Early estimates of Energy Balances

Reinoud Segers

Final technical report for Eurostat grant 832086 — 2018-NL-Energy

CBS Den Haag
Henri Faasdreef 312
2492 JP The Hague
P.O. Box 24500
2490 HA The Hague
+31 70 337 38 00

www.cbs.nl

project number

SLO
28 juni 2020

Index

1. Introduction

EU policy on energy and climate is increasingly considered as important and is now one of the priorities of the new European Commission. To support discussion on the progress timely data on the development in the energy system are needed. Thus far, annual data on the energy balances of each country are published by Eurostat approximately one year after the references year, which is considered as rather late. This report describes testing new mini-questionnaires of Eurostat for the Dutch situation, aiming at publishing a summary energy balance half a year after the reference year. In order to achieve this, the deadline for the Eurostat mini-questionnaires is set at the end of May.

This project (832086 — 2018-NL-Energy) is supported by a grant from Eurostat.

The national system for energy balances used to deliver an almost complete published energy balance in June on the year t-1. Data for this balance are also used for national provisional estimates of greenhouse gas emissions. Almost all main aggregates are mainly based on empirical data.

In order to be able to supply Eurostat with the data on the main aggregates of the energy balance we intended to adapt our current statistical process to have a complete integrated and consistent provisional data set at subsector level ("Statbase" or "stable point 5" called within Statistics Netherlands) available at May 15th, which could be used to supply Eurostat and our national users. In order to be able to have this Statbase available in time we need to sufficiently finalise data collection and editing from administrative sources and surveys by the end of April, creating 2 weeks for integration and final checks.

For three important subprocesses substantial efforts were foreseen to adapt our statistical process in such a way that we have more reliable and robust process to finalise this data collection in time:

1. The subprocess that delivers data for final consumption of natural gas and electricity for the residential, services and agricultural sector (based on an administrative sources: client files from grid operators).
2. The subprocess that delivers data for industrial companies with a simple energy balance: only final energy consumption (based on survey data).
3. The subprocess that delivers data for industrial companies with a more complex energy balance: final energy consumption, final non-energy consumption and autoproducer transformations (based on a combination of surveys).

Definitions, concepts and units of our national system differs on some essential parts from the Eurostat conventions. Before submitting data to Eurostat we always convert the data to comply with the international agreements. In order to have a quick and robust conversion process it was needed to design a special output module for the mini-questionnaires to automate the conversion process as much as possible.

In the end we compared the provisional data with the data send to Eurostat as part of the existing regulations and examined the magnitude and causes of the differences.

These 5 main challenges (improving three subprocesses, adapting output module and comparison of results) result in a project broken down into 5 different tasks. Each of these tasks is described in a separate chapter.

2. Task 1: Early estimates for the residential, services and agricultural sector from client files of grid operators

The data for these sectors are based on the client files from the grid operators. In principle, the annual raw data becomes available by the beginning of March, though in past years often data from one or two grid operators were late or initially of insufficient quality. The production process from raw data to final figures encompasses several steps. These steps involve amongst others linking to several other sources of admin and GIS data, advanced (text) detection and allocation algorithms. This process is combined with detailed plausibility checks, partly by visual inspection of individual records. By the month of July final energy figures on a low regional level can be derived from the resulting database.

An early estimate for a national total of these sectors was derived from an early stage of this production process. As a first step, to increase the chance of all data in time, we informally send an e-mail to our contacts from the grid operators to announce the formal letter asking for the data. This combination of formal and informal approach resulted in having timely raw data both for the reporting year 2018 and 2019.

To arrive at early estimates for the sectors services, households and agriculture, we started processing the raw administrative data for each grid operator as soon as we received the client files. This was different from earlier years, where we first combined all client files before further processing the data. In this way, we could add the delayed client files from one grid operator in a very late stage without processing all connections from the start.

