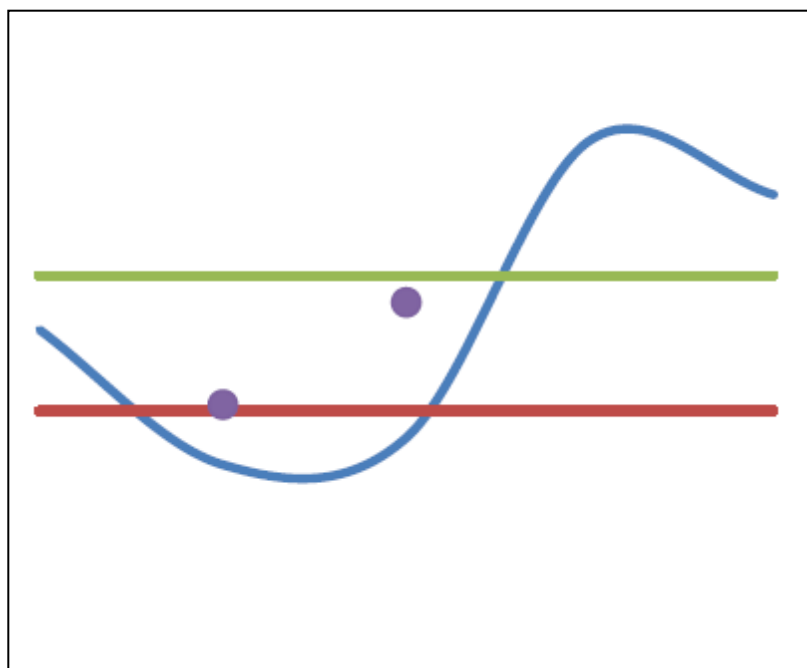


# Methodologies for identifying incomplete reporting of E-PRTR emission data with releases to air



**ETC/ACM Technical Paper 2014/10**  
**December 2014**

*Stephan Poupa, Katarina Mareckova, Katrin Seuss,  
Wim van der Maas, Jeroen Kuenen, Rianne Dröge, Chris Dore*



The European Topic Centre on Air Pollution and Climate Change Mitigation (ETC/ACM) is a consortium of European institutes under contract of the European Environment Agency  
RIVM Aether CHMI CSIC EMISIA INERIS NILU ÖKO-Institut ÖKO-Recherche PBL UAB UBA-V VITO 4Sfera

**Front page picture:**

A snapshot of an schematic illustration of actual emissions compared with calculated emissions, including a ration uncertainty (Figure 2 of this paper).

**Author affiliation:**

Stephan Poupa, Katarina Mareckova, Katrin Seuss (UBA-V),  
Wim van der Maas (RIVM),  
Jeroen Kuenen, Rianne Dröge (TNO),  
Chris Dore (Aether)

**DISCLAIMER**

This ETC/ACM Technical Paper has not been subjected to European Environment Agency (EEA) member country review. It does not represent the formal views of the EEA.

© ETC/ACM, 2015.

ETC/ACM Technical Paper 2014/10

European Topic Centre on Air Pollution and Climate Change Mitigation

PO Box 1

3720 BA Bilthoven

The Netherlands

Phone +31 30 2748562

Fax +31 30 2744433

Email [etcacm@rivm.nl](mailto:etcacm@rivm.nl)

Website <http://acm.eionet.europa.eu/>

## CONTENT

1	Introduction.....	5
2	General approach and limitations.....	7
3	Facility checks across time series .....	8
3.1	Approach .....	8
3.1.1	Cross pollutant completeness checks across the time series .....	8
3.1.2	Sector completeness checks.....	12
3.1.3	Cross pollutant ratio checks .....	13
3.1.4	Time series consistency checks .....	14
3.2	Conclusions / next steps.....	15
4	Cross pollutant analysis for a particular year .....	16
4.1	Approach .....	16
4.1.1	List of possible outliers or missing values from the cross pollutant test .....	16
4.1.2	List of possible outliers or missing values from the cross pollutant test with additional conditions .....	17
4.1.3	Quantitative relations.....	18
4.2	Conclusions / next steps.....	19
5	Assessment of missing facilities, in particular for some rarely reported pollutants .....	20
5.1	Approach .....	20
5.1.1	Typical results .....	22
5.1.2	Assessment of rarely reported pollutants.....	27
5.2	Conclusions.....	29
6	Using economic and employment statistics to review E-PRTR emissions data .....	32
6.1	Approach .....	32
6.1.1	Input Data .....	33
6.1.2	Methodology .....	34
6.1.3	Initial investigation of data tests .....	35
6.1.4	Emissions to air .....	36
6.1.5	Emissions to water.....	40
6.1.6	Consistency of the relationship across the time series .....	42
6.1.7	Conclusions on the analysis of 2011 emissions data.....	42
6.1.8	Methodology for the analysis of 2012 emissions data.....	43
6.2	Conclusions / next steps.....	45
	References.....	47
	Abbreviations .....	48
Annex 1	Manual applying cross pollutant checks on national and European databases	49
	Checking a national E-PRTR dataset, including the E-PRTR conversion tool.....	49
	Checking the European E-PRTR data base.....	52

Data files coming with this paper:

*Cross\_pollutant\_Check.xls*

*Cross\_Pollutant\_Check\_for\_European\_data\_set.mdb*

*EPRTRConversionTool.mdb*

*2012\_E-PRTR\_Data\_Review\_Aether\_issue1.xls*

*Cross\_Pollutant\_Check\_for\_Conversion\_tool.mdb*



## 1 Introduction

According to Regulation (EC) No 166/2006 concerning the establishment of a European Pollutant Release and Transfer Register (EU, 2006a), operators that undertake one or more activities specified in Annex I of the E-PRTR Regulation above the capacity threshold have to report their releases to air, water, land, off-site transfers of waste and releases of pollutants in waste water if these releases and transfers exceed the threshold values specified in Annex II of the Regulation. The Member States are obliged to submit this data to the European Commission annually since 2009 (covering 2007 data for the first time) following the implementation guidance (EU, 2006b). The E-PRTR currently contains data from 32 countries (EU-27 plus Norway, Iceland, Liechtenstein, Serbia and Switzerland) from more than 30,000 industrial facilities in 65 economic activities for up to 91 pollutants.

### Informal annual E-PRTR review

The EEA has carried out an informal review of E-PRTR data annually since 2009. Different European Topic Centres have been involved in carrying out the review (ETC/ACM, ETC/SCP, ETC/W). The informal review consists of different checks. The initial checks concentrate on checking the internal consistency of the reported E-PRTR data while the focus of the extended checks is to check consistency with data reported under other reporting obligations. The tests used during the informal review, which are based on the comparison of the submitted PRTR data with the data of the previous reporting round, have proven to be quite effective in identifying outliers and/or misreporting. However, these checks were not suitable for identifying missing facilities, pollutants and activities.

The informal review carried out in the year 2014 showed that out of the 31,405 E-PRTR facilities which reported for the year 2012 about 17,000 reported for all six reporting years (see Figure 1). This number will probably slightly increase after the resubmission of 2012 data because several countries did not succeed in reporting a complete data set for 2012 in March 2014. Nevertheless, Figure 1 reveals significant inconsistencies in the reporting of E-PRTR data by facilities across the years even if the possibility is taken into account that releases/transfers could have fallen below the reporting threshold in certain years.

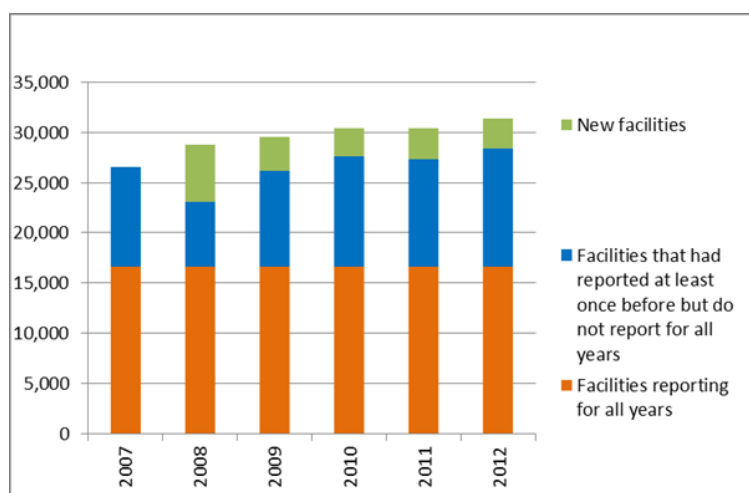


Figure 1 Number of facilities in E-PRTR and reporting consistency across the years

## **Triennial E-PRTR review**

In the year 2011, the Commission initiated a triennial review of E-PRTR data reported under the E-PRTR Regulation in order to assess the completeness and representativeness of E-PRTR data for the reporting years 2007 to 2009. At the same time, an analysis of the scope of E-PRTR was performed to evaluate whether the target has been fulfilled that 90% of the releases/transfers of Annex I facilities are covered by the Regulation. All findings of the review were presented in the report *“Three years of implementation of the E-PRTR”* (Umweltbundesamt et al 2012).

The results of the triennial review showed that reporting in 2011 had improved compared to 2009. All E-PRTR countries reported satisfactory data sets, in particular for air. The statistical analysis of the data sets showed that the reporting of 36 standard pollutants reached 90% of the total mass and that the reporting of other pollutants showed significant variability. The quality of the data on releases to water was considered acceptable, but 27 pollutants were reported by less than ten facilities. Reporting of transfers to water was considered incomplete. Regarding waste, E-PRTR covers only waste transfers (hazardous and non-hazardous) from major individual facilities. The statistical analysis showed that the waste thresholds allowed reporting of only about 39% of hazardous waste and 17% of non-hazardous waste. Reporting of releases to land was considered non-satisfying.

To conclude, both the formal and the triennial reviews of E-PRTR data have indicated that there are still data gaps in E-PRTR data in terms of missing facilities, pollutants and activities. Therefore, further work to identify these data gaps is of high priority for the European Commission and the EEA. In 2014, the ETC/ACM developed methodologies to identify incompleteness of E-PRTR data at the facility level. This methodology report describes the developed tests in detail.

## 2 General approach and limitations

The methods presented in this report are based on the experience gained through the informal reviews 2009-2013 and through the triennial review of E-PRTR data in 2011. They are designed to identify reporting gaps and incompleteness of reporting at facility level, in particular missing facilities and/or missing reports<sup>1</sup>.

Potential methods for the indication of incompleteness of E-PRTR data comprise:

- Facility checks across the time series
- Cross pollutant analysis (for a particular year)
- Comparison with other datasets (e.g. on large combustion plants (LCPs), statistical data, data on urban wastewater treatment plants)
- Expert judgment
- Other

The following methods have been implemented so far and are described in more detail in this methodology report:

- Facility checks across the time series
- Cross pollutant analysis (for a particular year)
- Comparison with other datasets

The methods developed so far cover only releases to **air**. The checks might be extended to cover other media in the future or new checks might be designed for other media. In terms of activities the focus of the checks has been put on LCPs and the chemical industry. In terms of pollutants the following air pollutants have been considered as priority pollutants:

- Pollutants reported under the National Emissions Ceiling Directive (NECD):
  - NO<sub>x</sub>, NMVOC, SO<sub>x</sub>
- PM<sub>10</sub>
- CO<sub>2</sub>

Heavy metals, PCDD/PCDF and HCl were considered as second priority pollutants. In addition, the reporting of rarely reported pollutants was analysed.

It is important to stress that all the performed tests can only indicate **potentially** missing releases of certain pollutants or missing facilities in certain years. The tests will always be limited due to the absence of activity data or any technology information of the facility other than E-PRTR activity or NACE codes. The test results have to be further analysed by country experts with more in-depth knowledge on the facilities who can then either confirm whether releases/facilities are really missing or whether there is a different explanation for missing releases/facilities.

---

<sup>1</sup> The facility is included in the E-PRTR database but not all pollutants were reported or releases/transfers were not reported for all years.

## 3 Facility checks across time series

### 3.1 Approach

The methodology developed by UBA-V aims at identifying facilities which do not report in particular years and/or do not report all pollutants in particular years. Previous work that the ETC/ACM has performed on the E-PRTR dataset indicated that a number of countries have not updated their 2007 datasets. Therefore, 2007 E-PRTR data was excluded from the time series analysis and the tests were carried out for 2008-2011 E-PRTR data.

Four different types of checks have been developed:

- **Cross pollutant completeness checks across the time series**  
This check verifies whether the reporting of two selected pollutants that were reported by an E-PRTR facility is consistent over the full time series.
- **Sector completeness checks**  
This check verifies whether E-PRTR facilities of a typical size with specific main activities and/or NACE codes (specific technology) have reported the expected pollutants for the different years.
- **Time series consistency checks**  
This test detects inconsistencies in the time series of a release/transfer of any pollutant into a defined medium, which might indicate incomplete reporting. This test targets large emitters.
- **Cross pollutant ratio check**  
This test checks whether the calculated quantity ratio of two defined pollutants exceeds a predefined threshold for the different years.

First, all the checks listed above are performed at the facility level. Then, the output of the checks is stored in a table together with a flag that indicates whether the check has been passed or failed and a message text in case the test has failed.

It has to be noted that although the results of the checks may also allow the detection of outliers, the detection of outliers is not the objective of the developed checks because outliers should already be detected during the informal E-PRTR review and communicated to countries.

#### 3.1.1 Cross pollutant completeness checks across the time series

This test checks whether the reporting of two pollutants by a facility is consistent across the time series. In so far this test differs from the cross pollutant analysis described in chapter 4 which carries out cross pollutant checks for one single year. Pollutant 1 defines a 'indicator pollutant' and Pollutant 2 defines the pollutant which is tested (i.e. checked). 'Indicator pollutant' refers to the pollutant that has been selected in order to test for the reporting of another pollutant. It is assumed that there is a relation between the indicator and the test pollutant, e.g. the CO<sub>2</sub> emission of fuel combustion in a power plant is the indicator pollutant and the related test pollutant is the SO<sub>x</sub> emission that co-exists with the CO<sub>2</sub> emission. It has



to be stressed that there might be exceptions to this assumed relation, e.g. in case only natural gas is used as fuel there will be no SO<sub>x</sub> releases. In this case, the test will nevertheless flag that SO<sub>x</sub> has not been reported. This is why the test results have to be further analysed by country experts who have the detailed knowledge on the facilities.

The check is performed only for those facilities with releases of the indicator pollutant above a defined threshold for at least one of the tested years (e.g. x times the Annex II threshold). The threshold for the inclusion of facilities into the test is based on expert judgement.

### Algorithm

- p1** indicator pollutant release
- p2** test pollutant release
- RU** Ratio Uncertainty (e.g. 50% or 100%, based on expert judgement)
- y** reporting year

For each reporting year where both pollutants, p1 and p2, are reported we calculate the:

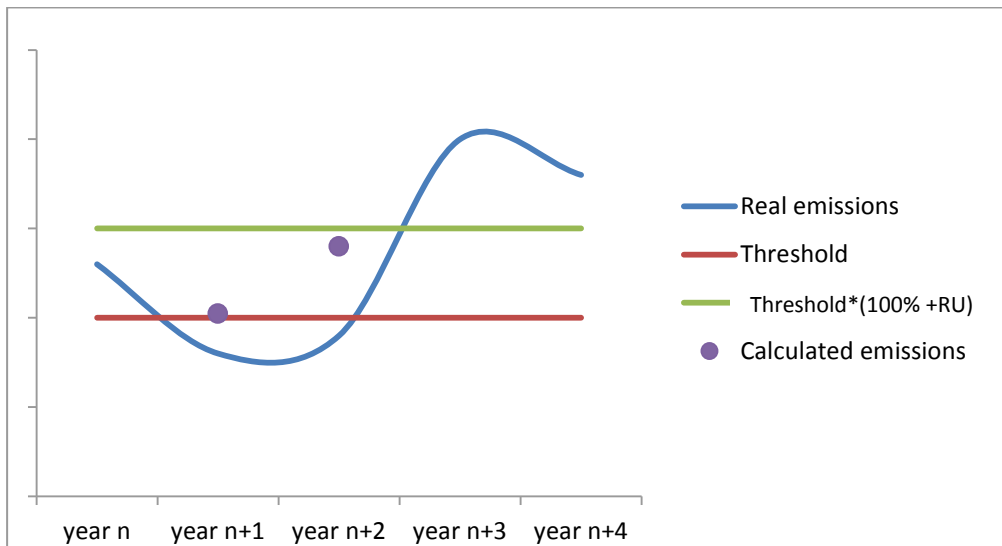
- ratio:  **$R_y = p1_y/p2_y$**
- average ratio:  **$R_{avg} = \text{average of } R_y \text{ of the tested years}$**

For each year where p1 is reported, but p2 is not reported we calculate:

- potential emission of p2:  **$emi_{pot} = R_{avg} * p1$**
- If  **$emi_{pot} > \text{threshold}(p2) * (100\% + RU)$** , then the test flags a result

The ratio uncertainty (RU) is used to account for the potential uncertainty of the p1/p2 ratio for two given pollutants, expressed as an absolute value in relation to the threshold. The RU has to be defined on the basis of expert judgements and is indicated in Table 2.

