. Under session IFS, Geir Inge Gundersen performed the presentation Census 2020 – a draft roadmap.

5 Deliverables

When relevant, a table with publications, products or other relevant outputs/deliverables which are due at the timing of the report should be included, together with their delivery date.

Deliverable	Delivery date	Description
Objective 2–3	2020.10.30	Sample survey on utilisation of inorganic fertiliser and manure in 2018 – questionnaire
Objective 4	2019.10.30	Agenda for Nordic meeting in Stockholm
Objective 4	2019.10.30	Census 2020 – a draft roadmap

6 Subcontracting

No sub-contracts involving third parties exist for this grant agreement.

7 Findings

Objective 1.1.

From the analysis of the data sources Sample Survey of Agriculture 2018 and A-ordningen it is concluded that questions relating to working hours in IFS could not be replaced by information from A-ordningen.

Based on a test of agricultural holdings linked to A-ordningen in 2019, we can expect that approximately 25 per cent of the holdings have relevant information in A-ordningen. The share is some less than expected.

The automated editing and imputation have been accurately tested on the Sample Survey of Agriculture 2018 relating to *Labour force and other gainful activities* with an acceptable outcome. Furthermore, automated editing and imputation have been used on the Fertiliser Survey 2018 with satisfactory outcome.

Objective 1.2

The new sample design is determined to improve the efficiency of the total sample size with respect to sample more agricultural holdings which are important for the accuracy of the compiled statistics. Another improvement of the new sample design is to ensure that sub-totals are improved. The population of agricultural holdings is divided into strata by size, location of the farm in counties and types of farming.

Objective 2

Results from the Fertiliser Survey 2018 indicates that treatment of manure at agricultural holdings is infrequent in Norway. Composting of manure occurred on approximately 10 % of the holdings with own manure. Other treatments were insignificant.

The Norwegian Agriculture Agency is managing a subsidy scheme covering holdings delivering livestock manure to biogas plants. In 2018, only 35 holdings were included. A total of 6 biogas plants received livestock manure, of which 4 were connected to agricultural holdings.

When reviewing the questionnaires from the Fertiliser Survey 2018, we detected that some holdings had replied that they did not have livestock while they at the same time had applied subsidies for livestock. Nearest neighbour methodology was implemented to complete the data. The method will be used when editing the IFS 2020 module Animal housing and manure management.

Objective 3

The fertiliser Survey 2018 provided information on:

- Area of agricultural land grazed in 2018 divided by category of land
- Number of various animals grazing by grazing time and type of pasture (agricultural land or outlying fields)

Furthermore, information was searched in written sources concerning energy requirement per day for various animals and nutrient content in grass of different maturity. Calculations based on these data have resulted in:

- Nutrient uptake per animal category and grass maturity
- Total nutrient uptake from pasture by grazing animals
- Total nutrient uptake from pasture distributed by type of agricultural land

8 Action list of future activities

Preliminary figures for IFS 2020 will be based on the new sample design. When final figures from the Census are available, an evaluation of the sample design will be done and adjusted if necessary. Further, the annual sample survey in agriculture will be based on the new sample design from 2022 and onwards.

A-ordningen will be used in 2023 and 2026, connected to labour force and other gainful activities. We will look into the possibilities to increase the utilisation of this register. It would be of interest to further develop the method for editing and imputation, e.g. by testing machine learning based on results from IFS 2020.

To reduce the extent of unit non-response, Statistics Norway can use more resources to follow up and motivate farms to respond. An analysis of which sub-groups of agricultural holdings more easily fail to complete a questionnaire will give Statistics Norway a better basis for implementing measures to reduce unit non-response. It would be of interest to further develop the method for editing and imputation, e.g. by testing machine learning based on results from IFS 2020.

9 Conclusions

Objective 1.1.

In IFS 2020, information from A-ordningen will be pre-printed in the questionnaire as a suggestion for the respondents, and respondents will be enabled to edit prefilled information. Furthermore, Statistics Norway will use A-ordningen in editing and to detect missing response.

A system for editing and imputation, using the program R is tested on data from the Sample Survey of Agriculture 2018. This system will be further tested and used in the preparation of data in the Agricultural Census 2020:

- Firstly, a preliminary version will be used when preparing preliminary data
- Secondly, the lessons learned from this work will be used to adjust the preliminary version before preparing the remaining data in the Agricultural and Forestry Census 2020

Due to this step of modernisation of the agricultural statistics, we expect that the respondent burden will be reduced for holdings with pre-printed information from A-ordningen, as well as increased quality in reported information on labour force. We also expect a rise in quality of the labour force data because A-ordningen should be used in the editing of these variables.

