Benchmarks to determine baselines for mitigation action under the Article 6.4 mechanism

Discussion Paper

Umwelt Bundesamt

DEHSt
Deutsche
Emissionshandelsstelle

Editorial information

Publisher
German Emissions Trading Authority (DEHSt)
at the German Environment Agency
Bismarckplatz 1
D-14193 Berlin
Phone: +49 (0) 30 89 03-50 50
Fax: +49 (0) 30 89 03-50 10
emissionstrading@dehst.de
Internet: www.dehst.de/EN

Status: April 2019

Authors
Jürg Füssler, Quirin Oberpriller (INFRAS), Vicky Duscha, Sascha Lehmann, Marlene Arens (Fraunhofer ISI)

Performing Organisation
INFRAS
Binzstrasse 23
8045 Zürich, Switzerland

Fraunhofer Institute for Systems and Innovation Research ISI
Breslauer Str. 48
76139 Karlsruhe, Germany

Wuppertal Institute for Climate, Environment, Energy
Döppersberg 19
42103 Wuppertal, Germany

On behalf of the German Environment Agency
Completion date March 2019

Environmental Research of the Federal Ministry for the Environment, Nature Conservation,
Building and Nuclear Safety
Project number: 3717 42 504 0

Cover image: Tkemot/Shutterstock.com

Abstract

Baseline setting with globally applicable and stringent benchmarks may be an important instrument for scaling-up market mechanisms under Article 6. Defining baselines based on BAU and NDC targets is challenging as targets are uncertain, sometimes unclear or have only a limited scope. Benchmarks promise to be an efficient and simple solution. This builds on the rationale that stringent benchmarks lead to baselines that are automatically below both BAU and to an emission trajectory that is compliant with the host-country's NDC target. This analysis indicates that even though there are sub-sectors with medium or large potential for benchmarking (e. g. industry process emissions), most emission sources cannot be covered by global benchmarks. The reason is that goods and services are heterogeneous, and emissions tend to depend on exogenous local factors. Furthermore, if stringent benchmarks are used in many sectors expected carbon prices may not be on a level that would trigger additional action. Benchmarking is therefore barely the silver bullet with respect to crediting baselines under the Paris Agreement. However, if one shifts the focus from global benchmarks towards a more efficient and robust approach of setting project-specific crediting baselines, the large body of methodological approaches and reference values from emission trading systems and the CDM can be of use nevertheless.

Kurzbeschreibung

Die Festlegung von Baselines mit weltweit anwendbaren und strengen Benchmarks kann ein wichtiges Instrument zur Ausweitung der Marktmechanismen gemäß Artikel 6 sein. Die Definition von Baselines auf Basis von BAU- und NDC-Zielen ist eine Herausforderung, da die Ziele unsicher, manchmal unklar oder nur einen geringem Geltungsbericht haben. Benchmarks versprechen hierfür eine effiziente und einfache Lösung. Der Grund ist, dass strenge Benchmarks zu Baselines führen, die automatisch sowohl unterhalb der BAU liegen, als auch zu einer Emissionstrajektorie, die dem NDC-Ziel des Gastlandes entspricht. Auch wenn es Teilbereiche mit mittlerem oder großem Benchmarking-Potenzial gibt (z. B. Emissionen aus industriellen Prozessen), zeigt diese Analyse, dass die Mehrheit der Emissionsquellen nicht durch globale Benchmarks abgedeckt werden kann. Waren und Dienstleistungen sind heterogen und die Emissionen hängen tendenziell auch von exogenen lokalen Faktoren ab. Darüber hinaus dürfte der erwartete CO_2-Preis in vielen Sektoren nicht auf einem Niveau liegen, das zusätzliche Maßnahmen auslösen würde, wenn strenge Benchmarks verwendet werden. Benchmarking ist daher kaum der Königsweg, um die Probleme der Baseline-Setzung im Rahmen des Paris Abkommens zu lösen. Wenn man jedoch den Fokus von globalen Benchmarks auf einen effizienteren und robusteren Ansatz zur Festlegung projektspezifischer Baselines verlagert, kann der große Bestand von methodischen Ansätzen und Referenzwerden aus den Emissionshandelssystemen und dem CDM dennoch von Nutzen sein.

Content

List of Tables

Tabellenverzeichnis

List of Figures

Abbreviations

CBA	Cost-benefit analysis
BAT	Best Available Technology
BAU	Business as usual
BOD	Biological Oxygen Demand
BF	Blast furnace
BOF	Basic oxygen furnace
BREF	Best Available Techniques Reference Documents
CAFE	Corporate Average Fuel Economy
CBDRC	Common but differentiated responsibilities and respective capabilities
CCS	Carbon Capture and Storage
COD	Chemical oxygen demand
CDM	Clean Development Mechanism
DEHSt	Deutsche Emissionshandelsstelle
DRI	Direct reduced iron
CHP	Combined heat and power
CO_2	Carbon dioxide
CO_2eq	CO_2 equivalent
CSI	Cement Sustainability Initiative
EAF	Electric arc furnace
ETS	Emission Trading System
EU ETS	European Emission Trading System
GHG	Greenhouse Gases
GNR	Getting the numbers right
GWP	Global Warming potential
HBEFA	Handbook Emission Factors for Road Transport
HFC-23	Fluoroform
IED	Industrial Emissions Directive
IPCC	Intergovernmental Panel on Climate Change
IPPC	Integrated Pollution Prevention and Control
IRR	Internal Rate of Return
LCTPI	Low Carbon Technology Partnerships initiative
MRV	Measurement, Reporting, Verification
N_2O	Nitrous oxide
NDC	Nationally determined contribution
PFC	Perfluorcarbon
PMR	World Bank's Partnership for Market Readiness
UNEP	United Nations Environment Programme
UNFCCC	United Nations Framework Convention on Climate Change
WSA	World Steel Association
WWTP	Waste water treatment plants

Summary

Baseline setting with globally applicable and stringent benchmarks may be an important instrument for scaling-up market mechanisms under Article 6. Defining baselines based on business-as-usual (BAU) and nationally determined contributions (NDC) targets is challenging as targets are uncertain, sometimes unclear or have only a limited scope. Benchmarks promise to be an efficient and simple solution. This builds on the rationale that stringent benchmarks lead to baselines that are automatically below both BAU and to an emission trajectory that is compliant with the host-country's NDC target. The analysis focuses on global benchmarks in the sense of simple reference values for specific products and services that are independent of a specific country and may be applied globally or on the level of groups of countries (e. g. low/middle/high income countries).