Figure 2 illustrates the application of the RU. In the example, the real emissions are close to the threshold for all years and below the threshold for the years n+1 and n+2 and therefore they were not reported for these two years. The calculated emissions for years n+1 and n+2 lie above the threshold, however, not above the threshold including its additional RU and as such leading to none flagging. The use of a RU therefore avoids that the test provides unrealistic flagging of false incompleteness test results.



**Figure 2: Real emissions compared with calculated emissions and the role of Ratio Uncertainty (RU)**

If the cross pollutant check detects incompleteness finding for one of the tested years the check delivers a message according Table 2.

**Table 1 Test messages for cross-pollutant checks**

Test result	Message
If p2 is missing in all tested years	"P2 is not reported"
If p2 is not reported in some of the tested years	"P2 missing pot.quantity: year(s)"
If p1 is missing in some of the tested years	"P1 missing: year(s)"

Based on the methodology explained above, Table 2 provides the implemented cross pollutant checking criteria used for flagging of possible incompleteness.

**Table 2 List of implemented cross pollutant checks**

ID	Medium	Test name	Test description
<b>A</b>	Air	CO <sub>2</sub> /NO <sub>x</sub> completeness for large CO <sub>2</sub> emitters	<p>It is expected that large CO<sub>2</sub> emitters (&gt;2 mio. t CO<sub>2</sub> = 20 x threshold in at least one of the tested years) should also report NO<sub>x</sub>. The test returns a message if either no NO<sub>x</sub> releases have been reported for a given year(s) or the time series for CO<sub>2</sub> releases is incomplete.</p> <p>If NO<sub>x</sub> is reported in any year(s) but not for all years the average NO<sub>x</sub>/CO<sub>2</sub> ratio is calculated and the NO<sub>x</sub> releases for the missing year(s) are estimated by means of the ratio. If the estimated NO<sub>x</sub> releases are higher than 2 x the NO<sub>x</sub> threshold (&gt;200 t) the test flags a potential problem for the respective year.</p>
<b>B</b>	Air	CO <sub>2</sub> /SO <sub>x</sub> completeness for large CO <sub>2</sub> emitters	<p>It is expected that large CO<sub>2</sub> emitters (&gt;5 mio. t CO<sub>2</sub> = 50 x threshold in at least one of the tested years) should also report SO<sub>x</sub> releases. The test returns a message if either no SO<sub>x</sub> releases have been reported for a given year(s) or the time series for CO<sub>2</sub> releases is incomplete.</p> <p>If SO<sub>x</sub> is reported in any year(s) but not for all years the average SO<sub>x</sub>/CO<sub>2</sub> ratio is calculated and the SO<sub>x</sub> releases for the missing year(s) are estimated by means of the ratio. If the estimated SO<sub>x</sub> releases are higher than 2 x the SO<sub>x</sub> threshold (&gt;300 t) the test flags a potential problem for the respective year.</p>
<b>C</b>	Air	SO <sub>2</sub> /Hg completeness for large SO <sub>2</sub> emitters	<p>It is expected that large SO<sub>2</sub> emitters (&gt;3 000 t of SO<sub>2</sub> = 20 x threshold in at least one of the tested years) should also report Hg. The test returns a message if either no Hg has been reported for a given year(s) or the time series for SO<sub>2</sub> releases is incomplete.</p> <p>The same principle as in Test A and B applies. The threshold for missing Hg releases is 20kg (2 x threshold).</p>
<b>D</b>	Air	SO <sub>2</sub> /NO <sub>x</sub> completeness for large SO <sub>2</sub> emitters	<p>It is expected that large SO<sub>2</sub> emitters (&gt;3 000 t of SO<sub>2</sub> = 20 x threshold in at least one of the tested years) should also report NO<sub>x</sub>. The test returns a message if either no NO<sub>x</sub> releases have been reported or the time series for SO<sub>2</sub> is incomplete.</p> <p>The same principle as in Test A and B applies. The threshold for missing NO<sub>x</sub> = 100 t (2 x threshold).</p>

<b>E</b>	Air	SO <sub>2</sub> /PM <sub>10</sub> completeness for large SO <sub>2</sub> emitters	It is expected that large SO <sub>2</sub> emitters (>3 000 t of SO <sub>2</sub> = 20 x threshold in at least one of the tested years) should also report PM <sub>10</sub> . The test returns a message if either no PM <sub>10</sub> releases have been reported or the time series for SO <sub>2</sub> is incomplete.  The same principle as in Test A and B applies. The threshold for missing PM10 is 75 t (1.5 x threshold).
<b>F</b>	Air	SO <sub>2</sub> /Dioxin completeness for large SO <sub>2</sub> emitters	It is expected that large SO <sub>2</sub> emitters (>3 000 t of SO <sub>2</sub> = 20 x threshold) should also report Dioxin. The test returns a message if either no Dioxin releases have been reported or the time series for SO <sub>2</sub> is incomplete.  The same principle as in Test A and B applies. The threshold for missing Dioxin is 0.15 g (1.5 x threshold).

The list of the tests (combination of pollutants) can be extended as needed. However, as 91 pollutants are listed in E-PRTR Annex II prioritising is essential because the output of the tests should finally be checked by country experts without becoming that cumbersome.

### 3.1.2 Sector completeness checks

For specific main activities and NACE codes a certain technology and typical plant size are assumed by expert judgment and it is expected that these plants emit certain pollutants. The test flags all facilities that have potential gaps in reporting based on these specific assumptions.

The checks deliver flagging messages at possible incompleteness as given in Table 3.

**Table 3 Test messages for sector incompleteness check**

Test result	Message
If pollutant is missing in all tested years	"Pollutant x is not reported"
If pollutant is missing in some of the tested years	"Pollutant x missing: year(s) y <sub>1</sub> , y <sub>2</sub> ..."

Based on the methodology explained above the sector incompleteness checks of Table 4 have been implemented.

**Table 4 List of implemented sector incompleteness checks**

ID	Medium	Test name	Test description
G	Air	SO <sub>x</sub> from refineries main activity 1.(a)	The test identifies any incomplete reporting of SO <sub>x</sub> from oil refineries.
H	Air	NO <sub>x</sub> for cement plants: NACE 23.51 -Manufacture of cement	The test identifies any incomplete reporting of NO <sub>x</sub> from cement plants.

### 3.1.3 Cross pollutant ratio checks

This check tests whether the calculated quantity ratio of two defined pollutants exceeds a predefined threshold, e.g. whether the ratio of NO<sub>x</sub> and CO<sub>2</sub> air releases exceeds a certain ratio. The assumed ratios are based on expert judgement. This test targets fuel or waste combustion plants of any size and Table 5 provides its test parameters.

**Table 5: Test parameters of cross pollutant ratio check**

Parameter	Description
Medium	Medium: air, water, soil
Nominator pollutant	Test pollutant 1 (p1)
Denominator pollutant	Test pollutant 2 (p2)
Maximum ratio threshold	The maximum tolerated ratio calculated as quantity of nominator pollutant/quantity of denominator pollutant (p1/p2)

The check delivers flagging messages at possible incompleteness as given in Table 6.

**Table 6: Test message for cross pollutant ratio check**

Test result	Message
If the ratio of the nominator pollutant and the denominator pollutant (p1/p2) exceeds the maximum ratio threshold	“Ratio exceeds with a max. in year x (y times higher than the max expected ratio)”

Based on the methodology explained above, Table 7 gives the implemented cross pollutant ratio check.

**Table 7 Implemented cross pollutant ratio check**

ID	Medium	Test name	Test description
I	Air	NO <sub>x</sub> /CO <sub>2</sub>	Check if the ratio of NO <sub>x</sub> /CO <sub>2</sub> exceeds 1/200.

### 3.1.4 Time series consistency checks

This check aims at detecting inconsistencies in the time series of a release/transfer of any pollutant into a defined medium, which might indicate incomplete reporting for (a) specific year(s), compared to other years. This test targets large emitters because only facilities whose lowest reported value in the time series of any reported pollutant is above a defined threshold (i.e. the minimum threshold multiple) have been included in the test. If the ratio of the maximum reported quantity in relation to the minimum reported quantity (i.e. the ratio threshold) exceeds a defined threshold the facility is flagged for possible reporting inconsistency. Both the minimum threshold multiple and the ratio threshold have been defined on the basis of expert judgement. The test is performed for all reported pollutants and according the parameters of Table 8.

**Table 8: Test parameters of time series consistency test**

Parameter	Description	Assumed value (expert judgement)
Medium	Medium: air, water, soil	
Minimum threshold multiple	Indicates the multiple of the Annex II pollutant threshold that the lowest reported quantity of a pollutant has to exceed in any of the tested years so that the facility is included in the test.	20
Ratio threshold	Maximum tolerated ratio of highest quantity in relation to lowest quantity reported. If this ratio is exceeded the facility is flagged.	10

The check delivers flagging messages at detected possible inconsistency as given in Table 9.

**Table 9: Test messages for the time series consistency checks**

Test result	Message
If pollutant(s) is/are missing in any of the tested years	“Failed for: pollutant(s) x; missing year(s) y <sub>1</sub> , y <sub>2</sub> ...”
If the ratio threshold is exceeded	“Failed for: pollutant(s) x <sub>1</sub> , x <sub>2</sub> ...”

Based on the methodology explained above, Table 10 gives the implemented time series consistency check.

**Table 10: Implemented time series consistency check**

ID	Medium	Test name	Test description
J	Air	Time series consistency	<p>List all facilities/pollutants which show gaps in reporting or high outliers. All pollutants are tested for which a threshold multiple &gt; x 20 has been reported in any year.</p> <p>Facilities are flagged when a ratio of the maximum value/minimum value of tested pollutants &gt; 10 or when any year in the time series of a tested pollutant is missing.</p>

Table 11 presents an example of test results for 2012 data. The column ‘Total’ includes the total number of facilities that were checked for that year. The columns ‘Positive’ and ‘Negative’ include the number of facilities with a positive (i.e. consistency detected) and a negative (i.e. possible inconsistency detected) test result. Their addition equals the total. The percentage of the total tested facilities with negative tests results is given in the column ‘Negative %’. The column ‘With missing year(s)’ includes the number of facilities which did not report all tested pollutants for the complete time series. The column ‘With high variance’ indicates that the highest reported quantity is more than ten times larger than the lowest reported quantity. The sum of these two possible incompleteness causes equals the negative number of facilities.

**Table 11: Example with summary test results of a time series consistency check**

Test	Min. threshold multiple	Ratio threshold	Number of facilities					
			Total	Positive (consistent)	Negative (possibly inconsistent)	Negative %	With missing year(s)	With high variance
Releases into air	20	10	2 289	785	1 504	66%	1 300	204

### 3.2 Conclusions / next steps

The implemented tests have proven to be reliable for the identification of potential non-reporting if the facility has reported at least once since 2009. However, since the test results can only flag *potential* non-reporting the test results always have to be checked by national experts with knowledge on the facilities. To identify facilities which have not been reported under E-PRTR in any year other methods have to be used, in particular the comparison with other data in combination with expert judgement.

The tests for the medium air could be further elaborated in the future. Additional tests for air could be added for additional pollutants and the parameter used in the present tests could be revised based on the experience gained. Furthermore, the checks could be extended to include tests for other media, e.g. water.

## 4 Cross pollutant analysis for a particular year

This method was developed under the ETC/ACM subvention 2011 for task 1.1.1.3 (E-PRTR data review). It describes and underpins the methodology of a cross pollutant check for industrial emissions reported by E-PRTR facilities for *a particular year*. In so far, the analysis differs from the cross pollutant completeness check described in chapter 3.1.1, which focuses on the cross pollutant completeness across the time series. The check runs against the E-PRTR MS-Access database that countries can use to create their delivery to E-PRTR and it runs against the database downloadable from the EEA to check all the countries.

### 4.1 Approach

The purpose of this cross pollutant test is to list possible outliers or missing values based on the emission of *another* pollutant, taking the EC Regulation 166/2006 Annex I *activity* and in some cases the main NACE code (economical *sector*) into account. For example facilities with the Annex I activity ‘Thermal power stations and other combustion installations’ are expected to report releases of CO<sub>2</sub> together with releases of NO<sub>x</sub>. Given a certain quantity of reported indicator pollutant CO<sub>2</sub> (the *source* pollutant) there is a minimum and maximum amount of NO<sub>x</sub> expected (the *resulting* pollutant). Emissions of the resulting pollutant outside this range or missing values are flagged by the test. In general, the emission factors used are the maximum and minimum values found for the given activity in the EMEP/EEA Guidebook (EMEP 2009) and IPCC Guidelines (IPCC 2006). In total, 174 cross pollutant relations have been integrated in the test. Parts of the tested relations are the same but have been tested for different activity codes.

#### 4.1.1 List of possible outliers or missing values from the cross pollutant test

The cross pollutant emission check has three steps:

1. Select the emission that could be checked for each facility from the E-PRTR database
2. Calculate for each selected emission of the source pollutant the resulting likely emission range of the resulting pollutant
3. Decide to give a warning when the resulting emission is missing or out of range, taking into account the reporting threshold and possible accidental release of the resulting pollutant

The cross pollutant check results in a list of possible outliers or missing values, of which an example is shown in Figure 3. It includes for each facility the emission for a certain *source* pollutant (here ‘Source Pollutant Name’ and ‘Source Emission’) and for a certain resulting pollutant (‘Resulting Pollutant Name’ and ‘Reported Resulting Emission’). The value for the *resulting* pollutant is compared to the expected range of emission for the resulting pollutant (‘Min Resulting Emission Calc’ and ‘Max Resulting Emission Calc’) and flagged when it is out of range or missing. Furthermore, it is indicated why an emission is flagged, e.g. the reported emission is lower than the expected minimum or missing, or the reported emission is higher than the expected maximum.



Medium	Air										
Country	x										
Year	x										
FacilityID	FacilityName	AnnexI ActivityCode	NaceMainEconomic ActivityCode	Source PollutantName	Source Emission	Resulting PollutantName	Reported ResultingEmission	MinResulting EmissionCalc	MaxResulting EmissionCalc	ReasonFor Flagging	
12345	Abcd	1.(a)	12.34	Abc	12345	Def	12345	123	45678	Below minimum expected or missing	
										Above maximum expected	

Figure 3 Example of the list of possible outliers

#### 4.1.2 List of possible outliers or missing values from the cross pollutant test with additional conditions

This test is comparable with the test above but adds an additional condition of another pollutant, the *conditional* pollutant. For example, where the source pollutant CO<sub>2</sub> stands for a certain amount of energy being used, the accompanying PM<sub>10</sub> indicates the type of fuel: natural gas (with little or no PM<sub>10</sub>) or liquid and solid fuel (with substantial amounts of PM<sub>10</sub>). This gives an opportunity to fine-tune the emission factors of the resulting pollutant.

The cross pollutant check with additional conditions also results in a list of possible outliers or missing values, of which an *example* is shown in Figure 4. It includes the same as the list in the figure above, but in addition it describes the conditional pollutant and its reported emission ('Conditional Pollutant Name' and 'Reported Pollutant Emissions').

Medium	Air											
Country	x											
Year	x											
FacilityID	FacilityNm	AnnexI ActivityCd	NaceMainEconomic ActivityCd	Source PollutantNm	Source Emission	Conditional PollutantNm	Reported PollutantEmission	Resulting PollutantName	Reported ResultingEmission	MinResulting EmissionCalc	MaxResulting EmissionCalc	ReasonFor Flagging
12345	Abcd	1.(a)	12.34	Abc	12345	Ghi	1234	Def	12345	123	45678	Below minimum expected or missing
												Above maximum expected

Figure 4 Example of the list of possible outliers from the cross pollutant check with additional conditions

A complete list of the derived relations is included in the MS-Excel table "Cross\_Pollutant\_Check.xls", which is provided separately with this methodology report. The relationships have only been derived for releases to air and most attention has been paid to the largest sectors and the most important pollutants (see Chapter 2). A MS-Access database version of this dataset including the queries to run the test and the results is also made available ("Cross\_Pollutant\_Check\_for\_Conversion\_tool.mdb" for national data checking and "Cross\_Pollutant\_Check\_for\_European\_data\_set.mdb" for checking the European dataset).