More use of automated editing and imputation will have at least two advantages:

- The editing will be more efficient with less use of human resources
- More editing and imputation by a given set of computer-based rules will lead to a process which is more coherent and punctual since it is better documented and less dependent of human resources

Objective 1.2.

Due to improved sample design, the number of holdings selected for the future annual sample survey in agriculture is stipulated to approximately 6 000 compared with approximately 8 000 in the previous sample design.

A general procedure will be developed of the Division of methods to make it easy to calculate the coefficient of variation along with compiling the statistics. This achievement will improve the analysis of the figures disseminated.

Objective 2

Despite of some differences, experience from the Fertiliser Survey 2018 has been of great value when creating the IFS questionnaire and in planning the editing routine.

Except for composting, further treatment of livestock manure is insignificant in Norway.

Objective 3

The recommendation is to use the factors for “Grass mixed meadow, medium maturity” when reporting data to Nutrient Budget in Eurostat. This step will lead to a more accuracy and complete calculation of the Nutrient Budget when output of N and P from grazing is included.

Bibliography

References relating to chapter 1.1.2 Automated editing and imputation:

1. Pannekoek, Zhang (2015) "Optimal adjustment for inconsistency in imputed data" https://eprints.soton.ac.uk/391013/
2. Hagesæther, Zhang (2008) "Statistical registers by restricted neighbour imputation – an application to the Norwegian Agriculture Survey" https://www.iaos-isi.org/papers/CS_4_1_Hagesher_Zhang.pdf

References relating to chapter 3 Data on crop from grazing relevant to nutrient budgets:

1. Harstad O. M. & Vangen O. 2015. "Råprotein", Store Norske Leksikon. Accessed 2020.05.13. https://snl.no/r%C3%A5protein.
2. Hegrenes A. % Asheim L. J. 2006. "Verdi av fôr frå utmarksbeite og sysselsetting i beitebaserte næringar", NILF Notat 2006-15
   "Norfor". Accessed 2020.05.13. http://feedstuffs.norfor.info/

List of tables

List of figures

10 Annexes

Annex 1. Animals per calculated manure unit

Type of animal	No. of animals per calculated manure unit
Adult horses	2
Dairy cow	1
Beef cow	1,5
Other cattle	3
Sheep and goats fed during winter	7
Breeding sows and boars	2,5
Slaktegriser[1]	18
Laying hens	80
Broilers[1]	1 400
Pullets[2]	550
Breeding turkeys	40
Turkeys for fattening[1]	240
Breeding ducks	40
Ducks for fattening[1]	300
Breeding geese	20
Geese for fattening[1]	150
Breeding rabbits	40
Rabbits for fattening[1]	600
Breeding foxes including cubs	25
Breeding minks including cubs	40

[1]The number of calculated manure units is based on the number of slaughtered animals per year.
[2]The number of calculated manure units is based on the number of delivered animals per year.

Annex 2. Counties in Norway including 2019 and from 2020

Counties – 2019	Counties[1] 2020 –
01 Østfold	30 Viken
02 Akershus	
06 Buskerud	
03 Oslo	03 Oslo
04 Hedmark	34 Innlandet
05 Oppland	
07 Vestfold	38 Vestfold og Telemark
08 Telemark	
09 Aust-Agder	42 Agder
10 Vest-Agder	
11 Rogaland	11 Rogaland
12 Hordaland	46 Vestland
14 Sogn og Fjordane	
15 Møre og Romsdal	15 Møre og Romsdal
16 Sør-Trøndelag	50 Trøndelag
17 Nord-Trøndelag	
18 Nordland	18 Nordland
19 Troms	54 Troms og Finnmark
20 Finnmark	

[1]Counties as from 2020 is not always the exact aggregate of the former counties in 2019. Some municipalities or parts of municipalities have been transferred to another county.

Annex 3. Farm types

1. Specialist cereals, oil-seeds and protein crops
2. General field cropping
3. Specialist horticulture and specialist permanent crops
4. Specialist dairying
5. Specialist cattle – rearing and fattening
6. Cattle – dairying, rearing and fattening combined
7. Specialist sheep
8. Goats and other grazing livestock
9. Specialist granivores
10. Mixed cropping, mixed livestock and mixed crops – livestock

Sample survey on utilisation of inorganic fertilizer and manure 2018

The respondents filled in use of fertilizer and manure for one of the categories of agricultural crops:

- Without agricultural area in use
- Meadows on arable land
- Meadows on surface-cultivated land
- Infield pastureland
- Crops for green fodder and silage
- Barley
- Oats
- Spring wheat
- Winter wheat
- Oil seeds
- Potato
- Onion
- Cauliflower and broccoli
- Carrot

Use of fertilizers

Barley as an example

How many decares of the area of barley was fertilized at least once with inorganic fertilizer, livestock manure or other organic fertilizer in 2018?