To categorize each connection to the correct sector, we used the information available from last year for all existing connections. This was only doable by using a data processing pipeline that we developed last year. To detect sector changes of existing connections, and to categorize the approximately 230,000 new connections, we used two business registers (ABR[1] and Locatus[2]) and name recognition algorithms. The name recognition algorithms used were significantly improved in accuracy and performance last year. At last, the connections with the largest supplies were checked by hand, and the obtained aggregates at NACE 1 digit level were compared with the figures from earlier years.

This task resulted in achieving milestones 1 and 6 as outlined in the project proposal “Early estimates for the residential, services and residential sector from client files of grid operators, first cycle and second cycle.”

[1] This is a businessregister by Dutch Statistics based on data from the Tax Authority and the Chamber of Commerce
[2] This is a commercial business register which includes only the service sectors

3. Task 2: Early estimates for industrial companies with a simple energy balance

The data for these companies are based on a survey, which is a census design for the big companies or companies using rare energy commodities, a sample survey for intermediate companies and model estimates for the small companies. The problem to have data that are timely and of sufficient quality is that several companies send in their data late and/or with poor quality or even do not send data at all.

We intended to gain one or two weeks by sending the survey a little earlier, but this turned out to be impossible due to the complexities with the interaction with the business register of our institute.

A second measure (adapting survey design), however, was successful. We have split the population into two groups:

- Large companies, companies with rare energy commodities or very large energy consumptions (census survey) (about 500).
- Intermediate companies (sample survey) (about 1900).

The division makes it possible to focus the attention in the period before the early estimates on the first group, with more efforts for data editing and the ability to check if an estimate of the previous year is available in case of lacking response of sufficient quality.

For the early estimates of reference year 2018 this measure was not really needed, because half May 2019 we already had a very good response rate of about 90% checked data for both groups, which was almost the same as used for the final data (though with improved data checking/editing).

However, for the early estimates of reference year 2019 splitting the population proved to be useful, because we had problems in sending out the survey as planned and the data checking a.o. due to change in staff. Half May 2020 sufficiently checked response was 53% for the larger companies and 47% for the intermediate.

For the larger companies missing data were supplemented with previous year data, which was not possible for only 25 companies. We manually checked these companies and did not observe a company known for a very large energy consumption. For intermediate companies raising was applied as was already implemented in this sample survey.

For 2019 the accuracy of the early estimate for the industrial companies with a simple energy balance will probably be less than for 2018 due to the much lower amount of data used. To tackle any possible severe inaccuracies we also checked the outcome at the level of subsectors of the industry after aggregating with companies with a complex energy balance. We did not observe any very remarkable trends.

This task resulted in achieving milestones 2 and 7 as outlined in the project proposal “Early estimates for industrial companies with a simple energy balance, second cycle, first cycle and second cycle”.

4. Task 3: Early estimates for industrial companies with a more complex energy balance

They energy data for these data come from several monthly and annual census surveys for about 100 companies. The challenge for these group of companies, with sometimes a quite complex energy balance, is to have timely validated data. In the old situation we had until the end of the month May to have the data sufficiently ok, in the new situation we need to have to data of sufficient quality by the end of April. Especially for the companies that are subject to an annual survey this is a substantial increase in timeliness, because response time for annual surveys according to the Dutch statistical law is two months, implying that we expect annual raw data only by the beginning of March.

This means this that we will have to design tools and documents to support efficient validation which could be carried out by different persons.

Especially for renewable energy we use a lot of administrative sources. In principle these sources are sufficiently timely available. However, in practice partners sometimes have delays in sending the (correct and complete) data. Part of this task is more intensive communication with data suppliers in order to ensure timely data delivery.

We started developing a new data editing tool for the survey for industrial companies with a more complex energy balance. Unfortunately, this tool is not finished yet, because of underestimated complexities in the interaction with central IT system of energy statistics, complications in the organization of development of IT tools and reduced availability of crucial staff.