See Annex 1 for some instruction on how to apply the "Cross\_Pollutant\_Check.xls" and the "Cross Pollutant Check for Conversion tool.mdb" for national data checking and "Cross Pollutant Check for European data set.mdb" for checking the European dataset.

### 4.1.3 Quantitative relations

The *quantitative* relations are provided in a cross pollutant table (also included in MS-Excel workbook "Cross\_Pollutant\_Check.xls" of section 4.1.2). The content of the cross pollutant table provides the basis for the methodology with almost 175 relations between a reported pollutant (the *source* pollutant) and an expected *resulting* pollutant, given the Annex I activity and in some cases the NACE sector within this Annex I activity. Optionally, there can be a *conditional* pollutant, which narrows down the expected range of the resulting pollutant. Table 12 presents the format of the spreadsheet.

**Table 12 Overview of information in the cross pollutant table**

Heading	Description	Example
AnnexIActivityCode	Activity code	1.(a)
AnnexIActivityName	Activity description	Mineral oil and gas refineries
NACEMainEconomicActivityCode	NACE code	
NACEMainEconomicActivityName	NACE description	
ReleaseMediumCode	Emissions to a compartment (water or air)	AIR
SourcePollutant	Pollutant initiating the rule	Carbon dioxide (CO <sub>2</sub> )
ResultingPollutant	Pollutant whose value is checked	Sulphur oxides (SO <sub>x</sub> /SO <sub>2</sub> )
MinFactor	Likely minimum factor (resulting pollutant emission divided by source pollutant emission)	0.194%
MaxFactor	Likely maximum factor (resulting pollutant emission divided by source pollutant emission)	5.26%
ConditionPollutant	Pollutant whose value is a condition for check	Particulate matter (PM <sub>10</sub> )
ConditionValue	Likely minimum factor for the condition	0.00625%
ConditionSourcePollutant	Pollutant initiating the condition for check	Carbon dioxide (CO <sub>2</sub> )
ReportingThreshold	Reporting threshold for the resulting pollutant (kg/year)	150 000 kg/year
Reference	Reference for factors	EMEP/EEA Guidebook, 2009 and IPCC Guidelines 2006. Min factor for "heavy fuel oil", Max factor for "hard coal", Condition for "derived gases"

There are two main constraints in deriving the quantitative relations and in performing a cross pollutant check: the limited amount of available information and the reporting thresholds.

- a) The limited amount of information is a constraint in deriving the quantitative relations. The only data available are the reported emissions (only above reporting threshold) and the *AnnexIActivity* and the *NACEMainEconomicActivity*. Other information might be useful in describing the relations, like the type of process, the fuel type, the production or capacity and the fuel consumption and use.

- b) The reporting thresholds limit the amount of reported emissions (compared to the total emissions) and the number of pollutants reported. In particular, smaller facilities often report only one pollutant. In that case, a cross pollutant check is impossible.

## 4.2 Conclusions / next steps

The cross pollutant check can identify outliers in emissions (compared to other pollutants) but it can also identify possible reporting gaps. Possible reporting gaps are identified when the reporting of a pollutant is expected (based on the reporting of other pollutants), while the facility did not report it. Therefore, it is also useful for checking the completeness of reporting. The check can be easily expanded and takes little time to run.

For most pollutants, only a few percent of the facilities report emissions that are not within the expected range. Most of the facilities with (one or multiple) emissions out of the expected range can be found in the sector 1.c.

As mentioned before, the releases that were flagged by this check are not necessarily wrong. Nevertheless, it needs to be checked whether specific circumstances apply for the facilities with flagged releases, which will explain the deviation from the expected emission range. These explanations could and should also be used to improve the Cross Pollutant Check.

## 5 Assessment of missing facilities, in particular for some rarely reported pollutants

This chapter describes a simple test conducted by RIVM/TNO to assess possible incompleteness of E-PRTR by comparing reported emissions with reported production statistics. The purpose of this comparison is to identify the NACE sectors<sup>2</sup> for which production statistics are available for a certain country (from the Eurostat Prodcom database), but the European E-PRTR dataset does not contain any facility reports for that country. It is an assessment of the European E-PRTR completeness that has not been performed earlier.

The work has been triggered especially because for some rarely reported emissions there is very low consistency in reporting emissions in the European E-PRTR dataset (as identified by Umweltbundesamt et al 2012). Some specific checks for these pollutants are included at the end of this chapter, however the methodology presented here can be applied for all pollutants.

The rarely reported pollutants checked for here, are listed in Table 13.

**Table 13** Pollutants to air that are rarely reported under E-PRTR

Pollutant	Activity	Threshold (release to air)
1,1,1-trichloroethane	4a, 4b, 4c, 5c, 9c and 9e	100 kg/year
1,1,2,2-tetrachloroethane	2a, 2b, 4a, 4b, 4c, 4d, 4e and 5a	50 kg/year
Asbestos	3d	1 kg/year
Halons	2c, 2d, 2e, 2f, 4a, 4b and 5a	1 kg/year
Hexachlorobenzene (HCB)	2c, 2e, 2f, 4a, 4b, 4c, 5a, 5b, 5c, 5d, 5f and 9c	10 kg/year
Trichlorobenzenes (TCBs) (all isomers)	4a, 4b, 4c, 5a, 5g and 9c	10 kg/year
Anthracene	1d, 2b, 2d, 3c, 4a, 4d and 6c	50 kg/year

### 5.1 Approach

In order to make a completeness assessment the following information has been collected to perform the comparison:

- *Production* statistics (Prodcom) per NACE Rev.2 sector per country for the year 2011 from Eurostat (downloaded from Eurostat website in January 2014)
- *Number of facilities* in E-PRTR per NACE sector per country for the year 2011 (E-PRTR v5.1 was used)

In terms of countries, the comparison of these datasets covers the EU27, Iceland and Norway because only for these countries both datasets were available.

---

<sup>2</sup> Wherever NACE is mentioned in this chapter, NACE Rev.2 is meant.

<http://ec.europa.eu/eurostat/en/web/products-manuals-and-guidelines/-/KS-RA-07-015>

The comparison is performed based on the 3 and 4 digit NACE Rev.2 codes. The number of sectors is reduced to the sectors which are most relevant for the rarely reported pollutants in terms of emissions (as identified by the E-PRTR Guidance Document<sup>3</sup>). NACE sectors included in the test are shown in Table 14.

Before the actual assessment, a check has been performed to assess whether Member States have implemented the correct NACE sector codes for each of the E-PRTR activities. Therefore, Table 14 also includes the link between the selected NACE sectors and E-PRTR activity codes.

If the main part of the E-PRTR activity codes and NACE codes are not related as given in Table 13, they are excluded from further analysis with regard to their completeness. The result of the check on valid relations for the year 2011 is shown in Table 14.

**Table 14 Result of the NACE sector definition check in E-PRTR for the year 2011**

NACE sector code	NACE sector name	E-PRTR activities code	Number of facilities		% included
			included	excluded	
16.1	Preservation of wood	6.c	4	2	67%
19.1	Coke ovens	1.d	60	7	90%
20.1	Manufacture of basic chemicals and fertilizers	4.a, 4.b, 4.c	1288	132	91%
20.2	Manufacture of pesticides	4.d	11	1	92%
21.1	Manufacture of pharmaceuticals	4.e	114	3	97%
23.5	Manufacture of cement	3.c	1215	14	99%
24 <sup>4</sup>	Manufacture of iron and steel	2.b, 2.c, 2.d	1164	222	84%
<b>TOTAL</b>			<b>3856</b>	<b>381</b>	<b>91%</b>

Table 14 shows that for most NACE sectors, 90% or more of the facilities reported under E-PRTR are correctly related to NACE sector codes. Only for NACE code 16.1 (Preservation of wood) the number of facilities covered is substantially lower.

As a next step, an assessment has been done of the completeness of the E-PRTR dataset reported for a year, by comparing the number of facilities reporting for each NACE sector and country with the number in both the Prodcom database and the E-PRTR database. This comparison was made initially at the level of 4-digit NACE codes (the highest level of detail), checking for each 4-digit NACE code whether there are facilities reporting in the E-PRTR dataset for the same NACE code. The 4-digit NACE sector codes are explained in Table 15. The link between the NACE 4-digit codes and E-PRTR activity codes are provided in the E-PRTR database. There is a large variety of E-PRTR activities reported compared with each of the NACE codes, in particular of the chemical industry (NACE 20.11 to 20.17). In that case for

<sup>3</sup> European Commission 2006 (EU, 2006b)

<sup>4</sup> In this assessment, only the ferrous metal sectors are taken into account, i.e. NACE codes 24.1, 24.2, 24.3, 24.51 and 24.52.

the assessment only those facilities have been taken into account where the links between NACE 3-digit level and E-PRTR activity codes are as those provided in Table 14.

**Table 15 Explanation of the 4-digit NACE sector codes used in this study**

4-digit NACE	NACE sector name
16.10	Sawmilling and planning of wood
19.10	Manufacture of coke oven products
20.11	Manufacture of industrial gases
20.12	Manufacture of dyes and pigments
20.13	Manufacture of other inorganic basic chemicals
20.14	Manufacture of other organic basic chemicals
20.15	Manufacture of fertilisers and nitrogen compounds
20.16	Manufacture of plastics in primary forms
20.17	Manufacture of synthetic rubber in primary forms
20.20	Manufacture of pesticides and other agrochemical products
21.10	Manufacture of basic pharmaceutical products
23.51	Manufacture of cement
23.52	Manufacture of lime and plaster
24.10	Manufacture of basic iron and steel and ferro-alloys
24.20	Manufacture of tubes, pipes, hollow profiles and related fittings, of steel
24.31	Cold drawing of bars
24.32	Cold rolling of narrow strip
24.33	Cold forming or folding
24.34	Cold drawing of wire
24.51	Casting of iron
24.52	Casting of steel

### 5.1.1 Typical results

This chapter shows typical results from this assessment, in this case based on data for the year 2011. Figure 5 shows a typical result for each country included in the comparison, the number of NACE 4-digit sectors which are included in both Prodcum and E-PRTR (green bars). The blue bars show those sectors where Prodcum reports data only as confidential and the comparison could not be made. The red bar shows the missing part, for these NACE 4-digit sectors a production amount is given in Prodcum, but no facilities report air emissions in E-PRTR. For the grey bars there are no data in Prodcum (no actual production and no note of confidentiality) and subsequently also no facility reporting in E-PRTR.

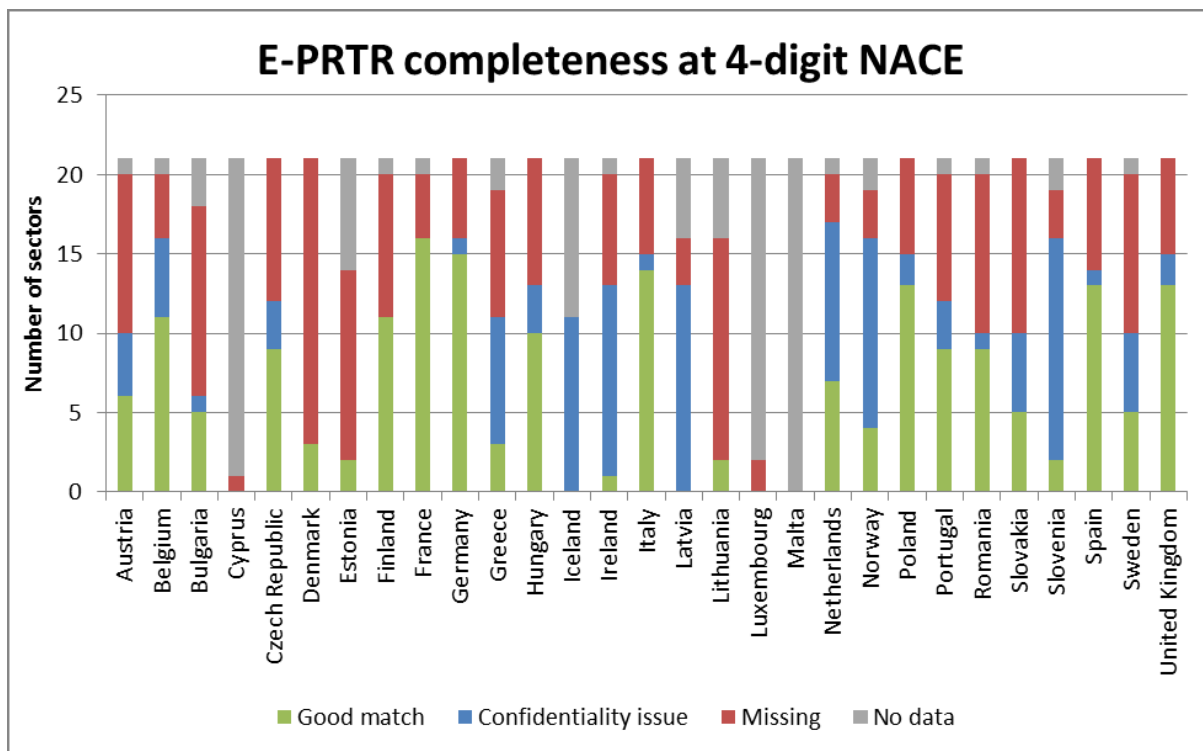


Figure 5 Completeness assessment per country of E-PRTR dataset (2011) using Eurostat Prodcom production data, aggregated at 4-digit NACE

Figure 5 shows that the E-PRTR coverage differs a lot between countries. The grey bars (which indicate no production in Prodcom and no release reports in E-PRTR) are typical for the small countries (Malta, Cyprus, Luxembourg) where production in most of the NACE 4-digit industries does not occur.

Figure 6 shows results from the same dataset along a different axis: the number of countries reporting per NACE 4-digit sector code. Again, the green bar illustrates where there is production reported in Prodcom and also reporting of air releases in E-PRTR (i.e. a good match). The blue bars indicate that in Prodcom the reporting is confidential, while the red bars show where E-PRTR reporting is missing. The grey bars indicate no production reported in Prodcom and no facility reports in E-PRTR.

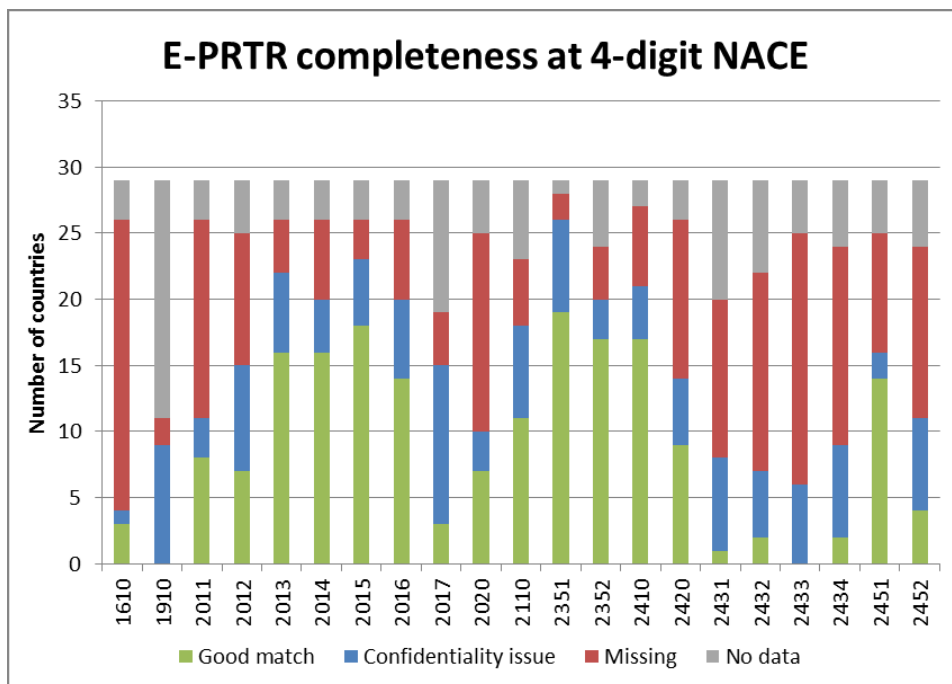


Figure 6 Completeness assessment per NACE 4-digit sector code of the E-PRTR dataset (2011) using Eurostat PRODCOM production data, aggregated by country and at 4-digit NACE

As shown in Figure 5 and Figure 6 there are several cases where Prodcom data show that production exists, but no emission reports are included in E-PRTR (the blue and red stacks). This could be due to incompleteness in the E-PRTR dataset, but this is not necessarily the case. There are a few possible causes for the identified gaps:

1. An E-PRTR facility has not been reported, although it should have been reported.
2. An E-PRTR facility has been reported, but different NACE codes are used for the E-PRTR reporting and the Prodcom production reporting.
3. An E-PRTR facility has not been reported because the production process does not result in a release or transfer above the threshold.