Include fertilizers spread during growing season and in the autumn.

Did you use inorganic fertilizer on the area of barley? ☐ Yes ☐ No

Decares of barley fertilized at least once with inorganic fertilizer: __________ decares

Did you use manure on the area of barley? ☐ Yes ☐ No

Decares of barley fertilized at least once with manure: __________ decares

Did you add water to the manure before spreading? ☐ Yes ☐ No

Include water from stirring of manure, precipitation, wash water from milking system etc.

What kind of equipment were used when spreading the manure?

Broadcasting with	Per cent of the manure spread
☐ Tank wagon	☐ %
☐ Tow hoses	☐ %
Band spreading with tank wagon	
☐ Trailing hose	☐ %
☐ Trailing shoe	☐ %

Band spreading with tow hoses
- Trailing hose %
- Trailing shoe %

Injection with tank wagon
- Shallow injection %
- Deep injection %

Other equipment
- Tank wagon etc. with cannon for long spreading %
- Rear discharge spreader %

Incorporation time for broadcasted manure:
- Manure incorporation within 1 hour %
- Manure incorporation within 1-4 hours %
- Manure incorporation within 4-12 hours %
- Manure incorporation after 12 hours %

- No incorporation into the soil %

Mark type of fertilizers used on the largest parcel of barley in 2018:

Inorganic fertilizer
- NPK 8-5-19 micro
- NPK 12-4-18 micro
- NPK 18-3-15
- NPK 20-4-11
- NPK 21-6-6
- NPK 22-2-12
- NPK 22-3-10
- NPK 25-2-6
- NPK 27-3-5
- OPTI-NK 22-0-12
- OPTI-START 12-23-0
- OPTI-PK 0-11-21
- OPTI-P 0-20-0
- OPTI-NS 27-0-0
- SULFAN 24-0-0
- Kalimagnesia patentkali 24/25
- Kalimagnesia 49 PCT 60ER
- Kaliumsulfat 41 PCT
- Kalksalpeter/NITRABOR
- Urea 46-0-0

Domestic manure
- Cattle, slurry
- Cattle, liquid manure 1-4%
- Cattle, solid manure
- Cattle, deep litter
- Cattle, liquid manure <1%
- Pig, slurry
- Pig, solid manure
- Pig, deep litter
- Pig, liquid manure <1%
- Sheep, slurry
- Sheep, solid manure
- Sheep, deep litter
- Goat, slurry
- Goat, solid manure
- Goat, deep litter
- Laying hens, slurry
- Laying hens, solid manure
- Broilers, slurry
- Broilers, solid manure
- Manure from horses

Other organic fertilizer

Organic and inorganic fertilizers

- ☐ Silage effluent
- ☐ Sewage sludge
- ☐ Green manure
- ☐ Meat bone flour
- ☐ Marihøne Pluss 8-4-5
- ☐ Marihøne 4-1-2
- ☐ Helgjødsel 12-2-10
- ☐ Helgjødsel 18-1-10
- ☐ Helgjødsel 18-1-2
- ☐ Helgjødsel 30-1-1

For each type of fertilizer, mark time for spreading the fertilizer, area fertilized and amount of fertilizer per decare:

	Time of spread				
	Spring	Summer	Autumn	Area	Kg per decare
Type of fertilizer	☐	☐	☐		

Manure imported to the agricultural holding in 2018

Did the holding import manure from other holdings in 2018?? ☐ Yes ☐ No

Amount of imported manure by slurry and solid manure:

Slurry	☐ tonnes ☐ m^3	
Solid manure	☐ tonnes ☐ m^3	

Use of the imported manure:

Used on agricultural area of the holding		%
Used for energy production, compost etc.		%

Use of manure from own agricultural holding in 2018

Were there domestic animals on the holding in 2018? ☐ Yes ☐ No

Percentage distribution of own domestic manure in 2018:

Refer to stored manure. Excluding manure dropped during grazing.