Benchmarking is a term widely used and describes a standard or set of standards that can be used as a point of reference for evaluating performance or level of quality against peers. In the context of Article 6.4., we consider benchmarking as a comparison of the performance with respect to either GHG or CO_2 emissions and in some cases also energy.

Different levels can be used to define a benchmark. Examples are the average performance level, the average of the top 20%/10%/ x% best performers (e. g. applied under the EU ETS (10%) and partly in CDM (20%)), the best achieved level or the best available level (see e. g. pmr 2017).

Benchmarks are often used as a management tool to monitor company performance, but they are also used more and more often in energy and climate policy. The best-known example of this is the EU ETS, for which 59 product benchmarks have been developed. These benchmarks are used to determine the free allocation of allowances to industrial installation operators. Fallback benchmarks for emissions from heat and fuel consumption have also been developed. Other countries such as South Korea and Switzerland also use benchmarks in their emissions trading systems.

In contrast to emissions trading, where a benchmark is used to determine the number of allowances allocated, in a crediting system the benchmark is used to define the baseline emissions. The amount of credits that can be issued is subsequently determined by the difference between the benchmark and the installation's performance (which needs to be monitored). The fact that benchmarks are being used to define the baseline (and not the standard that should be achieved), also implies that some of the existing benchmarks are not suitable in this context. For example, a benchmark based on best available technology may not provide sufficient leeway to credit further emission reductions compared to that benchmark.

Paragraph 48c of the Modalities and Procedures for the CDM today explicitly allows for the use of benchmarking. However, it is rarely used in reality. Yet, in contrast to the CDM, where each project provides its own baseline based on the specifications within the methodology, the use of benchmarks for the definition of a baseline could increase transparency, reduce administrative costs and may help to increase environmental integrity of the crediting system by preventing over-crediting.

Several different potential sources for benchmark values may be considered:

- Best Available Techniques reference documents that have been adopted under both the IPPC Directive (2008/1/EC) and later under the Industrial Emissions Directive (IED, 2010/75/EU).
- Data from CDM projects for the relevant sectors.
- Product benchmarks, in the context of the EU ETS.
- Energy efficiency standards for specific jurisdictions.
- Data on carbon intensity as collected by sectoral organizations, such as in the Cement Sustainability Initiative or data on upstream emissions from the oil industry.

For the development of benchmarks under the EU ETS, 11 guiding principles were developed in advance (Ecofys and ISI 2008). They characterise what – from a theoretical point of view – good benchmarks should respect. However, while at first glance the EU ETS benchmarks and hence the underlying guiding principles seem to provide a good starting point for the discussion, in detail one faces difficulties in the application of benchmarks in the context of international crediting (rather than linking of ETS). First, not all installations, but only some, either new or significantly improved installations will apply the benchmark on a voluntary basis (self-selection of installations). Second, applying similar benchmarks in all countries may not necessarily be the best solution. For products competing in the global market, a level playing field seems indeed appropriate. However, in the international context, differences in the countries' national circumstances and the UNFCCC's principle of common but differentiated responsibilities and respective capabilities may call for factoring in the national context in the definition of the benchmark.

Based on DEHSt (2013), the following general criteria are necessary for the development of a benchmarking approach in an international context:

- Clear definition of system boundaries
- Adequate definition of key performance indicator/ benchmark
- Availability of data for determination of benchmark level
- Similar benchmarks for similar products in similar countries
- Benchmark levels should incentivise investment in no- and low-carbon technologies
- Improvement over time
- Availability of data for performance evaluation

In addition to these general criteria, the use of benchmarks under Article 6.4 also necessitates specific requirements. This means that the crediting baseline must not only be below BAU emissions, but also in line with national NDC targets. The benchmark must also be consistent with the Paris Agreement, which means that the benchmark should not incentivize the construction of plants that are not consistent with the emission reduction pathway necessary to meet the Paris agreements. Furthermore, the benchmark should also be applicable to a wide range of countries or different benchmarks will be developed for different categories of countries (e. g. by income). And finally, there should be a transparent and science-based process for determining benchmarks.

The main text looks at data sources and discusses possible types of benchmarks for the most relevant areas (Industry including energy use, Energy generation, Housing, Transport and Waste water management).
In addition, there are three detailed case studies on Cement, Steel and Waste Water Treatment.

The following table evaluates important criteria that define the suitability of benchmarks for selected sectors. The criteria are: the availability of data (for activity levels and benchmark data); the availability of global benchmark (vs. the need to differentiate them for specific countries or regions); and the carbon market contribution to profitability (in order to focus on projects that are additional compared to what would otherwise occur). Adequate levels of data availability are crucial, be it for benchmarks that are determined purely rule based (by use of a formula, e. g. as a certain percentile of the market performance) or by expert judgement (considering ambition and technological leaps). In general, data availability is much more limited in developing countries, in more informal industries and the residential sector. The Paris Agreement requires also the regular updating of benchmark values, for instance in sync with the 5-year NDC cycle to prevent lock-ins into technologies that may not be in line with the long- term goal of the Paris Agreement.

Table 1: Evaluation of criteria that define the suitability of benchmarks for selected sectors

(Sub-) sector	Activity data availability	Benchmark data availability	Availability of global benchmarks	Carbon market contribution to profitability
Industry energy use – product benchmarks	***	**	**	*
Industry energy use – other benchmarks	**	*	*	*
Industry process emissions	***	**	***	***
Energy generation	***	**	**	*
Housing	**	*	*	*
Transport – general	*	*	*	*
Transport – fuel efficiency standards	**	**	**	*
Waste water	**	*	*	***

key: * = low, ** = medium, *** = high
Source: Own analysis

The contribution of carbon markets to the profitability differs strongly between different sectors. For wind projects the impact of carbon revenues under the CDM added on average below three percentage points to its profitability (internal rate of return). On the contrary in e. g. landfill gas projects, where the high global warming potential of the avoided methane produces much more revenues from carbon markets, the impact on profitability is on average in the order of 14 to 15 percentage points (Cames et al. 2016). This problem is not benchmarking-specific. However, when identifying sectors that are most suitable for using benchmarks under Article 6.4, the economic attractiveness is be an important factor.