The second and the third cause result in the identification of possible gaps, while there are actually no E-PRTR facility reports missing. The next paragraphs give detailed analysis aiming to distinguish between the different possible causes of the identified gaps.

For the assessment of possibly falsely identified gaps as a result of cause 2, the production statistics from Prodcom are compared with the amount of E-PRTR facilities also at the level of 3-digit NACE sector codes. At this more aggregated sectoral level, it is expected that the number of falsely allocated NACE codes will be smaller. Figure 7 presents the same overview per country as Figure 5, but now for the comparison with the 3-digit NACE sector codes. When the grey bars (no data) are regarded as good matches, the relative amount of good matches increased slightly going from 4-digit to 3-digit NACE comparison, from 48% using 4-digit NACE to 55% using 3-digit NACE. When also the confidentiality issues are included, the match is 67% using 4-digit NACE and 69% using 3-digit NACE. This means, that even when



confidentiality issues in Prodcop are excluded from the analysis, still one third of the values in Prodcop is unaccounted for in E-PRTR.

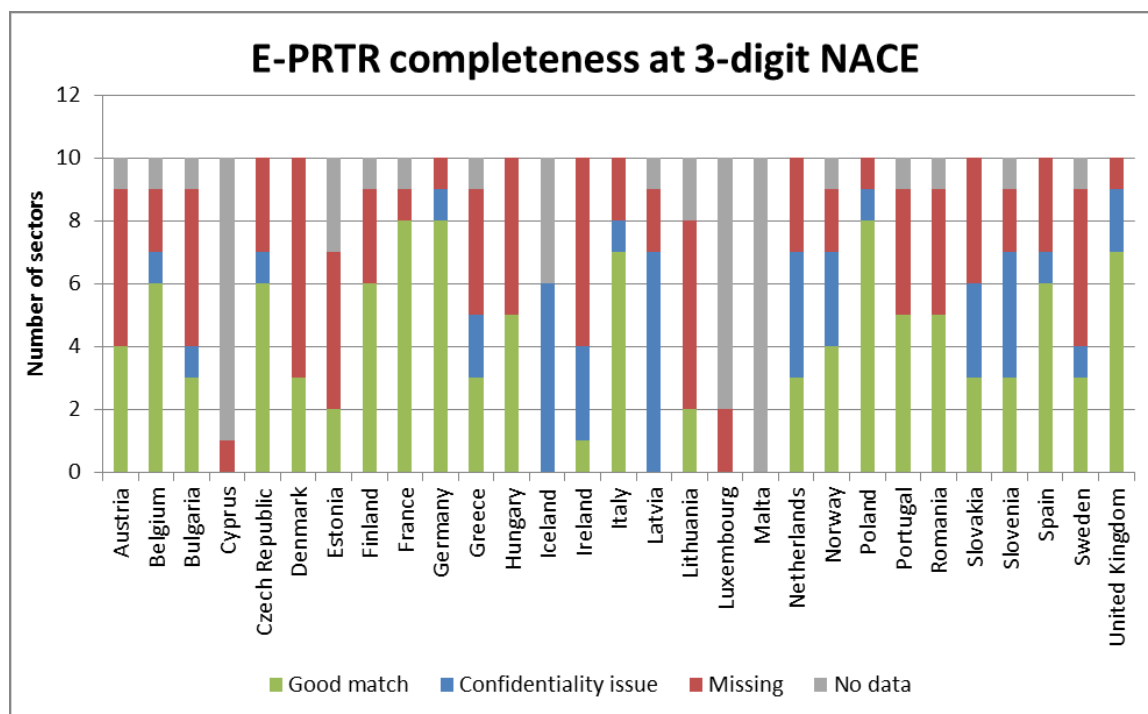


Figure 7 Completeness assessment per country of E-PRTR dataset (2011) using Eurostat Prodcop production data, aggregated at 3-digit NACE

Another reason for non-reporting of E-PRTR facilities may be that the emissions do not occur or lie below the threshold (cause 3). Therefore, we used in the analyses all the 4-digit NACE sectors of Table 15 in the screening of reported emissions.

As a first step, it we checked whether for all the 4-digit NACE sectors emission reports occur in the E-PRTR database. It was found that for each of these sectors, emissions are reported for selected pollutants and countries in the E-PRTR database. From this it is concluded that activities in all these sectors result in the release of emissions.

As a second step, an assessment can be made on how many different pollutants in the E-PRTR database are reported per country and per 4-digit NACE sector code. It is expected that the number of pollutants reported per 4-digit NACE is fairly constant throughout the countries, when assuming that the sector is relatively homogeneous and the size of installations is comparable between different countries, such that skewing from the threshold issue is prevented.

When analysing the dataset for the year 2011, we found that the number of pollutants per activity varies a lot between countries. This could be an indication for incomplete reporting, but may also be related to some releases being below the threshold. Table 16 includes all pollutants, not only the rarely reported emissions given in Table 13. Only country–NACE combinations with a production in the Prodcop data are included in Table 16. The

combinations without production or confidential production reported in Prodcop are coloured greyish and no emission reports are to be expected here. The zeros (0) in the table are essentially those cases previously identified as possible incompleteness cases (see Figure 5 and Figure 6).

**Table 16 Number of pollutants reported in E-PRTR dataset (2011) per 4-digit NACE and per country**

CountryName	1610	1910	2011	2012	2013	2014	2015	2016	2017	2020	2110	2351	2352	2410	2420	2431	2432	2433	2434	2451	2452	
Austria	0		0		0	1	5	0		1	0	5	1	15	0		0		0	0	0	
Belgium	2		3	0	11	16	5	14		1		14	8	28	0		0	0			2	
Bulgaria	0		0	0	4	0	6	0		0		5	3	0	0	0		0	0	0	3	
Croatia																						
Cyprus																						
Czech Republic	0		0		2	18	5	0		0	2	5	3	14	3	0	0	0	0	0	1	0
Denmark	0	0	0	0	0	2	0	0	0	1	0	6	0	0	0	0	0	0	0	0	0	0
Estonia	0		0	0	1	0	0	0		0		5	0	0	0						0	0
Finland	1		0	2	8	6	3	1	0	0	1	4	3	18	0	0	0	0	0	0	1	0
France	0		5	2	19	31	4	17	5	2	5	16	6	25	3	0	0	0	0	1	17	3
Germany	1		3	5	13	20	7	11	1	0	4	14	8	25	3	3	0	0	0	0	9	0
Greece	0		0				6			0		7	0	8	0		0	0	0			
Hungary	0		1		3	1	5	4		0	4	3	1	10	0	0		0	0	0	1	0
Iceland																						
Ireland	0					0	0			0	1			0	0						0	
Italy	0		1	0	10	20	5	3	4	0	5	10	3	21	5	0	1	0	0	0	2	4
Latvia	0		0			0																
Lithuania	0		0	0	0	0	7	0		0		5	0	0	0		0	0	0	0	0	
Luxembourg																						
Malta																						
Netherlands	0		6	1		12	11	18		3	0								0		4	
Norway	0		0		16								1	18	0							1
Poland	0		0	1	7	15	18	3		1	3	20	5	18	2	0	0	0	0	1	1	0
Portugal	0		0	0	6	3	3	8		0	0	11	1	11	1		0	0			4	0
Romania	0		0	0	2	1	5	2		0	1	7	1	7	1	0	0	0	0	0	0	0
Slovakia	0		0	0	0	0	5	10			0	5	3	17			0	0	0	0	0	0
Slovenia	0		0							0					6							2
Spain	0		1	3	9	11	5	7	0	0	4	14	7	20	6	0	0	0	0	0	5	15
Sweden	0			0	1	5		9		0			5	12	0	0	0	0	0	0	0	0
Turkey																						
United Kingdom	0		4	2	18	19	6	5	0	3	6	17	2		1	0	2	0	0	0	4	0

To illustrate how to interpret this information, Figure 8 shows the same data for 4 selected 4-digit NACE sectors. Since the issues with homogeneity of the sector and the thresholds of pollutants both influence the outcome, it is not possible to draw definite conclusions from this table and figure. However, when comparing per sector the number of pollutants reported in different countries, potential outliers can be identified which could be subject to more in-depth scrutiny at the level of an individual Member State.

For instance, for NACE 24.10 (Iron and Steel production) there are no emission reports for Bulgaria, Denmark, Estonia, Ireland and Lithuania, while production quantities are provided by Prodcop. These are potentially cases of incompleteness in E-PRTR reporting and could therefore be subject to further checking by the Member States.

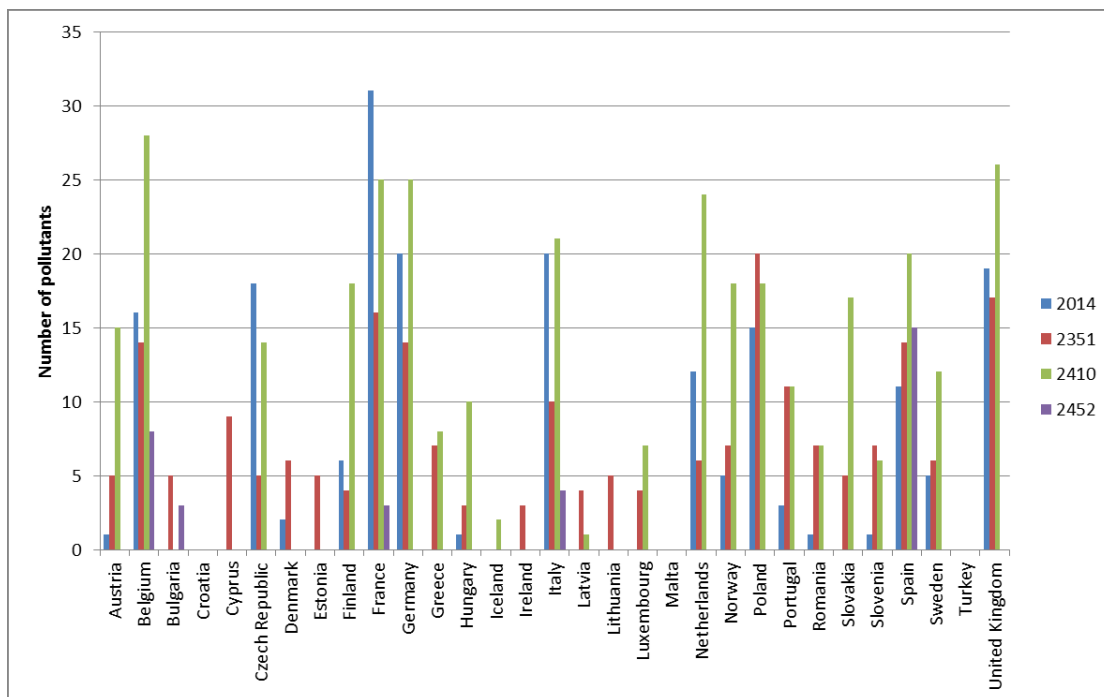


Figure 8 Number of pollutants reported per country for four selected NACE 4-digit sectors

### 5.1.2 Assessment of rarely reported pollutants

In this chapter the focus is on the rarely reported pollutants listed in Table 13. According to the E-PRTR Guidance Document (EC, 2006b), reporting of these pollutants is expected in the NACE 4-digit sectors provided.

Table 17 shows the reporting for these pollutants in E-PRTR for the year 2011, for three of the 4-digit NACE sectors provided in Table 15. Only for these NACE 4-digit sectors emissions have been reported, totaling 13 release reports. For the other 4-digit NACE sectors included in Table 15, no emissions are reported at all.

As can be seen from Table 17, asbestos is not reported at all in the year 2011. For anthracene there are four facilities reporting and for the other pollutants only one or two. The countries for which facility emissions are reported are Belgium (total 4), France (3), Germany (2) and Finland, Italy, Netherlands and the United Kingdom (all 1).

This table shows that reporting of these pollutants is very poor across Member States for the year 2011.

**Table 17 Number of facilities reporting for in E-PRTR for the year 2011 (all countries)**

NACE 4-digit sector	1,1,1-trichloro ethane	1,1,2,2-tetrachloro ethane	Anthracene	Asbestos	Halons	Hexachloro benzene (HCB)	Trichloro benzenes (TCBs) (all isomers)
20.13	0	0	0	0	0	2	0
20.14	1	1	0	0	2	0	1
24.10	1	1	4	0	0	0	0
<b>Total</b>	<b>2</b>	<b>2</b>	<b>4</b>	<b>0</b>	<b>2</b>	<b>2</b>	<b>1</b>

Another way of assessing the completeness of E-PRTR rarely reported pollutants is by comparing to *other emission reporting instruments*. Here we demonstrate it by comparing the E-PRTR emissions at national level to the reported national emission inventory under the Convention on Long-Range Transboundary Air Pollution (CLRTAP). From the selected rarely reported pollutants, the only pollutant that is also reported in CLRTAP national inventories is Hexachlorobenzene (HCB). To assess the consistency between reporting under E-PRTR and under CLRTAP, a cross-check has been made with CLRTAP reporting, also in view of the threshold in E-PRTR which equals 10 kg/year.

Data have been downloaded from the CEIP website in July 2014. The sum of the reported HCB emissions in the E-PRTR dataset 2011 for the 4-digit NACE codes 20.13 and 20.14 is compared to the HCB emissions from the total chemical industry (covered by NFR codes 1.A.2.c, 2.B.1-2.B.5, 2.E, 3.C) selected from the CLRTAP sectors; while for those for the 4-digit NACE codes 24.10, emissions from the iron and steel production sector from the CLRTAP sectors (covered by NFR codes 1.A.2.a, 2.C.1, 2.C.2) are selected for comparison.

Table 18 shows the HCB national totals reported in 2011, as well as totals for the chemical and the iron and steel sector. Values highlighted in red are above the threshold. The table shows that for many countries the national reported total is below the threshold, i.e. emission reports under E-PRTR cannot be expected. Looking at individual industries, Belgium, Bulgaria and Finland are the only countries where in the chemical and/or iron and steel industry total reporting is above the threshold.

In the E-PRTR dataset 2011, for HCB only at Belgium and Finland are facilities reporting in 4-digit sector NACE 20.13 (manufacture of other inorganic basic chemicals), which is consistent with their emission reports in the CLRTAP national inventory, as these are the only two countries where the reported emission for the total chemical industry exceeds the threshold for E-PRTR reporting, as shown in Table 18.

For the iron and steel sector, there is one country (Bulgaria) for which the total emissions according CLRTAP are exceeding the threshold; however no emission reports are in the E-PRTR database for 2011. This is a potential incompleteness in E-PRTR reporting from Bulgaria, but may as well be related to the threshold issue as the reported emission of 21.37 kg in the iron and steel sector for the national CLRTAP-inventory may result from the addition of emissions from multiple facilities.

**Table 18 National reporting of HCB in 2011 in kg (data from CEIP under CLRTAP) versus E-PRTR dataset 2011 (data for chemical industry, NACE 20.13 only)**

Country	CLRTAP Total of inventory	CLRTAP Iron/Steel	CLRTAP Chemical	E-PRTR NACE 20.13 code (Chemical industry)
Austria	37.44	3.91	0.12	
Belgium	40.95	18.3	10.57	68.7
Bulgaria	21.74	21.37	0	
Switzerland	0.21	0	0	
Cyprus	0.01	0	0	
Czech Republic	2.46	0	0.11	
Germany	3.46	0	0	
Denmark	2.60	0	0	
Estonia	0.18	0	0	
Spain	4.39	0.04	0	
Finland	33.19	0.07	18.8	18.8
France	15.97	0.01	0.01	
United Kingdom	24.23	0	0	
Croatia	0.13	0	0	
Hungary	2.11	0.05	0	
Ireland	1.16	0	0	
Iceland	0.04	0	0	
Italy	13.72	0	0	
Lithuania	0.21	0	0	
Latvia	0.32	0	0	
Malta	0.00	0	0	
Netherlands	1.59	0	0	
Norway	1.26	0.02	0.04	
Poland	12.59	0.93	0	
Portugal	0.76	0	0	
Romania	1.75	0.14	0	
Sweden	0.02	0	0	
Slovenia	0.56	0	0	
Slovak Republic	1.20	0.09	0	

## 5.2 Conclusions

This chapter 5 presents a new methodology for identifying possible incompleteness of air emissions reported in the E-PRTR database. By comparing with Eurostat Prodcom *production statistics* for selected NACE level 4 sectors and by Member State, country–sector combinations can be identified which appear to have production, but for which no emissions are reported in E-PRTR. These omissions may result from incomplete reporting, but also other reasons for the differences may exist, in particular cases involving the threshold issue.