Spread on agricultural area of the holding		%
Used for energy production or compost		%
Exported to other holdings for use on agricultural area		%
Delivered to other use off the holding		%
Remaining own manure on the holding end of 2018		

Manure storage on the holding

Storing of manure from **laying hens** after taken out from the laying hen housing
☐ Pits below animal confinement
☐ Outdoor on permeable ground
    ☐ Without cover
    ☐ With cover
☐ Outdoor on impermeable ground
    ☐ Without cover
    ☐ With cover

Temporary storing of manure from laying hens: ☐ Yes ☐ No
Kind of storing:
☐ Pits below animal confinement
☐ Outdoor on permeable ground
    ☐ Without cover
    ☐ With cover
☐ Outdoor on impermeable ground
    ☐ Without cover
    ☐ With cover

Bedding added to the manure from laying hens:
☐ Yes
☐ No

Kind of bedding used:
☐ Straw
☐ Sawdust
☐ Peat etc.

Storing of manure from **broilers** after taken out from the broiler housing
☐ Pits below animal confinement
☐ Outdoor on permeable ground
    ☐ Without cover
    ☐ With cover
☐ Outdoor on impermeable ground
    ☐ Without cover
    ☐ With cover

Temporary storing of manure from broilers: ☐ Yes ☐ No
Kind of storing:
☐ Pits below animal confinement
☐ Outdoor on permeable ground
    ☐ Without cover
    ☐ With cover

☐ Outdoor on impermeable ground
- ☐ Without cover
- ☐ With cover

Bedding added to the manure from broilers:
☐ Yes
☐ No

Kind of bedding used:
☐ Straw
☐ Sawdust
☐ Peat etc.

Holdings keeping cattle, pig, sheep, goat or horse

Mark for manure storage used in 2018
Include any temporary storage.

	Manure from:						
	Dairy cows	Beef cows	Other cattle	Pigs	Sheep	Goat	Horse
Pits below animal confinement, slurry	☐	☐	☐	☐	☐	☐	☐
Other slurry storage	☐	☐	☐	☐	☐	☐	☐
Pits below animal confinement, solid manure	☐	☐	☐	☐	☐	☐	☐
Outdoor on permeable ground, solid manure	☐	☐	☐	☐	☐	☐	☐
Outdoor on impermeable ground, solid manure	☐	☐	☐	☐	☐	☐	☐
Liquid manure (urine) tank	☐	☐	☐	☐	☐	☐	☐
Manure in deep litter systems, indoor	☐	☐	☐	☐	☐	☐	☐
Manure in deep litter systems, outdoor	☐	☐	☐	☐	☐	☐	☐

Slats between animal confinement and the pit below: ☐ Yes ☐ No

Cover used on tanks or ponds:

	Manure from: Cattle	Pig	Sheep
Impermeable roof	☐	☐	☐
Permeable cover such as straw or natural crust	☐	☐	☐
Artificial cover such as plastics or leca (light expanded clay aggregate)	☐	☐	☐
Without cover	☐	☐	☐

Kind of bedding used to sponge up moisture from manure:

☐ Straw
☐ Sawdust
☐ Peat etc.
☐ None

Livestock manure and animal housing

Amount of livestock manure stored by spring 2018, by type of storage:

Pits below animal confinement, slurry	☐	%
Other slurry storage	☐	%
Pits below animal confinement, solid manure	☐	%
Outdoor on permeable ground, solid manure	☐	%
Outdoor on impermeable ground, solid manure	☐	%
Liquid manure (urine) tank	☐	%
Manure in deep litter systems, indoor	☐	%
Manure in deep litter systems, outdoor	☐	%

Storage capacity for manure, by type of storage:

Pits below animal confinement, slurry	☐	months
Other slurry storage	☐	months
Pits below animal confinement, solid manure	☐	months
Outdoor on impermeable ground, solid manure	☐	months
Liquid manure (urine) tank	☐	months
Manure in deep litter systems, indoor	☐	months

Treatment of livestock manure in 2018

☐ Separating solid and liquid fractions
☐ Drying the manure
☐ Biogas treatment
☐ Composting
☐ None of the treatments used in 2018

Grazing and exercise yard

Grazing in outlying fields in 2018
- [ ] Yes
- [ ] No

Grazing on agricultural land in 2018
- [ ] Yes
- [ ] No

Agricultural land grazed at least once in 2018
- [ ] Meadows on arable land ____ decares
- [ ] Meadows on surface-cultivated land ____ decares
- [ ] Infield pastureland ____ decares
- [ ] Crops for green fodder ____ decares

Domestic animals in exercise yards in 2018
- [ ] Yes
- [ ] No

Number of animals grazing, weeks spent on grazing and weeks in exercise yard:

	Number of animals grazing on agricultural land	Weeks spent on grazing on agricultural land	Number of animals grazing in outlying fields	Weeks spent grazing in outlying fields	Number of animals in exercise yards	Weeks spent in exercise yards
Dairy cows						
Beef cows						
Other cattle						
Sheep 1 year and over						
Goat 1 year and over						
Horse						