While the analysis indicates an overall limited potential for global benchmarks, there are some quick wins in the form of global benchmarks related to industry process emissions. Here, the CDM has established robust and stringent benchmarks for baseline setting e. g. in N_2O abatement in nitric acid or adipic acid production, or for abatement of HFC23 emissions in the production of refrigerants. it may also be assumed that with these high GWP gases, the revenues from the transfer of emission reductions may provide a significant contribution to overall profitability and therefore lead to mitigation action beyond BAU.

In our case studies we derived that some other industries may be suitable for benchmarking, including cement or iron and steel. However, related emissions depend on local factors (such as quality of raw materials) and are thus difficult to implement on a global level. Here, baseline setting with approaches of intermediary complexity may be possible, building on proposed or approved CDM methodologies and EU-ETS guidance for product benchmarks. In practice, expected carbon prices may not be on a level that would trigger additional action in these sectors.

Also, the process that leads to the definition of benchmark values may be challenging to implement under an Article 6.4 mechanism. Providing benchmarks may open the door for loopholes or non-stringent values may result. A stringent and science-based process within the Article 6.4 supervisory body should facilitate the definition of adequate global benchmarks. In settings of weak governmental oversight, using benchmarks may be less adequate than conventional methodologies of baseline setting, where baselines are set on the basis of project specific parameters that are validated by independent third parties.

To summarize: Even though there are sub-sectors with medium to large potential for benchmarking, most emission sources cannot be covered by global benchmarks, because the goods and services are heterogeneous (e. g. “shoes”, “tonne-kilometers”) and emissions tend to depend also on exogenous local factors. Benchmarking is therefore barely the silver bullet to solve the issues with crediting baseline setting under the Paris Agreement.

However, if one moves from global benchmarks more towards standardized approaches to baseline setting, there is a large body of methodological approaches and reference values from Emission Trading Schemes and the CDM that can be used to define crediting baselines in a more efficient and robust way. Their use under Article 6 requires their further development including comprehensive data collection exercises that would allow for standardized approaches taking into account at least some regional, local or project specific factors.

Zusammenfassung

Die Festlegung von Baselines mit weltweit anwendbaren und strengen Benchmarks kann ein wichtiges Instrument zur Ausweitung der Marktmechanismen gemäß Artikel 6 sein. Die Definition von Baselines auf Basis von Business-as-usual (BAU)- und „nationally determined contributions“ (NDC)-Zielen ist eine Herausforderung, da die Ziele unsicher, manchmal unklar oder nur einen geringem Geltungsbericht haben. Benchmarks versprechen hierfür eine effiziente und einfache Lösung. Der Grund ist, dass strenge Benchmarks zu Baselines führen, die automatisch sowohl unterhalb der BAU liegen, als auch zu einer Emissionstrajektorie, die dem NDC-Ziel des Gastlandes entspricht. Die Analyse konzentriert sich auf globale Benchmarks im Sinne einfacher Referenzwerte für bestimmte Produkte und Dienstleistungen, die unabhängig von einem bestimmten Land sind und global oder auf der Ebene von Ländergruppen (z. B. Länder mit niedrigem, mittlerem oder hohem Einkommen) angewendet werden können.

Benchmarking ist ein weit verbreiteter Begriff und beschreibt einen Standard oder eine Reihe von Standards, die als Bezugspunkt für die Bewertung der Leistung oder des Qualitätsniveaus im Vergleich zu anderen verwendet werden können. Im Rahmen von Artikel 6.4 betrachten wir Benchmarking als Vergleich der Treibhausgas- oder CO_2-Emissionen und in einigen Fällen auch der Energie.

Zur Definition eines Benchmarks können verschiedene Ansätze verwendet werden. Beispiele sind das durchschnittliche Emissionsniveau, der Durchschnitt der besten 20 %/10 % oder x % (z. B. im Rahmen des EU-ETS (10 %) und zum Teil im CDM (20 %)), das beste erreichte Niveau oder das beste verfügbare Niveau (siehe z. B. pmr 2017).

Benchmarks werden häufig als Managementinstrument zur Überwachung der Unternehmensleistung eingesetzt, aber auch immer häufiger in der Energie- und Klimapolitik. Das bekannteste Beispiel dafür ist das EU-ETS, für das 59 Produktebenchmarks entwickelt wurden. Diese Benchmarks werden verwendet, um die kostenlose Zuteilung von Zertifikaten an Betreiber von Industrieanlagen zu berechnen. Darüber hinaus wurden Fallback-Benchmarks für Emissionen aus Wärme- und Kraftstoffverbrauch entwickelt. Auch andere Länder wie Südkorea und die Schweiz verwenden Benchmarks in ihren Emissionshandelssystemen.

Im Gegensatz zum Emissionshandel, bei dem ein Benchmark zur Bestimmung der Anzahl zugeteilter Zertifikate verwendet wird, wird in einem Kreditierungssystem der Benchmark zur Bestimmung der Baseline Emissionen verwendet. Die Höhe der kreditierten Emissionsreduktionen wird anschließend durch die Differenz zwischen dem Benchmark und den tatsächlichen Emissionen der Anlage bestimmt. Die Tatsache, dass Benchmarks zur Definition der Baseline (und nicht des zu erreichenden Standards) verwendet werden, bedeutet auch, dass einige der bestehenden Benchmarks in diesem Zusammenhang nicht geeignet sind. So bietet beispielsweise ein Benchmark auf der Grundlage der besten verfügbaren Technologie nicht genügend Spielraum, um weitere Emissionsreduktionen im Vergleich zu diesem Benchmark gutzuschreiben.

Paragraph 48c der Modalitäten und Verfahren für den CDM sieht heute ausdrücklich den Einsatz von Benchmarking vor. In der Realität wird er jedoch nur selten eingesetzt. Im Gegensatz zum CDM, bei dem jedes Projekt seine eigene Baseline auf der Grundlage der Spezifikationen innerhalb der Methodik bereitstellt, könnte jedoch die Verwendung von Benchmarks für die Definition einer Baseline von Vorteil sein: Sie kann die Transparenz erhöhen, die Verwaltungskosten senken und dazu beitragen, die ökologische Integrität der Kreditierung von Emissionsreduktionen zu erhöhen, indem eine zu hohe Vergabe verhindert wird.