At a significant amount of about 1/3 of the cases a production quantity is reported in Prodcum without emission reports in E-PRTR. Hence, these cases have been examined in more detail, attempting to further limit the number of cases and as such to identify the possible causes of flagging for possible E-PRTR reporting incompleteness. By *sector aggregating* to a comparison at NACE level 3, possible misreporting of E-PRTR emissions in the wrong NACE 4-digit sector (e.g. 20.14 instead of 20.13) can be tackled. This type of misreporting could be likely as some NACE 4-digit sectors are closely related. An analysis on the reported E-PRTR data for 2011 showed that this comparison indeed improves the relative amount of good corresponding cases, thus limiting the possible misreporting cases.

By comparing the *number of pollutants reported in each sector between countries*, we get another indication for possible incompleteness. This test assumes that if an activity takes place in a country, a certain number of pollutants are expected and this number is relatively constant throughout all countries. Despite that this test implicitly assumes that the sector is relatively homogeneous and the typical size of facilities is comparable between countries, the test may be able to help identifying possible cases of incompleteness.

More specifically for *rarely reported pollutants*, reporting of their emission in E-PRTR is very limited, due to thresholds not exceeded and incompleteness checks are difficult as other emission inventories do not register these pollutants. For example, in this study with the 2011 dataset for HCB there were only two emission reports for the NACE sectors considered; if all sectors are to be considered there would be three emission reports. By comparison with *national totals*, it becomes clear that the threshold value is such that almost no emission reports can be expected. For most EU Member States, the total national HCB emission reported under CLRTAP is lower than the E-PRTR threshold value. Only in 8 Member States the total reported HCB emissions exceeds the threshold for a single facility in E-PRTR, which explains why HCB is rarely reported. For the other rarely reported pollutants, no comparison with the national inventory was possible since these pollutants are not part of the national inventory submissions under CLRTAP.

Further work should focus on a way to further refine the list of flagging for possible incompleteness issues, or to identify priority sectors or pollutants where reporting improvements are necessary. The relevance of the threshold issue in this assessment varies largely between the NACE level 4 sectors. Perhaps it is possible to preselect those sectors where emissions are expected to exceed thresholds most prominently. Sectors for which emissions above threshold are to be expected could be identified in several ways. For example:

- If many countries report emissions for a certain sector, then it is more likely that emissions above threshold occur by facilities in these sectors. For example at the NACE sectors 20.13, 20.14, 20.15, 20.16, 23.51, 23.52 and 24.10 at least half of the countries report emissions.
- It might be possible to calculate implied emission factors for countries which report many pollutants and which have a high productivity. These implied emission factors

could be used to check whether emissions above the threshold could be expected for those countries that did not report emissions in a certain sector.

- The selection of sectors could also focus on sectors for which it is expected that the facilities are similar. For example, for NACE 24.10 (Iron and steel production) there is a large variation in the number of pollutants reported between countries. Since this sector is expected to consist of mostly large and similar type of installations the number of pollutants to be reported is expected to be relatively homogeneous.

When using the results of this assessment, one should bear in mind that only *potential* cases of incompleteness are to be identified or flagged. This assessment may however also overlook cases of incomplete reporting as well. It will always be necessary to check at Member State level, i.e. by the national experts, whether there is indeed a case of incomplete reporting in E-PRTR. However, by providing a list of detected possible cases of incomplete reporting, this QA/QC procedure contributes to help improve the quality of the reporting on E-PRTR emissions.

## 6 Using economic and employment statistics to review E-PRTR emissions data

Aether has undertaken a simple initial assessment of whether it is possible to generate a review methodology that uses economic and employment statistics. The aim was specifically to help identify non-reporting of emissions to the E-PRTR from selected countries, for selected pollutants and sources, as well as reviewing the quality of the data that is reported.

This initial work was undertaken on the 2011 emissions data, and a summary is presented in the following sections, indicating the tests which were considered viable in terms of identifying where Parties might be under-reporting emissions. In addition, an investigation was made on how variable the tests were across different years for a single country. The output from this work with the 2011 datasets differs from that previously reported because input datasets have been revised, and some errors addressed.

The work was then continued by applying specific tests to 2012 emissions data. Delivery of this part of the task had to be postponed from the initial schedule until the beginning of October 2014 as the economic and employment statistics were not available from Eurostat until then.

The output of this work is to identify specific data in the PRTR dataset which is not consistent with the main body of the data, and in particular data that suggest under-reporting of emissions. The EEA may then wish to use this information to approach Parties for an explanation as to why these data are substantially different to data from other Parties. In addition, recommendations are made on how the tests developed and presented here can be further refined and automated for future reporting cycles.

### 6.1 Approach

The basic premise of this work is that the emissions for a given pollutant from a given E-PRTR activity do not vary substantially per unit of economic activity, or per employee between countries. Consequently it should be possible to generate an economic and employment “indicator” (e.g. emissions per employee) and compare the results across the countries to identify outliers – and in particular those which appear to be considerably lower than other countries.

However, in practise the relationship between emissions and economic or employment data will give a large scatter for many pollutants and industrial sectors across the countries. This is simply a reflection of the different situations across the countries. It is also possible that the economic or employment datasets do not provide reliable or accurate information. It is therefore necessary to first undertake some investigative work to identify which pollutant, source and activity data combinations give a strong enough correlation to allow outliers to be identified with some certainty, and whether the available economic and employment data are robust enough for this analysis.

The following is a consideration of some of the assumptions that need to hold for this approach to be effective:



- First, it is assumed that **all industrial activities within each E-PRTR activity are similar**. It is, of course, entirely possible for there to be different activities within each E-PRTR activity, or different methods of production that give rise to differing emissions profiles, e.g. electric arc furnaces versus blast furnaces in iron and steel production. If there is a substantial geographical variation in these production methods then there may be valid reasons for countries differing in the emissions per unit of economic output or employee. As a result, the resolution of the available data needs careful consideration.
- Secondly, it is assumed that **the distribution of plant size or capacity within an E-PRTR activity is similar between different countries**. If there were many small plant in the E-PRTR activity, then a larger proportion of the emissions sources would be below the E-PRTR reporting threshold. It would therefore be possible for two countries of equal employment or economic “size” to give widely differing emissions reported to the E-PRTR. It should also be noted that larger plants tend to be able to draw on economies of scale, and therefore usually higher levels of output per employee. The requirement for similar plant size distribution within an E-PRTR activity is more likely to hold in some categories than others. For example, all plant undertaking activities in primary iron and steel manufacture are likely to be sufficiently large to require data to be reported to the E-PRTR. Consequently a reasonably consistent relationship between E-PRTR emissions and economic activity across different countries can be expected. However, for other sources, e.g. food and drink, it is possible for a substantial portion of the plants to be below the E-PRTR reporting threshold, resulting in a less consistent relationship between emissions in the E-PRTR and employment or economic activity across the countries.
- It is also assumed that the **efficiencies of the industrial activities are generally similar** across different countries. Whilst there may be variations in efficiencies (in terms of the number of people required for a unit of production or output), market forces typically require that industries are competitive in the international arena. Therefore there are not expected to be substantial differences across the countries.

### 6.1.1 Input Data

A number of data are available as indicators of economic activity. For this study “Value Added at Factor Cost” was sourced from Eurostat for both 2011 and 2012. This is referred to in this study as GVA<sup>5</sup> (Gross Value Added) data. This has the advantage of being suitably detailed at the industrial sector level, and in particular is available in a NACE code format that allows easy use with the E-PRTR emissions data.

Employment data for 2011 and 2012 was taken from Eurostat (Number of Employees as full-time equivalents (FTEs))<sup>6</sup>. These data provide national employment data by country split by NACE code.

---

<sup>5</sup> Eurostat indicator code V12150.

<sup>6</sup> Eurostat indicator code V16140.

It is possible to obtain employment data that is further sub-divided into a more detailed geographical scale (e.g. NUTS3). Use of these data was not explored for this study, but offers potential for more detailed studies to be undertaken in the future if considered to be of value.

It was necessary to wait until the beginning of October 2014 for the 2012 data to be available from Eurostat. However, despite waiting for this release date, there were still gaps evident in the economic and employment data. For example the dataset from Eurostat includes production statistics for France, but there is no GVA or employment data in the same industrial sectors. These gaps directly impact on the number of data points that can be included in the indicators.

### 6.1.2 Methodology

Emissions by NACE category for each country were obtained from the E-PRTR database from the EEA website for 2011 and 2012<sup>7</sup>. Employment data, GVA and production data for selected NACE categories were obtained from the Eurostat website. The NACE categories chosen were selected based on their size of emissions, and/or the expected quality of the data reported to the E-PRTR. The following E-PRTR activities (provided with corresponding NACE codes) proved to give the larger E-PRTR datasets whilst also providing a good match with the available employment and GVA data<sup>8</sup>:

- Manufacture of basic iron and steel and of ferro-alloys (C24.10)
- Manufacture of cement (C23.51)
- Manufacture of lime and plaster (C23.52)
- Manufacture of paper and paperboard (C17.12)
- Sewerage (E37.00)
- Water collection, treatment and supply (E36.00)

Eurostat revised some of the categorisation and naming of industrial sectors between the 2011 and 2012 datasets. However, data was handled to be as consistent as possible, and is presented with the source names and codes given above.

Pollutants emitted to air were selected to include: air quality pollutants, greenhouse gases, heavy metals and POPs. Sulphur dioxide was included as a pollutant that should be accurately reported to the E-PRTR datasets, and one which arises from iron and steel and the

---

<sup>7</sup> E-PRTR data downloaded from the V5 database on the EEA website: <http://www.eea.europa.eu/data-and-maps/data/member-states-reporting-art-7-under-the-european-pollutant-release-and-transfer-register-e-prtr-regulation-8>

<sup>8</sup>2011 and 2012 data from Eurostat table [sbs\_na\_ind\_r2]

[http://appsso.eurostat.ec.europa.eu/nui/show.do?query=BOOKMARK\\_DS-120933\\_QID\\_41A7F38A\\_UID\\_-3F171EB0&layout=TIME,C,X,0;GEO,L,Y,0;NACE\\_R2,L,Z,0;INDIC\\_SB,L,Z,1;INDICATORS,C,Z,2;&zSelection=DS-120933NACE\\_R2,C10;DS-120933INDICATORS,OBS\\_FLAG;DS-120933INDIC\\_SB,V12150;&rankName1=INDIC-SB\\_1\\_2\\_-1\\_2&rankName2=NACE-R2\\_1\\_2\\_-1\\_2&rankName3=INDICATORS\\_1\\_2\\_-1\\_2&rankName4=TIME\\_1\\_0\\_0\\_0&rankName5=GEO\\_1\\_2\\_0\\_1&sortC=ASC\\_-1\\_FIRST&rStp=&cStp=&rDCh=&cDCh=&rDM=true&cDM=true&footnes=false&empty=false&wai=false&time\\_mode=NONE&time\\_most\\_recent=false&lang=EN&cfo=%23%23%23%2C%23%23%23.%23%23%23](http://appsso.eurostat.ec.europa.eu/nui/show.do?query=BOOKMARK_DS-120933_QID_41A7F38A_UID_-3F171EB0&layout=TIME,C,X,0;GEO,L,Y,0;NACE_R2,L,Z,0;INDIC_SB,L,Z,1;INDICATORS,C,Z,2;&zSelection=DS-120933NACE_R2,C10;DS-120933INDICATORS,OBS_FLAG;DS-120933INDIC_SB,V12150;&rankName1=INDIC-SB_1_2_-1_2&rankName2=NACE-R2_1_2_-1_2&rankName3=INDICATORS_1_2_-1_2&rankName4=TIME_1_0_0_0&rankName5=GEO_1_2_0_1&sortC=ASC_-1_FIRST&rStp=&cStp=&rDCh=&cDCh=&rDM=true&cDM=true&footnes=false&empty=false&wai=false&time_mode=NONE&time_most_recent=false&lang=EN&cfo=%23%23%23%2C%23%23%23.%23%23%23)

cement industry. NO<sub>x</sub> and PM<sub>10</sub> were included as air quality pollutants for which there are both emissions and ambient concentration targets in international legislation.

For pollutants emitted to water, three of the most commonly used indicators for water pollution were included in the methodology assessment (emissions of nitrogen, phosphorus and total organic carbon).

The ten pollutants included in this study were:

- NO<sub>x</sub> (air)
- Sulphur dioxide (air)
- CO<sub>2</sub> (air)
- PM<sub>10</sub> (air)
- Mercury (air)
- PCBs (air)
- PAHs (air)
- Nitrogen (in water)
- Phosphorus (in water)
- Total Organic Carbon (in water)

National level E-PRTR emissions for each activity were divided by GVA or employment to obtain indicators for each E-PRTR activity. These indicators were compared between countries to identify patterns, consistency, outliers, and usability of the methodology. Not all E-PRTR activities were available for all pollutants, and some pollutant-activity combinations had very few countries reporting. These were discounted from this study.

For the 2011 data, scatter plots were produced mapping pollutant emissions against GVA or employment, to view the consistency of the relationship across the countries. In addition, relationships between emission and the Eurostat data were investigated for a single country across the time series. The results from this assessment on the 2011 datasets were used to decide which combinations of pollutant emissions and activity data could be used with the 2012 data to identify potential under-reporting.

Not all plots are presented in this report, but they are available in the accompanying Excel spreadsheet “2012\_E-PRTR\_Data\_Review\_Aether\_issue1.xls”.

### 6.1.3 Initial investigation of data tests

2011 emissions data have been compared against: GVA (millions of Euros), Employment data (expressed as full time equivalents, or “FTE”) and Production value (Turnover in millions of Euros) to determine whether there is enough consistency in the relationship across countries to identify outliers.

As the purpose of this assessment was to evaluate the relationship between the datasets, data points were not included in the statistical assessment where the GVA, Employment or Production data was zero. These gaps in the Eurostat data are significant in that they impact on the extent to which emissions data from the E-PRTR can be utilised in this study. However

it is important to ensure that these zero values are not included in the data analyses to avoid skewing the results.

Data points were included for where there were no emissions data from the E-PRTR, but GVA, Employment or Production data were available. The reason for this is that it highlights where emissions data may be missing in E-PRTR (the primary aim of this work). In addition, these points may represent E-PRTR activities where all plants are below the reporting threshold (i.e. zero E-PRTR reported emissions, but non-zero GVA, Employment or Production data). These points should therefore be included in establishing the relationship between emissions and activity data.

Some examples of the data are provided in the following sections as simple scatter plots. Linear lines of best fit with  $R^2$  values are included as a simple indication of the strength of the relationship between the datasets, although it is recognised that this has limited use, for reasons explained below.

#### 6.1.4 Emissions to air

The strength of the relationship between emissions and GVA or Employment or Production varies greatly between E-PRTR activities and pollutants, as indicated in Table 19.

**Table 19 Result from the regression analysis of air emissions with activity data (2011)**

Air Pollutant	Lowest and Highest $R^2$ Values	
NO <sub>x</sub>	0.38 (production of paper)	0.84 (production of cement)
Sulphur dioxide	0.21 (GVA of paper)	0.86 (employees in iron & steel)
CO <sub>2</sub>	0.15 (employees in iron & steel)	0.69 (GVA of iron & steel)
PM <sub>10</sub>	0.00 (employees in iron & steel)	0.87 (employees in cement)
Mercury	0.37 (production of cement)	0.53 (employees in iron & steel)
PCBs	0.01 (production of cement)	0.29 (production of iron & steel)
PAHs	0.00 (GVA from paper)	0.25 (GVA from cement)

Many of the low  $R^2$  values arise from datasets that have a significant amount of zero emissions reporting for non-zero activity datasets. This can result in two subsets, and the simply approach used here does not represent this well.