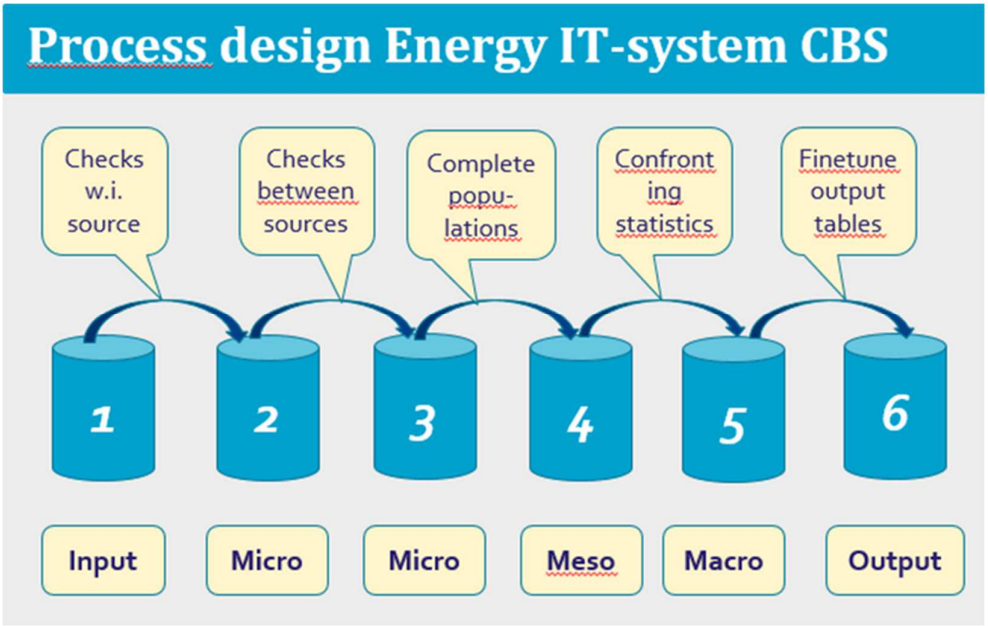
As a result we still had to check the data without the new tool, but by carefully looking directly at the (Excel) surveys and the build-in checks. We noticed that for the reference year 2019 the response rate half May was substantially improved compared to previous year. Apparently companies got used to the survey which was redesigned in 2018.

For renewable energy we have sometimes some trouble in obtaining timely (administrative) data from the municipal waste plants, but with intensive communication and sometimes using an alternative administrative source we obtained sufficient data in time.

The work done within this task thus far did not contribute to achieving milestones 3 and 7 (“Early estimates for industrial companies with a more complex energy balance, first and second cycle”). We achieved the milestones in alternative way (existing way of data checking) with the drawback that improving efficiency and robustness was not achieved. However, we did make substantial progress in developing the data editing tool and the expectation is that will be to use the tool for the next cycle.

5. Task 4: Output module to convert data from national system to Eurostat system

Recently we have started using a new IT system for the core of our energy statistics. This systems integrates all data at the level of individual companies (micro level) and data at the level of sub-sectors (meso level). Output of this systems are stable points 5 and 6 describing fulling consistent energy balance at the (detailed) sector level. The figure shows the process design of the IT-system which was tailor made for energy statistics, based on the general CBS process design for making statistics. A 'stable point' is the word we use for data which are 'in rest' after some kind of processing has been taken place.



We have designed tools to convert the data from these stable point 5 (with national definitions and classifications) to definitions and classifications of the mini-questionnaires of IEA and Eurostat (part of stable point 6). We could use the import facility of the IEA questionnaires to import the data to the IEA mini's and use the import facility of the Eurostat questionnaire to import the data from the IEA questionnaire. However, the additional data of the Eurostat mini-questionnaires had to be copied manually from the output file of our IT system into the questionnaire.