Es können mehrere verschiedene potenzielle Quellen für die Bestimmung der Benchmark-Werte in Betracht gezogen werden:

Referenzdokumente zu den besten verfügbaren Techniken, die sowohl im Rahmen der IPCC Richtlinie (2008/1/EG) als auch in der späteren Industrie-Emissionsrichtlinie (IED, 2010/75/EU) verwendet wurden

Daten aus CDM-Projekten für die relevanten Sektoren

Produktebenchmarks im Rahmen des EU-ETS

Energieeffizienzstandards für bestimmte Regionen

Daten zur CO_2-Intensität, die von verschiedenen branchenspezifischen Organisationen erhoben werden, wie beispielsweise in der Cement Sustainability Initiative oder Daten zu Upstream-Emissionen der Ölindustrie.

Für die Entwicklung von Benchmarks im Rahmen des EU-ETS wurden im Vorfeld 11 Leitprinzipien entwickelt (Ecofys und ISI 2008). Sie charakterisieren, was gute Benchmarks – aus theoretischer Sicht – berücksichtigen sollten. Während die EU-ETS-Benchmarks und damit die zugrundeliegenden Leitprinzipien auf den ersten Blick einen guten Ausgangspunkt für die Diskussion zu bieten scheinen, stößt man im Detail jedoch auf Schwierigkeiten bei der Anwendung von Benchmarks im Rahmen der internationalen Emissionskreditierung (und nicht bei der Verknüpfung des EHS). Erstens werden nicht alle Anlagen den Benchmark auf freiwilliger Basis anwenden, sondern nur einige, entweder neue oder deutlich verbesserte Anlagen (d.h. eine Selbstselektion der Anlagen findet statt). Zweitens ist die Anwendung ähnlicher Benchmarks in allen Ländern möglicherweise nicht unbedingt die beste Lösung. Sicherlich erscheinen für Produkte, die auf dem Weltmarkt konkurrieren, gleiche Wettbewerbsbedingungen angemessen. Im internationalen Kontext können jedoch Unterschiede in den nationalen Gegebenheiten der Länder sowie der Grundsatz des UNFCCC „common but differentiated responsibilities and respective capabilities“ eine Berücksichtigung ded nationalen Kontext bei der Definition von Benchmarks erfordern.

Basierend auf der DEHSt (2013) sind folgende allgemeine Kriterien für die Entwicklung eines Benchmarking-Ansatzes im internationalen Kontext notwendig:

- Klare Definition der Systemgrenzen
- Angemessene Definition der Key Performance Indicators / Benchmarks
- Verfügbarkeit von Daten zur Bestimmung des Benchmark-Levels
- Ähnliche Benchmarks für ähnliche Produkte in ähnlichen Ländern
- Benchmark-Niveaus sollten Anreize für Investitionen in kohlenstofffreie und -arme Technologien schaffen.
- Verbesserung im Laufe der Zeit
- Verfügbarkeit von Daten für die Leistungsbewertung

Zusätzlich zu den allgemeinen Kriterien erfordert die Verwendung von Benchmarks nach Artikel 6.4 auch spezifische Anforderungen. Das bedeutet, dass die Baseline nicht nur unter den BAU-Emissionen liegen muss, sondern auch im Einklang mit den nationalen NDC-Zielen. Der Benchmark muss zudem auch mit dem Paris Abkommen übereinstimmen, was bedeutet, dass der Benchmark nicht den Bau von Anlagen anregen sollte, die nicht mit dem Emissionsreduktionspfad übereinstimmen, der zur Erfüllung des Paris Abkommen erforderlich ist. Darüber hinaus sollte der Benchmark auch für ein breites Spektrum von Ländern gelten, oder es werden für verschiedene Kategorien von Ländern unterschiedliche Benchmarks entwickelt (z. B. nach Einkommen). Und schließlich sollte es einen transparenten und wissenschaftlich fundierten Prozess zur Festlegung von Benchmarks geben.

Der Haupttext befasst sich mit den Quellen und möglichen Arten von Benchmarks für die wichtigsten Bereiche (Industrie einschließlich Energienutzung, Energieerzeugung, Wohnen, Verkehr und Wasserwirtschaft). Darüber hinaus gibt es drei detaillierte Fallstudien zu Zement, Stahl und Abwasserbehandlung.

Die folgende Tabelle bewertet wichtige Kriterien, welche die Eignung von Benchmarks für ausgewählte Branchen definieren. Die Kriterien sind: die Verfügbarkeit von Daten (für das jeweilige Aktivitätsniveau und Benchmarks); die Verfügbarkeit von globalen Benchmarks (Im Gegensatz zu der Notwendigkeit, sie für bestimmte Länder oder Regionen zu differenzieren); und der Beitrag des CO_2-Marktes zur Rentabilität (um sich auf zusätzliche Projekte zu konzentrieren). Entscheidend ist eine angemessene Datenverfügbarkeit, sei es für Benchmarks, die rein regelbasiert ermittelt werden (d.h. auf der Basis von Formeln, wie z. B. die Bestimmung der Perzentile der Marktperformance) oder durch Expertenurteil (unter Berücksichtigung von Ambitionen und Technologiesprüngen). Im Allgemeinen ist die Datenverfügbarkeit in Entwicklungsländern, in eher informellen Branchen und im Wohnungssektor viel geringer. Das Paris Abkommen verlangt auch die regelmäßige Aktualisierung der Benchmark-Werte, beispielsweise im Einklang mit dem fünfjährigen NDC-Zyklus, um Lock-in in Technologien zu verhindern, die möglicherweise nicht mit dem langfristigen Ziel des Paris Abkommens übereinstimmen.