However, as this stage of the project is an initial assessment (and also because the size of the datasets is not very large), it was not considered necessary to undertake any more detailed investigations with the relationships showing very low  $R_2$  values.

Strong relationships were seen for:

- CO<sub>2</sub> from iron and steel and paper and paperboard with GVA;
- PM<sub>10</sub> from cement with employment data;
- SO<sub>2</sub> from cement with GVA, and iron and steel with employment data;

- NOx from cement, and iron and steel with GVA, employment and production data, and also paper production with GVA.

Some example figures are included below for illustrative purposes. It should be appreciated that points where emissions are zero, but GVA, Employment or Production data are non-zero have been retained in the datasets that are shown (i.e. there are points which lie on the x-axis). This is intentional, to show the extent to which potential non-reporting of emissions occurs in the plots (for reasons explained above), and varies from pollutant to pollutant. However, this does mean that the  $R^2$  values that are shown on the figures do not give a true representation of the regression analysis between the corresponding variables – because a subset of data known to represent a specific situation that is different to the main dataset has been included in the analysis.

Figure 9, Figure 10 and Figure 11 present NOx emissions, plotted against GVA, Employment and Production respectively. Compared to other pollutants, there are relatively few points lying on the x-axis, indicating that the reporting of NOx emissions for these E-PRTR activities is generally better than other pollutants. The relationships between emissions and the different activity datasets are generally better than for other pollutants as a direct result.

Iron and Steel production and Cement production give consistently better agreements with the economic and employment data than other E-PRTR activities.

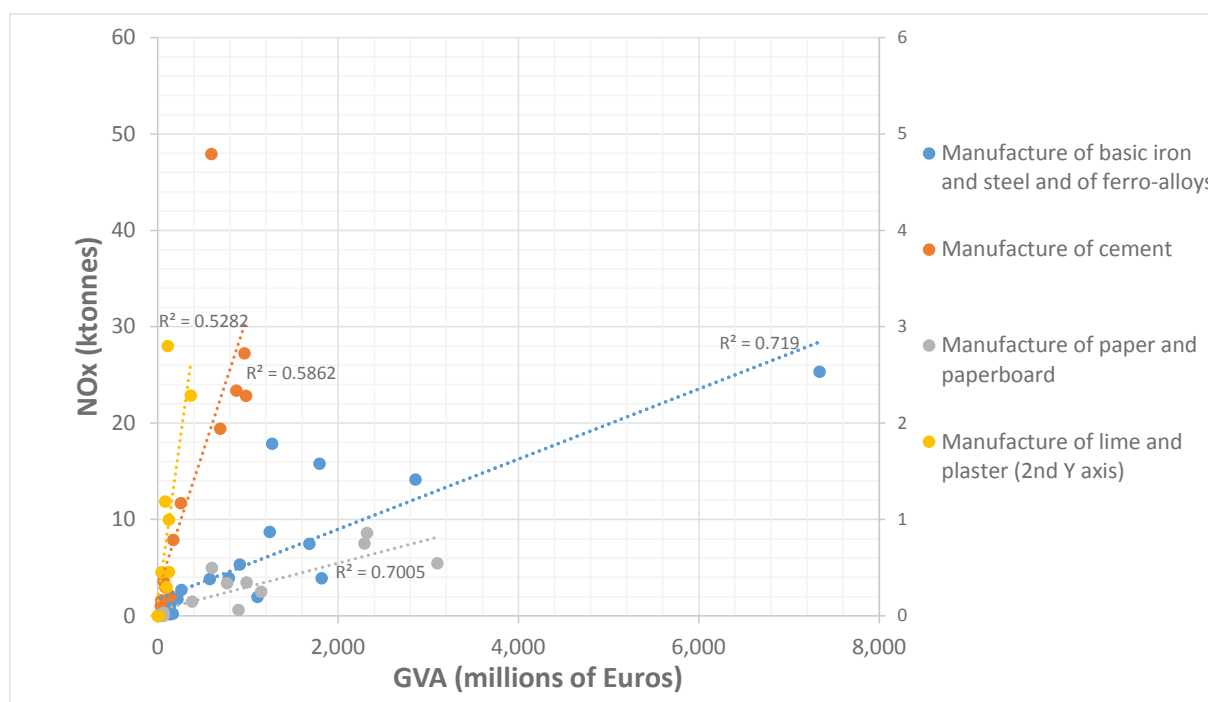


Figure 9 NOx emissions versus GVA for selected E-PRTR activities

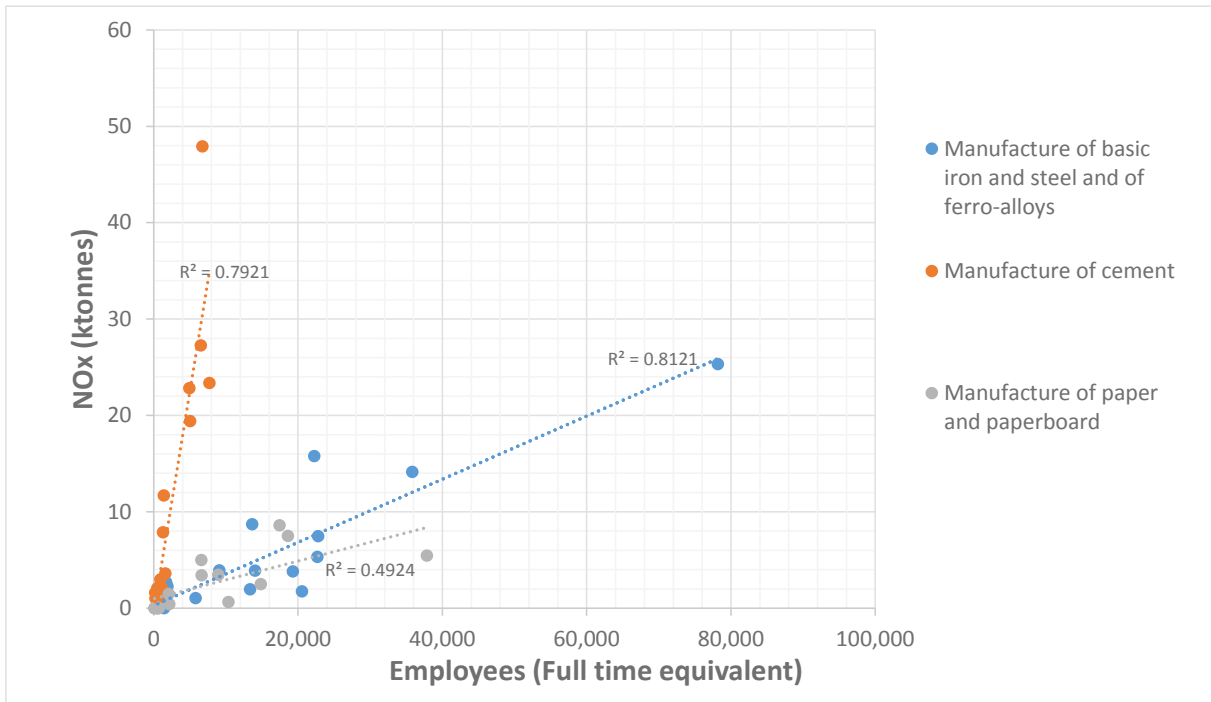


Figure 10 NOx emissions versus employment (FTE) for selected E-PRTR activities

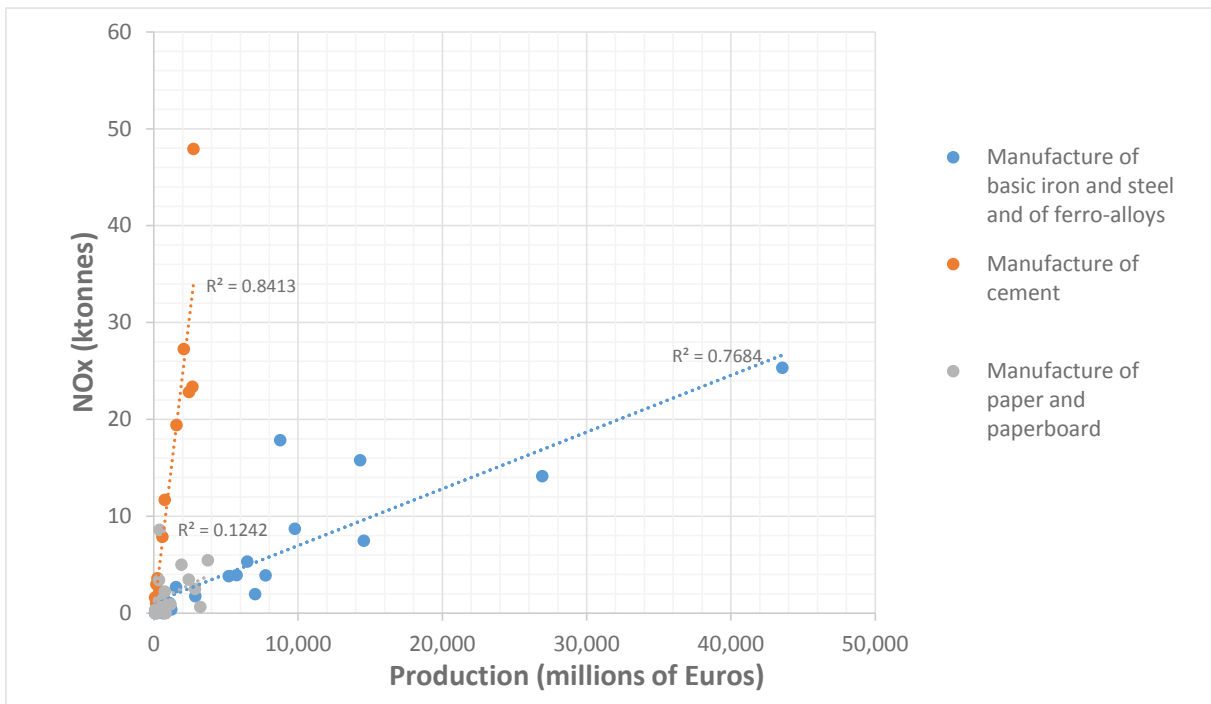
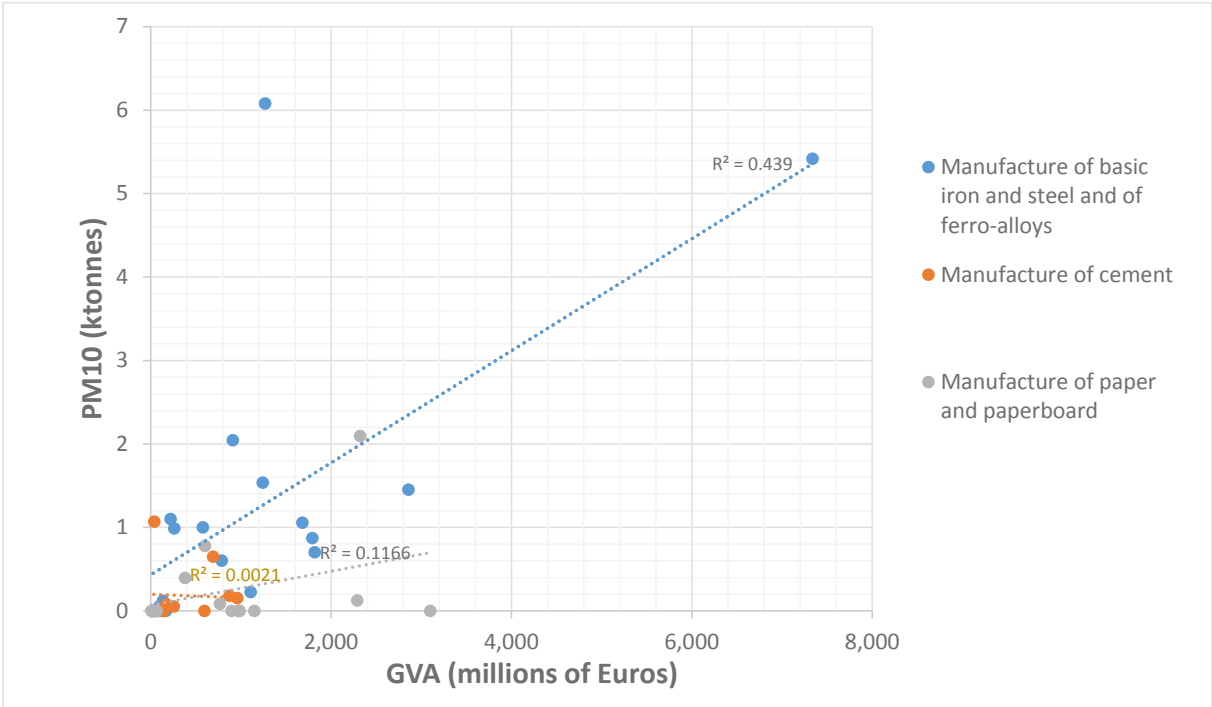


Figure 11 NO<sub>x</sub> emissions versus production for selected source E-PRTR activities

By way of contrast, Figure 12 shows emissions of PM<sub>10</sub> versus GVA. A significant number of points can be seen on the x-axis, indicating that zero emissions have been reported, even

where the corresponding E-PRTR activity makes a significant financial contribution to the country (in this case manufacture of paper, and to a lesser extent manufacture of cement). This suggests non/under-reporting of emissions from the cement and paper E-PRTR activities, however it is possible that there are valid explanations for this. For example it may be that the corresponding industrial sectors in these countries are dominated by smaller installations, which emit at levels below the E-PRTR reporting thresholds, meaning that the national total emission is proportionally lower than that observed from other countries. The points on the x-axis also skew the line of best fit and for this reason they are not a particularly helpful in the quantitative representation of the agreement between the two datasets.



**Figure 12 PM10 emissions versus GVA for selected E-PRTR activities**

There are instances where the E-PRTR activity and pollutant combination results in a very much reduced dataset. These examples are also subject to line of best fit being very much skewed by outliers. Figure 13 of PCBs versus GVA illustrates this point. Manufacture of cement has one non-zero value, and therefore does not give a particularly meaningful line of best fit. For iron and steel, the line of best fit is also significantly impacted by zero emissions values, and the point with greatest GVA, which also appears to be a questionable data point.

This illustrates the caution that is needed in using output from these analyses, which is why at this stage of the project they are used in an indicative way, only to guide the analysis undertaken on the 2012 data.

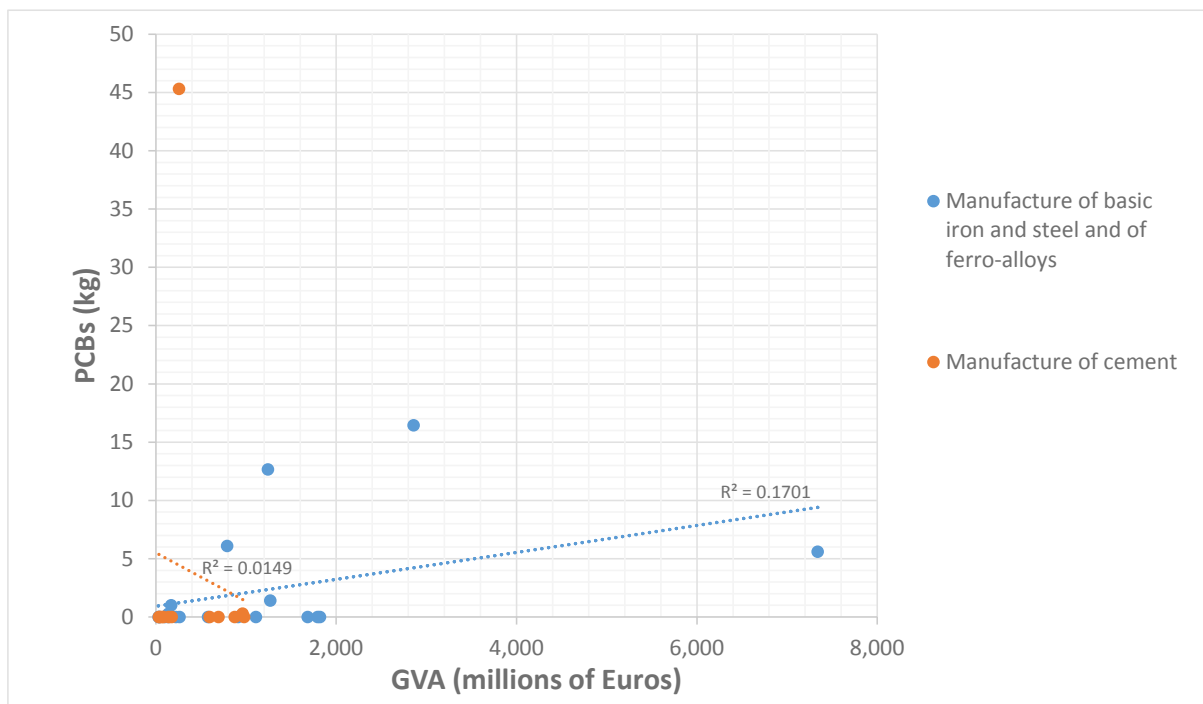


Figure 13 PCBs (kg) versus GVA for selected E-PRTR activities

### 6.1.5 Emissions to water

The relationships between emissions to water and employment or economic indicators (Table 20) are not as strong as those for emissions to air.