The whole process was to a large extent automated though some manual interferences were still needed and currently hard to avoid, because only part of the data can be processed via the IEA miniquestionnaires.

The Eurostat mini-questionnaires with references to or comparisons with previous annual and actual monthly data proved to be useful for validation and discovering issues were the automated process was still insufficient. Sometimes pragmatic repair actions (e.g. copying t-1 data for less important items) were needed.

We are gradually evolving to the situation where we have one central database with data on micro and meso level that will be used for making consistent monthly and annual publication, both final ones and early estimates. The mini questionnaires are fully integrated in this development. In 2019/2020 the output module migrated from Python 2 tot Python 3 and because the code for the miniquestionnaires is fully integrated the tool that is transition caused no special trouble for the mini-questionnaires. For the reference year 2019 we could use the output tool in the same way as for the reference year 2018.

Our experience is that in the current situation we still have to spend attention to data logistics and technical issues, because we have to cope within one month with both IEA and Eurostat miniquestionnaires with a three weeks difference in timing. In this three weeks also improved data becomes available which needs to be included as well, further complicating matters. Our task would be more simple (and less vulnerable for errors) if there would be one set of joint Eurostat-IEA miniquestionnaires with a facility to import all data, like for the annual questionnaires.

This task resulted in milestone 4 and 9: Output module to convert data from national system to Eurostat system, including checking and sending to Eurostat, first cycle and second cycle.

6. Task 5: Examine the differences between early estimates and final data send later

The main question is: how good are the early data?. We have examined the quality of the data for the reference year 2018 by converting the data to a summary energy balance and compared this with energy balance available on the website of Eurostat in June 2020. This balance is based in the joint annual questionnaires send by Statistics Netherlands at the end of November 2019 and some limited revisions afterwards (Table 1).

Table 1. Comparision of Energy balance form miniquestionnaires with energy balance from joint annual questionnaire for 2018 in PJ NCV

	Total	Coal	Oil	Gas	Renewables	Other	Nuclear	Electricity	Heat
Energy balance calculated from miniquestionnaires									
PrimaryProduction	1550		76	1.162	243	33	34	0	2,63
Imports	8514	348	6.271	1.769	20	9		96	0
Exports	6319	5	4.610	1.577	57	2		68	0
StockChanges	-45	1	29	-68	-6	0			
MarineBunkers	475		475		0				
AviationBunkers	170		170						
Total energy supply	3056	344	1120	1286	200	41	34	29	3
NetTransformationInput	436	312	32	405	124	37	34	-409	-98,06
EnergySector	200	8	85	58	0	0		34	13,78
DistLosses	24	0	0	0	0	0		19	4,11
TotFinalNonEnergyConsumption	529	0	428	100	0				
TotFinalEnergyConsumption	1881	24	587	723	76	4		383	84,27
FinConsumptionIndustry	563	24	133	212	7	1		126	60,41
FinConsumptionTransport	450	0	420	2	21	0		7	
FinConsumptionOtherSectors	867	0	34	509	48	3		250	23,87
Stat.difference	-13	0	-12	-1	0	0	0	1	
Energy balance form joint annual questionnaires									
PrimarProduction	1548	0	80	1.163	237	29	34	0	6
Imports	8621	349	6.319	1.827	21	8		96	0
Exports	6395	5	4.628	1.635	57	2		68	0
StockChanges	-44	1	30	-68	-6	0			
MarineBunkers	469		468	0	1				
AviationBunkers	170		170						
Total energy supply	3091	344	1163	1287	194	36	34	29	6
NetTransformationInput	441	310	56	393	120	34	34	-412	-94
EnergySector	203	8	86	63	0	0		34	11
DistLosses	23	0	0	0	0	0		19	4
TotFinalNonEnergyConsumption	537	0	435	101	0				
TotFinalEnergyConsumption	1881	25	592	716	74	2		389	84
FinConsumptionIndustry	573	25	137	214	8	0		130	59
FinConsumptionTransport	454	0	422	3	21	0		8	
FinConsumptionOtherSectors	855	0	33	499	45	2		251	25
Stat.difference	6	0	-6	13	0	0		-1	0
Difference									
PrimaryProduction	-3	0	4	1	-6	-5	0	0	3
Imports	107	0	48	58	1	-1	0	0	0
Exports	76	0	18	58	0	0	0	0	0
StockChanges	1	0	1	0	0	0	0	0	0
MarineBunkers	-6	0	-7	0	1	0	0	0	0
AviationBunkers	0	0	0	0	0	0	0	0	0
Total energy supply	36	0	43	1	-6	-5	0	0	3
NetTransformationInput	5	-1	24	-12	-4	-3	0	-3	4
EnergySector	3	0	1	5	0	0	0	0	-2
DistLosses	0	0	0	0	0	0	0	0	0
TotFinalNonEnergyConsumption	8	0	7	1	0	0	0	0	0
TotFinalEnergyConsumption	1	1	6	-7	-2	-2	0	6	0
FinConsumptionIndustry	10	1	5	2	1	-1	0	3	-1
FinConsumptionTransport	3	0	2	1	0	0	0	0	0
FinConsumptionOtherSectors	-13	0	-1	-10	-3	-2	0	2	1