Tabelle 1: Kriterien zur Eignung von Benchmarks für ausgewählte Branchen

(Teil-) Branche	Verfügbarkeit von Daten zum Aktivitätsniveau	Verfügbarkeit von Benchmarkdaten	Verfügbarkeit von globalen Benchmarks	Beitrag des CO_2 Markt zur Rentabilität
Energieverbrauch der Industrie – Produktbenchmarks	***	**	**	*
Energieverbrauch der Industrie – andere Benchmarks	**	*	*	*
Emissionen aus Industrieprozessen	***	**	***	***
Energieerzeugung	***	**	**	*
Wohnen	**	*	*	*
Transport – allgemein	*	*	*	*
Verkehr – Emissionsstandards	**	**	**	*
Abwasser	**	*	*	***

Legende: * = niedrig, ** = mittel, *** = hoch
Quelle: Eigene Analyse

Der Beitrag der CO_2-Märkte zur Rentabilität ist in den einzelnen Sektoren sehr unterschiedlich. Bei Windprojekten trugen die CO_2-Einnahmen im Rahmen des CDM durchschnittlich unter drei Prozentpunkten zur Rentabilität bei (interner Zinsfuss). Im Gegensatz dazu liegen die Auswirkungen auf die Rentabilität bei Deponiegasprojekten, bei denen das hohe Treibhauspotenzial des vermiedenen Methans deutlich mehr Einnahmen aus den CO_2-Märkten generiert, im Durchschnitt bei 14 bis 15 Prozentpunkten (Cames et al. 2016). Dieser Faktor ist nicht spezifisch für den Kontext des Benchmarkings. Bei der Identifizierung von Sektoren, die für die Verwendung von Benchmarks gemäß Artikel 6.4 am besten geeignet sind, kann jedoch die wirtschaftliche Attraktivität ein wichtiger Faktor sein.

Während die Analyse ein begrenztes Potenzial für globale Benchmarks offenbart, sind trotzdem einige quick wins in Form von globalen Benchmarks in Bezug auf die Emissionen von Industrieprozessen möglich. Hier hat der CDM robuste und strenge Benchmarks für die Baseline-Setzung z. B. bei der Reduktion Lachgas-Emissionen in der Salpetersäure- oder Adipinsäureproduktion oder bei der Reduzierung von HFC23-Emissionen bei der Herstellung von Kältemitteln festgelegt. Es ist auch davon auszugehen, dass bei solchen Gasen mit hohem Treibhauspotenzial die Einnahmen aus den Emissionsreduktionen einen wesentlichen Beitrag zur Gesamtrentabilität leisten und daher zu Minderungsmaßnahmen über den BAU hinausführen können.

Zudem könnten einige andere Branchen für Benchmarking geeignet sein, wie z. B. die Zementindustrie oder die Eisen- und Stahlindustrie. Die damit verbundenen Emissionen hängen jedoch stärker von lokalen Faktoren (wie der Qualität der Rohstoffe) ab und sind auf rein globaler Ebene schwieriger umzusetzen. Hier kann eine Baseline-Setzung mit Ansätzen von intermediärer Komplexität möglich sein, die auf vorgeschlagenen oder genehmigten CDM-Methoden und EU-ETS-Leitlinien für Produktebenchmarks aufbauen. In der Praxis liegt der erwartete CO_2-Preis jedoch möglicherweise nicht auf einem Niveau, um zusätzliche Maßnahmen in diesen Sektoren auszulösen.

Der Prozess, der zur Definition von Benchmarkwerten führt, kann im Rahmen eines Mechanismus nach Artikel 6.4 schwierig zu implementieren sein. Die Bereitstellung solcher Benchmarks kann zudem auch zu Schlupflöchern und nicht stringente Werte führen. Ein stringenter und wissenschaftlich fundierter Prozess innerhalb des Aufsichtsorgans von Artikels 6.4 sollte die Definition von angemessenen globalen Benchmarks ermöglichen. In Situationen schwacher staatlicher Kontrolle kann die Verwendung von Benchmarks weniger angemessen sein als herkömmliche Methoden der Baseline-Bestimmung, bei denen die Baseline auf der Grundlage projektspezifischer Parameter festgelegt wird, die von unabhängigen Dritten validiert werden.

Zusammenfassend lässt sich sagen, dass, auch wenn es Teilbereiche mit mittlerem und grossem Benchmarking-Potenzial gibt, die meisten Emissionsquellen nicht durch globale Benchmarks abgedeckt werden können, da die Waren und Dienstleistungen heterogen sind (z. B. „Schuhe“, „Tonnenkilometer“) und die Emissionen tendenziell auch von exogenen lokalen Faktoren abhängen. Benchmarking ist daher kaum der Königsweg, um die Probleme bei der Anrechnung der Baseline-Setzung im Rahmen des Paris Abkommens zu lösen.

Wenn man jedoch von globalen Benchmarks zu standardisierten Ansätzen für die Festlegung von Baselines übergeht, gibt es eine Vielzahl von methodischen Ansätzen und Referenzwerten aus Emissionshandelssystemen und dem CDM, mit denen sich Baselines effizienter und robuster definieren lassen. Ihre Verwendung nach Artikel 6 erfordert ihre Weiterentwicklung einschließlich umfassender Datenerhebungen, die standardisierte Ansätze unter Berücksichtigung zumindest einiger regionaler, lokaler oder projektspezifischer Faktoren ermöglichen würden.

1 Background – the role of benchmarks and BAT values (1-2p) (ISI)

Benchmarking is a term widely used and describing a standard or set of standards that can be used as a point of reference for evaluating performance or level of quality (see www.businessdictionary.com) against peers. The comparison of performance can be applied in many fields such as profitability, safety, energy-use or for climate change-related issues. In our case, a comparison of the performance with respect to either Greenhouse Gas (GHG) or often — more specific — CO_2 emissions and in some cases also energy is most useful and will be applied used in this study. Additional common definitions in the context of benchmarking are (adapted from Ecofys and Fraunhofer ISI, 2009):

- “Activity” refers to the commodity, the service provided or an activity the emission benchmark applies to (e. g. production of a commodity in tonnes, heated square meters, kilometres driven)
- “Activity level” refers to the amount of the activity
- “Impact” refers to the measured impact, here mainly GHG emissions, but partly also energy use
- “Benchmark level” refers to the level of the performance in terms of specific emissions per unit of a certain activity
- “EU ETS benchmark”/”EU ETS benchmark values” refers to the benchmarks and benchmark levels defined under the EU Emissions Trading System for Phase III

Different levels can be used to define a benchmark. Examples are the average performance level, average of the top 20%/10%/ x% best performers (applied under the EU ETS (10%) and partly in CDM (20%) when benchmarking is used), the best achieved level or the best available level (see e. g. pmr 2017). As will be shown later in the text, the setting of the benchmark level has a key role in the use of benchmarks.