Table 20 Result from the regression analysis of emissions to water with activity data

Air Pollutant	Lowest and Highest R <sup>2</sup> Values	
Nitrogen (water)	0.09 (production in paper)	0.68 (production in sewerage)
Phosphorus (water)	0.03 (GVA in iron & steel)	0.54 (production in sewerage)
TOC (water)	0.01 (GVA in iron & steel)	0.76 (production in sewerage)

As with the emissions to air, some combinations of datasets are affected by zero emissions points being included, giving a clear subset of the main dataset and impacting on the R<sup>2</sup> value.

Nitrogen emitted to water from a range of different sources is shown in Figure 14 and Figure 15 against employment.



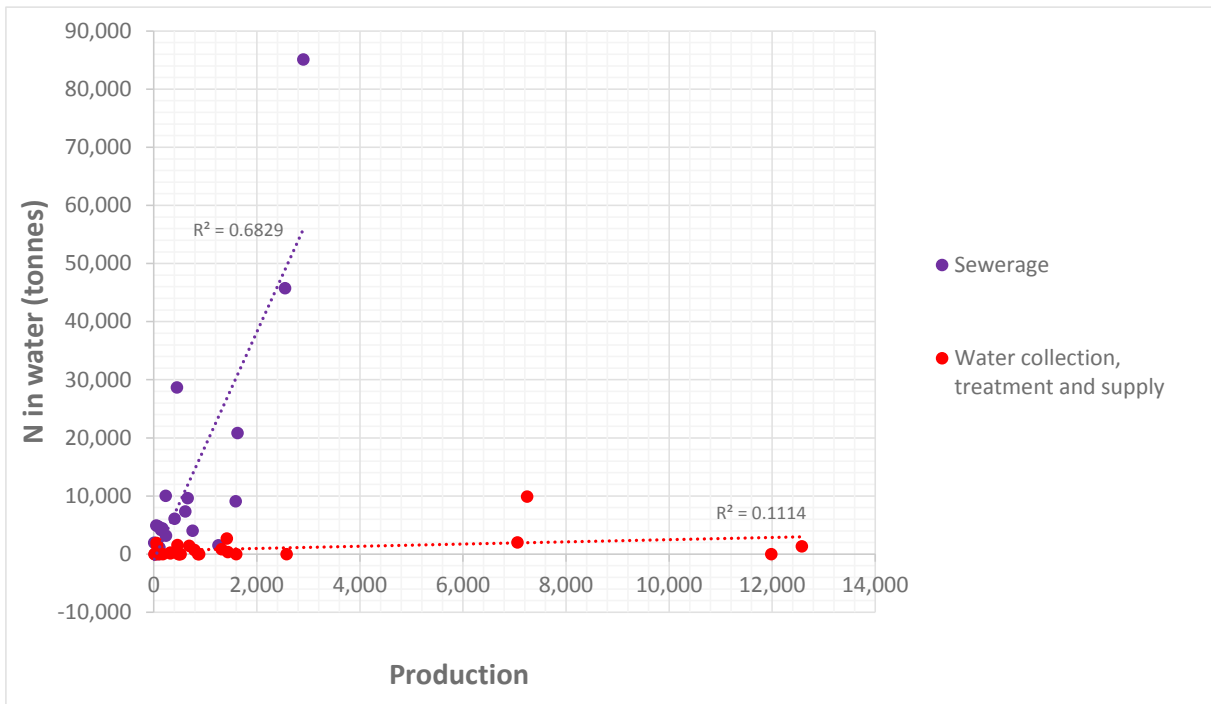


Figure 14 Nitrogen emitted to water versus production in sewerage and water treatment

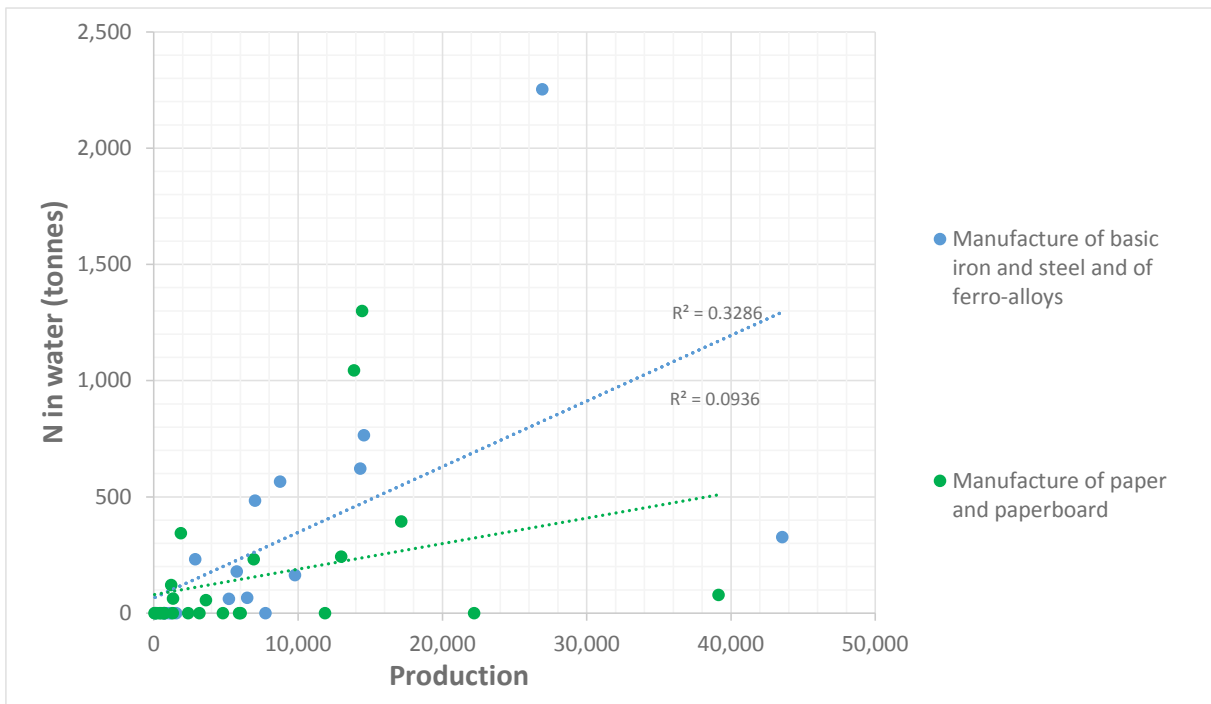


Figure 15 Nitrogen emitted to water versus production in iron & steel and paper

These plots also illustrate that the extent to which zero is reported for emissions is variable across the E-PRTR activities.

In general the emissions to water do not show such a strong agreement as the emissions to air. Also, emissions of Phosphorus give weaker relationships with activity data than those for Nitrogen emissions to water. Total Organic Carbon (TOC) gives stronger relationships with activity data, on a relative basis.

### 6.1.6 Consistency of the relationship across the time series

Data across a time series were also analysed. This was to investigate whether the relationships between emissions and activity indicators varied from year to year within the same country.

However, the datasets do not vary greatly from year to year, and as a result it is difficult to determine whether there is a particularly strong relationship between the emissions and corresponding activity data. An example is illustrated in Figure 16 for SO<sub>2</sub> emissions from the Cement industry in Italy. This pollutant, E-PRTR activity and country combination gave better R<sup>2</sup> values than many others, however the grouping of the data means that the line of best fit is heavily influenced by a limited portion of the dataset.

So, whilst the data does suggest that there is a reasonably consistent relationship between the emissions and the activity data, the uncertainty levels are high. Consequently it was decided not to take the investigation of the data on a time series basis any further.

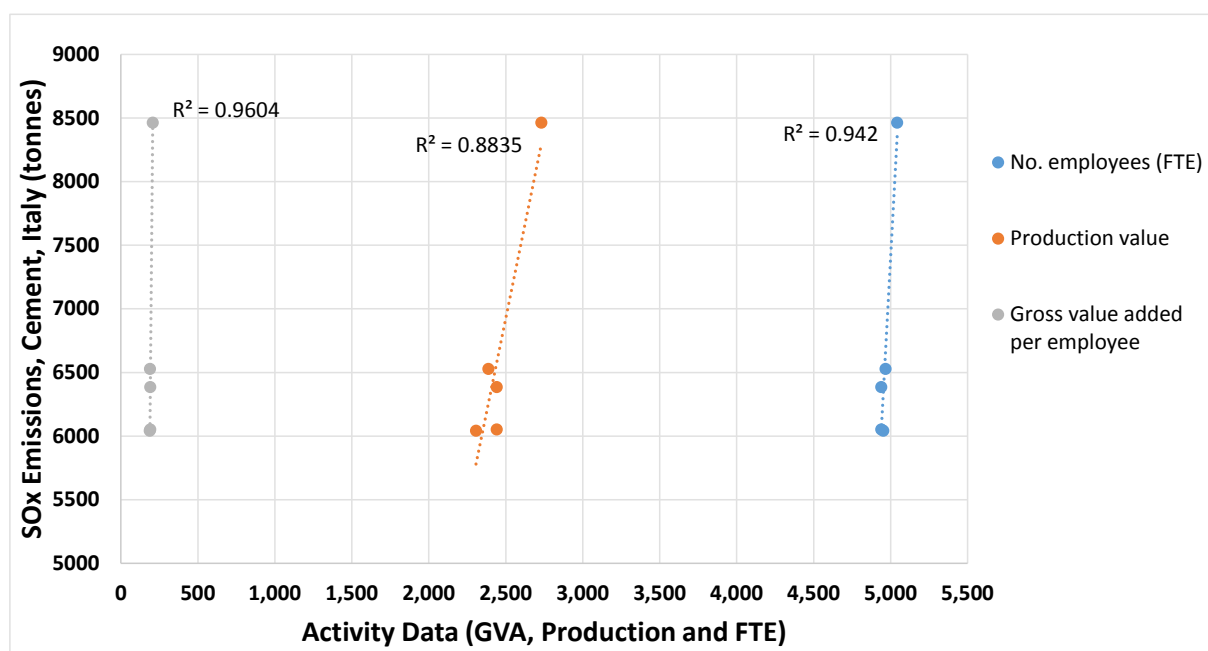


Figure 16 Emissions from the Italian cement industry versus activity data for 2008 – 2012

### 6.1.7 Conclusions on the analysis of 2011 emissions data

This relatively simple and short study has attempted to assess whether linear relationships exist between emissions and GVA, employment or production data. The time available for

this study has not allowed in-depth statistical analysis to be undertaken (although this is something that could be addressed in the future).

With the simple indicative analysis presented here, it is possible to identify the combinations of E-PRTR activity and economic or employment data that give the stronger correlations with each pollutant. It is also clear that the handling of data points where no emissions are reported in the E-PRTR, but where there is non-zero activity data, is important in establishing relationships between the different datasets.

The results here were considered to meet the intended aims of the initial scoping work, and hence it was considered sensible to apply this approach to study the 2012 emissions data.

In general, the use of employment data provides stronger relationships with emissions from the pollutants and E-PRTR activity investigated compared to the use of GVA data. However there are some cases (such as SO<sub>2</sub> and cement) where the use of GVA data provides a better relationship with E-PRTR emissions data.

The datasets for PCBs and PAHs were noted to be particularly poor. This may be due to plant emissions being below the reporting threshold, and therefore the emissions reported to the E-PRTR not being particularly representative of the total emissions from the corresponding industrial sectors for a country.

The data for some countries were noted to be repeatedly among the outliers. There are several reasons why this might be the case, and these are explored further in the results from applying this analysis to the 2012 E-PRTR datasets.

### **6.1.8 Methodology for the analysis of 2012 emissions data**

Emissions data for 2012 from the E-PRTR were processed using the same approach as described for the 2011 emissions dataset (see sections 6.1.3 and 6.1.4). However, the focus of the data analysis was to make a clear quantifiable identification of points which were considered to be outliers.

Emissions data were taken from the 2012 E-PRTR dataset, and activity data for 2012 were downloaded from the Eurostat website. The project schedule had to be postponed until these data were available from Eurostat, and even when they were published (early in October), the datasets were noted to be far from complete.

The input 2012 datasets were sourced as indicated in Table 21 below.

**Table 21 Sources of 2012 input data**

Input Dataset	Source
PRTR Emissions	E-PRTR dataset for 2012: eprtr_v5 database downloaded from EEA <a href="http://www.eea.europa.eu/data-and-maps/data/member-states-reporting-art-7-under-the-european-pollutant-release-and-transfer-register-e-prtr-regulation-8">http://www.eea.europa.eu/data-and-maps/data/member-states-reporting-art-7-under-the-european-pollutant-release-and-transfer-register-e-prtr-regulation-8</a>
Gross Value Added (GVA)	V12150: Downloaded from Eurostat website (October 2012). Value added at factor cost, Eurostat database (sbs_na_ind_r2)
Employment, as full time equivalent (FTE)	V16140: Downloaded from Eurostat website (October 2012). Number of employees in full-time equivalents, Eurostat database (sbs_na_ind_r2)
Industrial Production data (Production)	PRODCOM: Downloaded from Eurostat website (October 2012). Production data by NACE code.

The 2012 data were processed, and scatter plots of emissions versus activity data were generated for the same pollutants and E-PRTR activities as generated for the 2011 datasets. However some changes were made to the statistical analysis.

- A limited number of pollutant/activity combinations were investigated further in terms of regression analysis to determine statistical outliers. The selection was based on the work conducted on the 2011 data, and were the combinations shown to give the more robust correlations between emissions and activity.
- The line of best fit was forced through the origin. There is an inherent assumption in the statistical analysis that zero emissions will result from zero activity.
- Points falling on the y-axis were not included in the statistical analysis. These data points represent gaps in the GVA, employment and production datasets, and were excluded to avoid skewing the line of best fit.
- Points falling on the x-axis (zero emissions, but non-zero GVA or employment or production data) were included in the identification of outliers. These points may be illustrative of under or non-reporting to the E-PRTR. However it is possible that they arise from E-PRTR activities where all relevant facilities give emissions which are below the reporting threshold (resulting in zero emissions, but non-zero economic activity indicators).
- The data analysis was extended to include upper and lower confidence intervals on the line of best fit set at 99%. This offers a wider uncertainty range than applying a standard 95% confidence interval and was applied in order to account for the associated uncertainties in the data points, and the relationships being analysed as identified in the 2011 analysis. The value and statistical method requires review and improvement in the future because a relatively high proportion of the datasets are flagged as being outliers.

After processing the data in this way, it is possible to identify outliers according to one of three different criteria:

- **Upper outlier:** The data point represents an emission that is higher than that of the upper confidence interval. These points were not included in the data collated for the Party, as the intention is to identify data that may be missing from the E-PRTR.
- **Lower outlier:** The data point represents an emission that is lower than that of the lower confidence interval.
- **Zero emission:** No emissions are reported for the corresponding E-PRTR activity, but the GAV, employment or production data indicates that these activities do occur in the country.

As explained above, there may be valid reasons for the emissions at the E-PRTR activity level to be lower than that of the lower confidence interval. However, it is appropriate to collate these data and request an explanation from the relevant Party.

## 6.2 Conclusions / next steps

This methodology is still in the stages of development. However it has identified some outliers which warrant further investigation.

It is suggested that members of the ETC, or the EEA, approach country representatives and present the findings of this and other studies on the E-PRTR datasets. The results can be accompanied by a request that, for each occurrence, they explain the reasons for either zero emissions reporting or reporting of emissions that are considerably lower than is evident from other countries. It may be helpful for the Parties to compare and contrast the national total emissions of these sources with the data reported to the E-PRTR. This might help to indicate whether a substantial fraction of the sectoral emissions fall below the E-PRTR reporting threshold – or perhaps cases where there is genuine under-reporting to the E-PRTR.