We compiled the energy balance from the miniquestionnaires by using the caloric values from the 2018 joint annual questionnaires. Furthermore, we assumed that all oil products reported as final energy consumption for aviation were used for aviation bunkers, which is for the Netherlands a reasonable assumption, because the contribution of domestic aviation is less than 1 percent. For nuclear energy no miniquestionnaire is available. Therefore, we calculated the production and transformation input from the nuclear electricity production in the E&H miniquestionnaire and the average efficiency for the years 2013-2017.

In general the difference between the final data and the miniquestionnaires was typically one or two percent, also for the finale energy consumption which is relevant for calculating the share of renewables.

The largest difference was observed for import and exports of natural gas. This relates to improved data on import and export of LNG which used to be available only annually with a timing too late for the miniquestionnaires. Given the increased relevance of these flows we now have monthly data collection on these already included in M-1 data on natural gas.

A second major improvement in terms of difference PJ was related to imports and conversion processes of oil. The energy flows related to oil are very big in the Netherlands. Relatively small changes can be big compared to changes for other commodities. But also for oil the difference between the provisional data from the miniquestionnaires and the final data from the joint annual questionnaires was only 1 percent for almost all flows.

For renewables we observed a difference for final consumption and transformation input. The difference for final consumption was related to the revision of the household biomass data which occurs every six years after results of the six-annual survey become available. The difference for transformation input was of technical nature. Something was not entirely correct and time was lacking to observe and correct this. We checked whether this problem affected the related transformation output (see table 2), but this was not the case.

Table 2. Electricity and (sold) heat production by energy source 2018

Product	Unit	Total	Coal	Oil	Gas	Renewabl	Nuclear	Other
Miniquestionnaires								
Electricity	GWh	113.672	29.884	1.292	57.536	18.649	3.515	2.797
Heat	TJ	107.795	2.750	11.431	69.644	14.453	0	9.517
Joint annual questionaires								
Electricity	GWh	114.468	30.298	1.303	58.359	18.884	3.515	2.109
Heat	TJ	100.517	2.769	11.217	66.015	14.067	0	6.449
Difference								
Electricity	GWh	-796	-415	-11	-823	-235	0	688
Heat	TJ	7.278	-19	214	3.629	386	0	3.068

Apart from the energy balance also electricity and heat production is a relevant variable from (summary) energy statistics. We also compared the data from the miniquestionniares with joint annual questionnaires for this variable (table 2).