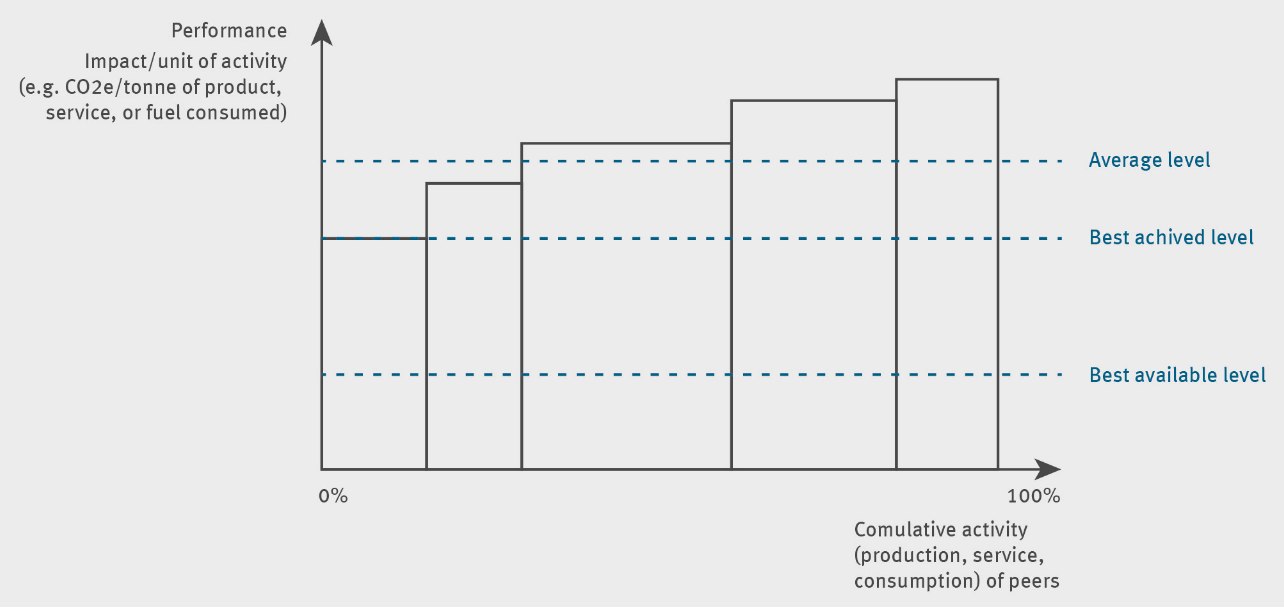

Figure 1: Overview of Design Options and their General Feasibility under the Paris Agreement

A different definition of “Best available technology/techniques” (BAT) can be found in the Industrial Emissions Directive (IED) of the European Union. As the documents developed under the IED are important later in the text, we introduce the definition here. The directive defines BAT as: „best available techniques means the most effective and advanced stage in the development of activities and their methods of operation which indicates the practical suitability of particular techniques for providing the basis for emission limit values and other permit conditions to prevent and, where that is not practicable, to reduce emissions and the impact on the environment as a whole.“ Further, IED specifies that the term „techniques includes both the technology used and the way in which the installation is designed, built, maintained, operated and decommissioned“ and that „available techniques means those developed on a scale which allows implementation in the relevant industrial sector, under economically and technically viable conditions, taking into consideration the costs and advantages, ... as long as they are reasonably accessible to the operator.“ This definition is much broader compared to the above example where „BAT“ refers to a technically possible, but not yet in reality achieved performance level. For the context of this report, we will use BAT in the stricter definition when referring to best available levels as defined by pmr (2017) and to BAT in the IED definition when referring to best available techniques as defined under the IED.

While benchmarks are often used as a management tool to optimize a firm’s performance, they are more and more commonly also used in energy and climate policy. The most prominent application of emission benchmarks can be found in the EU ETS. Since 2013, GHG benchmarks are used under the EU ETS for EU-wide harmonized free allocation of allowances to installations in industry. For that, in-depth sector studies have been prepared to define product benchmarks for 59 products regulated under the EU ETS along with benchmarks for heat and fuel (fallback approaches). In advance to the development of benchmarks for industry sectors, some member states had developed national benchmarks for power generation and combined heat and power generation for earlier years of the EU ETS as well as for allocation of allowances to new plants in industry. Also, other countries such as Korea and Switzerland are using benchmarks for free allocation of allowances in their national emissions trading systems.

The use of benchmarks in the context of the market-based mechanism under Art. 6.4 (crediting system) differs from the application of benchmarks under the EU ETS. Under a crediting system, the benchmark is used to define the baseline. The performance of a specific installation then needs to be monitored to determine the amount of credits that can be issued (determined by the difference between the benchmark and the installation’s performance). That is, the measurement of real performance allows for the issuance of credits. If the measured emissions are higher than the baseline, no credits are issued but no further action from the installation is required. In contrast, in the EU ETS benchmarks are used to determine in advance the free allocation of permits to installations. The reporting of emissions is used to determine the amount of allowances that need to be surrendered at the end of the year. If the installation emits more than the benchmark, then it has to compensate for this by buying additional allowances. Also, emissions in ETS may be reported for the installation as a whole, while the benchmarks can be defined on the level of sub-installations, i. e. only a part of the production process within an installation. In case of crediting, the performance must be reported for the same system boundaries as the benchmark.

The fact that benchmarks are being used to define the baseline, not the standard that should be achieved, also implies that some of the existing benchmarks are not suitable in this context. E. g. a benchmark based on best available technology may not provide sufficient leeway to credit further emission reductions compared to that benchmark.

Paragraph 48c of the Modalities and Procedures for the CDM today explicitly allows for the use of benchmarking. However, it is rarely used in reality. In contrast to the CDM, where each project provides its own baseline based on the specifications within the methodology, the use of benchmarks for the definition of a baseline could increase transparency, reduce administrative costs and may help to increase environmental integrity of the crediting system by preventing over-crediting.

Benchmark values are in general more stringent than baseline values that have been used in the CDM, which were often derived from historical performance. Even though Paragraph 48c of the Modalities and Procedures for the CDM foresees the use of benchmarks, they have not been very often used (see also Section 2). With the Paris Agreement, all countries have committed themselves to NDCs. In this new situation, baseline setting needs to be informed by the NDCs targets and only mitigation action that goes beyond the NDCs emissions levels may be used for international transfers under Article 6 (see Schneider, Fuessler, et al. 2017). This makes baseline setting for host countries more difficult.