This work has shown that there is value in using both the GVA and employment data with the E-PRTR emissions data to generate indicators that help to identify anomalies with the reported emissions data, and in particular help to identify where there may be occurrences of non-reporting. Discussions with countries can therefore focus on the suspected areas of non-reporting, allowing resources to be used more efficiently.

However, it must be acknowledged that this assessment has been relatively simple in its approach, and hampered by poor data capture in Eurostat activity datasets. It would be helpful to further refine the methodology, and also introduce automation as far as possible, to help with the identification of outliers and non-reporting.

It is recommended that this work is continued and developed. A number of possible improvements warrant further investigation, and are listed below. They are listed in decreasing order of priority, with a qualitative indication of the efforts required:

1. **Use of more complete or larger datasets** (high priority, average investment): The poor data capture in the Eurostat datasets has reduced the extent to which the emissions data can be assessed. It may be necessary to investigate other sources of data, which can also be made available to timescales which better suit the timelines of this project. In addition, it would be simple to extend the regression analysis to include points across all of the years of the E-PRTR data. This would help to better establish the relationships between the emissions and the GVA, employment and production data.
2. **Screening data for emissions below the E-PRTR reporting thresholds** (high priority, small investment): Some data points are flagged because zero emissions have been reported. However, it may be that the line of best fit suggests that the E-PRTR activity emissions are predicted to be below the E-PRTR reporting threshold. These data should be identified and filtered out of the final outlier dataset.
3. **The use of other statistical tools/analysis** (high priority, medium investment): The analysis presented here has used simple lines of best fit, regression analysis and standard deviations to identify outliers. There are other statistical tools and approaches which may be more appropriate for identifying outliers.
4. **Analysis at a more detailed spatial resolution** (medium priority, large investment): It may be possible to improve the spatial resolution of the analysis with employment data. Regional employment data is available (NUTS 2 and 3), and this could be used to conduct the analysis on a regional scale rather than at a national scale. This type of approach could then be linked with the work of other teams in this Topic Centre task, which have been conducted at a plant by plant level. This might help to identify plants which have emissions that fall below reporting thresholds for some years.
5. **The use of geographical clustering** (lower priority, medium investment): Data analysis could consider whether some form of geographical clustering could be used with the national level data. In determining the expected relationship between emissions and an indicator dataset for a specific country, more weight could be given to neighbouring countries. These neighbouring countries are expected to have industrial activities that are more comparable than countries which are further away. A larger dataset would be needed for this type of analysis.
6. **The use of non-linear relationships** (lower priority, small investment): All analysis here has assumed a linear relationship between emissions and GVA, employment or production data. It may be that simple non-linear functions provide a more reliable assessment e.g. to account for economies of scale. However, larger datasets are required to test this.

## References

EMEP (2009): *EMEP/EEA air pollutant emission inventory guidebook 2009*. EEA Technical report No 9/2009. <http://www.eea.europa.eu/themes/air/emep-eea-air-pollutant-emission-inventory-guidebook>

IPCC (2006): *2006 IPCC Guidelines for National Greenhouse Gas Inventories*. <http://www.ipcc-nggip.iges.or.jp/public/2006gl/>

ETC/ACM (2011): *E-PRTR Cross Pollutant Check Methodology*, ETC ACM working paper 2011/18, December 2012. [http://forum.eionet.europa.eu/etc-acm-consortium/library/subvention-2014/task-deliveries-ap2014/1211-e\\_prtr-review-work-and-support/kick-meeting/e-prtr-cross-pollutant-check-methodology-background-report-and-manual-v3](http://forum.eionet.europa.eu/etc-acm-consortium/library/subvention-2014/task-deliveries-ap2014/1211-e_prtr-review-work-and-support/kick-meeting/e-prtr-cross-pollutant-check-methodology-background-report-and-manual-v3) (portal with access restrictions)

EU (2006a): *Regulation (EC) No 166/2006 of the European Parliament and of the Council of 18 January 2006 concerning the establishment of a European Pollutant Release and Transfer Register and amending Council Directives 91/689/EEC and 96/61/EC*. Available at: <http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32006R0166&from=EN>

EU (2006b): *Guidance Document for the implementation of the European PRTR*. 31 May 2006. [http://prtr.ec.europa.eu/docs/EN\\_E-PRTR\\_fin.pdf](http://prtr.ec.europa.eu/docs/EN_E-PRTR_fin.pdf)  
(or [http://ec.europa.eu/environment/industry/stationary/eper/pdf/en\\_prtr.pdf](http://ec.europa.eu/environment/industry/stationary/eper/pdf/en_prtr.pdf) (from web page: <http://ec.europa.eu/environment/industry/stationary/eper/implementation.htm>))

Umweltbundesamt et al. (2012): *Final report – Three years of implementation of the E-PRTR. Supporting study for the European Commission*. Reference: ENV.C.3/SER/2010/0056. [http://ec.europa.eu/environment/industry/stationary/eper/pdf/Final%20report\\_20120605.pdf](http://ec.europa.eu/environment/industry/stationary/eper/pdf/Final%20report_20120605.pdf)

## Abbreviations

EC.....	European Commission
EEA.....	European Environment Agency
EMEP .....	Co-operative programme for monitoring and evaluation of the long-range transmissions of air pollutants in Europe
E-PRTR .....	European Pollutant Release and Transfer Register
ETC/ACM .....	European Topic Centre on Air Pollution and Climate Change Mitigation
ETC/SCP .....	European Topic Centre on Sustainable Consumption and Production
ETC/W.....	European Topic Centre on Water
EU .....	European Union
IPCC.....	Intergovernmental Panel on Climate Change
LCP .....	Large combustion plant
NACE.....	Nomenclature statistique des activités économiques dans la Communauté Européenne - a European industry standard classification and nomenclature system consisting of a six digit code
NECD.....	National Emission Ceilings Directive
PRODCOM.. .....	survey, with an at least annual frequency, for the collection and dissemination of statistics on the production of industrial (mainly manufactured) goods, both in value and quantity terms, in the European Union (EU). It is abbreviated from the French <b><i>Production Communautaire</i></b> .
RT.....	Release/transfer
RU .....	Ratio uncertainty



## Annex 1 Manual applying cross pollutant checks on national and European databases

*The MS-Excel workbooks and MS-Access databases coming with this document are included separately on the internet web page where this paper is released, aiming to assure easy access to the materials covered in this paper. Annotations of file, sheet, table and query names used this Annex follow as much as possible those use in these materials.*

### Checking a national E-PRTR dataset, including the E-PRTR conversion tool

This section explains the use of the Cross Pollutant Check and the Conversion Tool database together, in order to execute the emission checks on a national dataset.

The Cross Pollutant Check has been developed with the purpose of being integrated with the Conversion Tool to allow for a simple, yet consistent method of checking emissions per facility.

The available data consists of a database **Cross\_Pollutant\_Check\_for\_Conversion\_tool.mdb** and a spreadsheet **Cross\_Pollutant\_Check.xls**, both available from the EEA website and at the web page with this paper.

The E-PRTR Conversion Tool is made public in the form of an MS-Access database. It is meant to be used as a tool for member states to validate their national facility reports before they upload to the Central Data Repository (CDR).

Information on how to use the Checking and Conversion tool and which system requirements it needs can be found at:

<http://www.eionet.europa.eu/schemas/eprtr/EPTRUserManual.pdf>

The tool itself has to be downloaded from:

<http://www.eionet.europa.eu/schemas/eprtr/conversiontool>

Before performing the Cross Pollutant Check, the cross pollutant check data need to be imported into the country's Conversion Tool database. After that, the predefined queries in the Conversion Tool database can be used to perform the cross pollutant check.

Follow these steps:

## A) Import the E-PRTR Cross Pollutant Check to the country's conversion tool database

- 1) Make sure that the country's Conversion Tool database **EPRTTRConversionTool.mdb**, the database **Cross\_Pollutant\_Check\_for\_Conversion\_tool.mdb** and the spreadsheet **Cross\_Pollutant\_Check.xls** are downloaded and accessible.

It is advised to save the files in the same directory:

EPRTTRConversionTool.mdb
Cross_Pollutant_Check_for_Conversion_tool.mdb
Cross_Pollutant_Check.xls

- 2) Open the country's Conversion Tool database **EPRTTRConversionTool.mdb**.
- 3) Import the spreadsheet **Cross\_Pollutant\_Check.xls** as a linked table into the Conversion Tool database **Cross\_Pollutant\_Check\_for\_Conversion\_tool.mdb**.  
When using the import wizard make sure to:
  - a. Allow files of the type spreadsheet to be linked
  - b. Select the worksheet '**Cross\_pollutant\_check**',
  - c. tick 'First Row Contains Column Headings' and
  - d. confirm the suggested linked table name: 'Cross\_Pollutant\_Check' (please pay attention to the underscores between the words).
- 4) Import the queries from the database **Cross\_Pollutant\_Check\_for\_Conversion\_tool.mdb** into the country's **EPRTTRConversionTool.mdb** database.  
The checks are predefined in the queries and are named:

Name of query
01_select_emissions_per_facility_1_sourcepollutant
01_select_emissions_per_facility_2_conditionpollutant
01_select_emissions_per_facility_3_resultingpollutant
02_coupling_with_cpc_NACE
03_coupling_with_cpc_NACE_null
04_emissions_check_1
05_emissions_check_2
06_emissions_check_3
07_emissions_check_4
08_end_result

## B) Perform the E-PRTR Cross Pollutant Check

To perform the E-PRTR Cross Pollutant Check itself, the only query in database **EPRTRConversionTool.mdb** that one needs to execute is query **'08\_end\_result'**. (The other queries facilitate conditions for properly obtaining the results in query **'08\_end\_result'** and it is not necessary to execute them separately).

Query **'08\_end\_result'** provides the emissions per facility that – according the checking criteria defined in the imported excel sheet – are out of plausible range and better be subject to further scrutiny and inquiries on possible incompleteness.

Other reported emissions **not** presented in query **'08\_end\_result'** are either not checked, or within the expected emission ranges as predefined in the **Cross\_Pollution\_Checks.xls**.

The following information is displayed in the query **'08\_end\_result'**:

- The company ID, company name, facility name, city, the previous emission reporting year, the economic activity (Annex and NACE) and the medium in which the emissions are released (columns 1 to 9);
- The checked pollutants (ResultingPollutant) and their reported non-accidental emission quantity (ReportedResultingEmission), columns 10 and 11;
- On the right side of the reported emission of the checked pollutants the expected minimum and maximum emission of the checked pollutants is shown (columns 12 and 13) and the source pollutant which is used to derive the expected minimum and maximum emission (column 14). Next to these, in column 15, the reporting threshold is displayed;
- The final column, named 'CheckTotal', shows whether the emission of the checked pollutant is beneath the minimum or above the maximum of the calculated emission ranges.

If the emission is flagged as being beneath the minimum of the range, then it means that either no emission was reported (however an emission above threshold is expected), or that an emission was reported, however it was below the expected range.

If the emission is flagged as being higher than the maximum, then it means that the reported emission exceeds the calculated range.

Both occurrences should be an incentive to initiate further inquiries about the emissions at the facilities in question.

## Checking the European E-PRTR data base

This section explains how to use the Cross Pollutant Check in the downloadable European E-PRTR data set published on the EEA web site, in order to execute the cross pollutant emission checks on all European data and for all reporting years.

The Cross Pollutant Check has been developed with the purpose of being integrated into the European data set to allow for a simple, yet consistent method of checking emissions per facility.

The European data set is made public as a MS-Access database (as '**eptr\_v<x.x>.mdb**') and as a text formatted file ('**eptr\_v<x.x>.csv**'), both with a versioning code v<x.x> (e.g., v6.2) representing the sequence of releases at the EEA web site.

The data is available at the EEA web site:

[http://www.eea.europa.eu/data-and-maps/data/ds\\_resolveuid/34bb6ad353744412a1704902faf3497b](http://www.eea.europa.eu/data-and-maps/data/ds_resolveuid/34bb6ad353744412a1704902faf3497b)

(permalink to the latest published version of the reported E-PRTR data for Europe)

The database contains the E-PRTR data covering the latest reporting from 2007 to 201x by EU Member States, Iceland, Liechtenstein, Norway, Switzerland (EFTA-4) and Serbia.

The available data for checking reference consists of a database called **Cross\_Pollutant\_Check\_for\_European\_data\_set.mdb** and a spreadsheet called **Cross\_Pollutant\_Check.xls**, both available from the EEA web site and at the web page with this paper.

Before performing the Cross Pollutant Check, the cross pollutant check data stored in database **Cross\_Pollutant\_Check\_for\_European\_data\_set.mdb** and in the spreadsheet **Cross\_Pollutant\_Check.xls** need to be imported into the latest European E-PRTR data base **eptr\_v<x.x>.mdb**. After that, the predefined queries originally stored in **Cross\_Pollutant\_Check\_for\_European\_data\_set.mdb** can be used to perform the cross pollutant check.

Follow these steps:

### **A) Import the E-PRTR Cross Pollutant Check to European data set (e.g. eptr\_v6.2.mdb)**

- 1) Make sure that the European data set database **eptr\_v<x.x>.mdb**, the Cross Pollutant Check database **Cross\_Pollutant\_Check\_for\_European\_data\_set.mdb** and the Cross Pollutant Check spreadsheet **Cross\_Pollutant\_Check.xls** are downloaded and accessible. It is advised to save the files in the same directory.

eptr_v<x.x>.mdb
Cross_Pollutant_Check_for_European_data_set.mdb
Cross_Pollutant_Check.xls

- 2) Open the European data base **eptr\_v<x.x>.mdb**.
- 3) Import the cross pollutant check spreadsheet as a linked table into the European data set. When using the import wizard make sure to:
  - a. Allow files of the type spreadsheet to be linked
  - b. Select the worksheet '**Cross\_pollutant\_check**',
  - c. Tick 'First Row Contains Column Headings'
  - d. Confirm the suggested linked table name: 'Cross\_pollutant\_check'
- 4) Import the queries from the database **Cross\_Pollutant\_Check\_for\_European\_data\_set.mdb** into European database **eptr\_v<x.x>.mdb**.

The queries in which the checks are prepared are named:

Name of query
01_select_emissions_per_facility_1_sourcepollutant
01_select_emissions_per_facility_2_conditionpollutant
01_select_emissions_per_facility_3_resultingpollutant
02_coupling_with_cpc_NACE
03_coupling_with_cpc_NACE_null
04_emissions_check_1
05_emissions_check_2
06_emissions_check_3
07_emissions_check_4
08_end_result

## B) Perform the E-PRTR Cross Pollutant Check

To perform the E-PRTR Cross Pollutant Check itself, the only query that in database **eptr\_v<x.x>.mdb** that one needs to execute is query **'08\_end\_result'**.

(The other queries are used to facilitate the result in query **'08\_end\_result'** and it is not necessary to execute them separately).

Query **'08\_end\_result'** provides the emissions per facility that – according the checking criteria predefined in the imported excel sheet - are out of plausible range and better be subject to further scrutiny and inquiries on possible incompleteness.

Other reported emissions **not** presented in query **'08\_end\_result'** are either not checked, or are within the expected emission ranges as predefined in the **Cross\_Pollution\_Checks.xls**.

The following information is displayed in the query **'08\_end\_result'**:

- The company ID, company name, facility name, city, country, reporting year, the main economic activity (Annex and NACE) and the medium in which the emissions are released (columns 1 to 10);
- The checked pollutants (ResultingPollutant) and their reported non-accidental emission quantity (ReportedResultingEmission), columns 11 and 12;
- On the right side of the reported emission of the checked pollutants the expected minimum and maximum emission of the checked pollutants is shown (columns 13 and 14) and the source pollutant which is used to derive the expected minimum and maximum emission (column 15). Next to these, in column 16, the reporting threshold is displayed;
- The final column, named **'CheckTotal'**, shows whether the emission of the checked pollutant is beneath the minimum or above the maximum of the calculated emission ranges.

If the emission is flagged as being beneath the minimum of the range, then it means that either no emission was reported (however an emission above threshold is expected), or that an emission was reported, however it was below the expected range.

If the emission is flagged as being higher than the maximum, then it means that the reported emission exceeds the calculated range.

Both occurrences should be an incentive to initiate further inquiries about the emissions at the facilities in question.