For electricity the differences are typically about 1 percent, for heat the differences are larger. A main reason for this is that electricity is part of a monthly statistics tailored to timely cover the most important sources, whereas for heat annual surveys are used whose analysis is not complete yet in May. A second reason is that observing heat is more complex than electricity, because the definition of heat flows measured by companies is often not the same as definition of heat flows required for energy statistics. For policy purposes the data on renewable heat are especially important and for these data the difference was 3 percent.

This task resulted in milestone 5 “Examine the differences between early estimates and final data send later”.

7. Conclusions and recommendations

We were able to send the completely filled five mini-questionnaires before the deadline, both for the reference year 2018 and 2019.

We derived a summary energy balance from the mini-questionnaires for 2018 and compared this balance with the balance published by Eurostat and derived from the joint annual questionnaires send to Eurostat in the end of 2019. We believe that differences are acceptable and think that the provisional summary energy balance would be accurate enough to be useful for policy makers.

In particular data deviations for data related to calculating the share of renewable energy data were limited. For national purposes we calculated and published the share of renewable energy for 2018 from the same underlying data (Statistics Netherlands, 2019) and this share (7.4 percent) was close to the share reported via SHARES in December 2020 (Eurostat, 2019) (7.3 percent).

We adapted the existing statistical process to have early data for all main elements of the energy balance. The type of adaptation was different for different subprocesses. For the companies in industry with a complex balance we worked on a new data editing tool and for the companies in industry with a simple balance we adapted the sample design to have more robust raising in case of low response which may be the case for early estimates. For the other sectors we designed a method to arrive early estimates from the client files from the grid-operators. For the technical phase of actually filling the questionnaires we adapted our output-tool to automatically produce the desired numbers from the central database. All these adaptations proved to work, except for the data editing tool which was not ready in time. However, we expect to be able to use the tool the next cycle.

Experiences from the two cycles learned that it is even more important to thoroughly organize and plan our regular work, because little delay is possible. Delays of two or three weeks in the statistical process half a year before the May deadline (preparing surveys, contacting admin data owners) already may affect quality of the early estimates.

At the end time for solving data issues (which will always be there) could be gained if we would have to spend less time on data logistics for both the IEA mini-questionnaires (with a deadline in the same period as we work on the Eurostat questionnaire) and Eurostat mini-questionnaires. Therefore, we would welcome joint Eurostat-IEA mini-questionnaires with a facility to automatically import data instead of the current copy paste method.

We observe that there is also a national demand for early estimates for energy balances. Main applications are use for early estimates of greenhouse gas emissions (both for national and international reporting), input for the models of the annual National Energy Outlook and public debate on the progress regarding the share of renewable energy. For reasons of clarity for users and efficiency at our side it is desirable to have a fixed, limited number of release moments for annual energy statistics. Logically, these moments are internationally harmonised.

For the final data there is already a well established moment of publication: December, based on the same data as send to Eurostat/IEA at the end of November and also used as input for

official reporting of greenhouse gas emissions. We think it would be good to try achieve a second harmonised moment for publications of early estimates of energy statistics and balances before summer. For greenhouse gas emissions member states already have to report early (T-1) estimates before 31th July (EU 2013). These estimates contain a sectoral break down that more or less corresponds to the sectoral break down of the miniquestionnaires, implying that comparable data are needed.

The purpose of the Eurostat mini-questionnaires is to calculate a summary energy balances and also a calculation of the share of renewable energy is a user need. To facilitate these type of calculations Eurostat may develop a tool like SHARES or the Energybalance builder that imports mini-questionnaires and automatically calculates an energy balance and the share of renewable energy.

8. References

EU (2013) Regulation 525/2013 on monitoring and reporting greenhouse gas emissions and for reporting other information at national and Union level relevant to climate change

Eurostat (2019) SHARES summary results 2018

Eurostat (2020) Energy Balances in the MS Excel file format (2020 edition)

Statistics Netherlands (2019) Share of renewable energy up to 7.4 percent, news release, May 2019.