Also, in many developed and developing countries NDC targets are not sufficiently ambitious to meet the Paris Agreement’s 1.5/2° target (UNEP Emissions Gap report 2017), or they are not clearly formulated and often do not have economy wide coverage. In this situation, the NDC target cannot be used as a point of reference for determining the crediting baseline.

Under these circumstances, a simple approach may be to use ambitious performance benchmarks for baseline setting. Such benchmarks may be stringent enough that while they allow for the participation in international transfers, they keep a significant part of the mitigation impact in the host country so as not to endanger the meeting of its own NDC because of its engagement in international transfers (and the related corresponding adjustments).

In the following, we provide an overview of potential sources for benchmark values for different key sectors (Section 2), provide criteria for good benchmarks (3), further analyse the practical feasibility of a benchmark approaches under Article 6.4 in three example sectors (4), identify suitable approaches and sectors for benchmarks under Article 6.4 (5) and conclude with findings and recommendations (6).

2 Sources and types of benchmarks

Several different potential sources for benchmark values may be considered:

- BREF documents (Best Available Techniques (BAT) reference documents) that have been adopted under both the IPPC Directive (2008/1/EC) and later under the Industrial Emissions Directive (IED, 2010/75/EU). Each BREF document provides "information on a specific industrial/agricultural sector in the EU, on the techniques and processes used in this sector, current emission and consumption levels, techniques to consider in the determination of the best available techniques (BAT) and emerging techniques."[1] Sectors such as housing, transport or residential waste are not covered as these do not cause industrial emissions. Large emitting industrial sectors are covered by BREF documents. However, the age of the documents varies significantly (see Appendix 1 for an overview of BREF documents and their publication date).[2]
- Data from CDM projects for the relevant sectors.
  - Currently the CDM has more than 115 methodologies and several thousand registered projects both of which can be a valuable source of standardized parameters, reported emission intensities or at least the relevant literature.
  - Approved standardized baselines of the UNFCCC cover grid emission factors, the charcoal sector, rice mills and cultivation, landfill gas capture and flaring, cookstoves and wastewater treatment for specific countries.
- Product benchmarks, in the context of the EU-ETS. Product benchmarks have been developed for 52 products from different industrial sectors, namely mineral oil refining, mineral wool, aluminum, cement, lime, gypsum, glass, ceramics, iron and steel, pulp and paper and chemicals.
- Energy efficiency standards for specific jurisdictions.
- Data on carbon intensity as collected by sectoral organizations, such as in the Cement Sustainability Initiative (CSI) or data on upstream emissions from the oil industry. These have the advantage that the data are available globally, but the quality of such industrial data needs to be scrutinized.

In the following we discuss the potential of benchmarks from those and other sources for the following sectors:

- Industry including energy use
- Energy generation
- Housing
- Transport
- Waste water treatment

For each of these sectors, we identify suitable benchmark values and explore their specific sector context. This allows later for the analysis on which sectors and benchmark types are best suited for their potential use in the Article 6.4 mechanism.

2.1 Industry incl. energy use

The use of benchmarks in industry has a long tradition as it is often used as a management tool to assess and improve the economic performance of installations. In addition, many (multi-national) enterprises are experienced in comparing their own performance against that of competitors. A number of consultants such as Solomon Associates, Plant Service International or SRI management consulting provide services that allow access to anonymized benchmarks if own data sets are provided.

Looking at sources for greenhouse gas emissions in industry, two main sources exist:

---

[1] http://eippcb.jrc.ec.europa.eu/reference/ (09.04.2018)

[2] The OECD's "Best Available Techniques for Preventing and Controlling Industrial Chemical Pollution" is a project that aims to assist governments to implement policies and practices that embody BAT (or similar concepts) to prevent and control industrial emissions. It is separated into three activities (compilation of policies in four member and three partner countries, exchange of experience and finally evaluation of effectiveness). So far there exists only a report concerning the first part. The focus is not on deriving BAT values per se but to help in the process of creating appropriate laws. It has therefore limited value for deriving carbon related benchmarks.

- Consumption of fossil fuels, mainly for heat but also for electricity production
- Process emissions (e. g. in cement or steel or glass production).

The most prominent case for GHG benchmark application in industry is the EU ETS. It uses the following cascade of benchmark-types:

- Product benchmarks (output-related) that define emissions related to the production of one unit of the final or intermediate product (in units of tCO_2eq/product); 52 product benchmarks have been developed for the free allocation of allowances under the EU ETS for the main emitting sectors: coking, sintering and iron and steel production (most importantly: hot metal), non-ferrous metals, cement clinker, lime, glass, tiles and bricks, pulp, paper and board, chemicals and refinery products.
- Fall-back benchmarks can be used if a homogeneous output for a product benchmark is not available (e. g. in the chemical industry, which has various outputs). For fall-back benchmarks, there are the following options:
    - A heat benchmark related to the amount of heat used in the process (in units of tCO_2eq/kWh)
    - A fuel benchmark related to the amount and type of fuel used (in units of e. g. tCO_2eq/tonne lignite or tCO_2eq/liter of oil).

Product benchmarks are preferred as they allow to harness the mitigation potential due to process efficiency and fuel switches. Heat benchmark allow only for the mitigation potential due to fuel switches. And fuel benchmarks do not incentive any mitigating potential and as such shall only be used for minor cases.

The product benchmarks in the EU-ETS are performance benchmarks and are based on so-called benchmark curves. They have been developed based on data collected from all installations within the EU producing a specific product. Based on information on production and emissions in the – then – most recent historic years, the benchmark level has been set to the average GHG emissions of the 10% best performers in the EU. In contrast, the heat benchmark is a technology benchmark, determined on the use of natural gas.

For being able to provide a limited number of benchmarks that cover a large number of installations throughout the EU so called “sub-installations” were defined, intermediate products that can, but are not necessarily always traded between installations, but can also be produced in an integrated way. While this allowed to limit the number of product benchmarks used under the EU ETS, it also increased the necessary effort to collect the data needed to calculate production of the intermediate product as well as emissions related to the production of the intermediate product.

Besides the EU ETS, national voluntary agreements (Energy Efficiency Benchmarking Covenant) have been in place in the Netherlands and parts of Belgium to increase energy efficiency in industry. Participating companies were required to report their energy use. Benchmarking methodology needed to be developed by the participating companies and was to be verified and approved by an official verifying entity.

Besides the EU ETS benchmarks, also some industry sectors have been engaged in the collection of data for benchmark development. For refineries, Solomon Associates collect information on refineries and steam crackers, among others on GHG intensity of the individual installations. Firms that collect and provide data themselves have access to an anonymised data set on performance of other refineries and steam crackers. While data is not freely accessible to everyone, Solomon Associates was in the past open for cooperation to develop benchmarks in other contexts (e. g. under the EU ETS or for the US EPA).

Another association active in the development of reporting protocols and collection of data is the world cement association. Under the Cement Sustainability Initiative (CSI) data were collected from participating enterprises and a benchmark was developed based on the collected data for cement clinker or cement. The data set “Getting the numbers right” is freely accessible (aggregated information on all reporting enterprises). It contains information on 50% of global cement production outside of China. The CSI also submitted a methodology (NM0302)[3] to the CDM executive board based on the data collected.

---

[3] https://cdm.unfccc.int/methodologies/PAmethodologies/pnm/byref/NM0302

Similar to the CSI, a standardised protocol for data collection and reporting was developed by the World Resources Institute and the World Business Council on Sustainable Development for aluminium producers, focusing on PFC emissions. Two CDM methodologies were developed based on the standardised protocol (AM0059 and AM0030).[4]

Other documents providing information on BAT or performance of installations are the BREF documents developed under the IPPC Directive on behalf of the European Commission. They contain information – among others – on energy consumption and GHG emissions. While they are available for the most energy-intensive sectors and provide a detailed view on installations in the EU, they were found not to be suitable for the development of benchmarks under the EU ETS for two reasons: the BREF documents developed in "the first round of the Seville-process (1996–2010)" are very heterogeneous and differ in size, content and age (some state specific energy consumption of technologies, others merely state ranges or do not provide information at all) and background and status of information contained in the documents is not always clear. Despite these shortcomings that prevented use of the information for benchmarks under the EU ETS, they provide a good starting point for information on installations and technology available in the EU. Also, there are updating and improvement processes in place that may make the documents more relevant for benchmarking purposes in the future. If available, BREF documents will be considered in the case studies in section 4.

Under the Asia-Pacific-Partnership, the task force on steel collected information on state-of-the-art clean technologies in the participating countries and published those information in the form of a handbook. Technologies are grouped by production steps and the state of the technology is ranked between mature and emerging. However, technology information in the document are limited and in particular no information is available on energy consumption or GHG emissions.

2.2 Energy generation

Energy generation relates to electricity production as well as heat production in centralized heating plants (in contrast to local heat production in industry or housing). In contrast to industry, the final products (electricity, and – although to a lesser extent – heat) are very homogenous. We concentrate on electricity generation.

Electricity generation relevant for GHG benchmarking includes coal- and lignite-fired power plants, gas-fired power plants, in some cases oil-fired power plants and, to a certain extent also biomass- and waste-fired power plants. All other power generation (e. g. from nuclear or renewables) is – broadly speaking – independent of direct CO_2 emissions. In addition, a differentiation needs to be made between electricity generation only or combined heat and power (CHP) plants. While in aggregate, efficiency of CHP plants is higher and hence emissions are lower, emissions for electricity may be higher compared to conventional electricity generation, depending on the method to calculate heat and power's share in emissions.

The definition of electricity-benchmarks is normally output-related to not differentiate for efficiency of the plant. More interesting is the question of whether a technology-differentiated benchmark is applied (e. g. one for coal, one for lignite, one for gas) or not and how existing abatement options are taken into account. Renewable energies, biomass, nuclear but also – with slight restrictions – the combination of fossil-fuels with carbon capture and storage allows for CO_2-free electricity generation. Hence, the level of the benchmark merely reflects the level of ambition of the target that should be met along with the incentive that should be provided by the market-based mechanism to invest into low-carbon technologies.

---

[4] https://cdm.unfccc.int/methodologies/DB/CHNLRVLNEAM438MR5400YQDS3CPC50 and http://cdm.unfccc.int/methodologies/DB/PKA23BNEYGINU7U4FBINDNYP1F1EU8

2.3 Housing

Greenhouse gas emissions in the housing sector are primarily related to heating and cooling, the provision of warm water (drinking, shower, etc.) and even the provision of electricity for appliances. In the following, we will concentrate on heating and cooling.

Heating

Benchmarks may be derived from:

- Statistics of energy consumption data of the housing stock for the country or region considered or similar geographical areas (regarding climatic and economic conditions). Such data is available for already existing buildings, mainly through energy bills. A benchmark corresponds to a certain quantile.
- Consumer data from energy providers (especially gas or district heating)
- Building norms[5] for buildings relevant for the country in question or similar countries (regarding climatic and economic conditions). Building norms normally include one or both of the following:
    - Country-specific requirements of useful energy demand[6] in kWh per energy reference area [kWh/m^2] for new buildings[7]
    - heat transfer coefficients (u-values) [W/(m*K)] of specific components for new or retrofitted buildings[8]
- A micro-economic investment model (weighing investment costs against energy savings)
- National greenhouse gas NDC target for the building stock that can be broken down to a benchmark value per adequate activity metric such as heated square meter.

The greenhouse gas emissions related to heating can be mitigated at three major control elements:

1. The isolating properties of the building **envelope** (roofs, walls, windows, doors, floors, etc),
2. The energy efficiency of the **heating system**, and
3. The type of the **fuel** used.

There are additional factors that influence greenhouse gas emissions, e. g. the climate, the indoor temperature, the shape of the building, etc. Yet, these factors are not considered in this context.

Related to the three major control elements, different benchmark scopes can be defined:

Benchmarks that comprise only **the building envelope** are mainly related to building norms. Projects may thus claim emission reductions if a building's useful energy demand or specific components' u-values are better than these norms.[9] To calculate greenhouse gas reductions, the useful energy savings have to be converted to primary energy savings (through a parameter that reflects the heating system's energy efficiency). Second, the primary energy has to be multiplied with the emission factor of the building's energy carrier. The first step is complex, as a building's useful energy demand cannot be directly measured. One has to take the final energy demand (energy bills) and use assumptions on the energy systems efficiency. Transforming improvements of u-values into final energy demand is even more errorprone, as this depends, among other things, on the condition of the building's other components and most importantly on use patterns.[10]

---

[5] For example, ISO 52016-1:2017 (internationally), SIA 380/1 (Switzerland), MuKEn (Switzerland), ANSI/ASHRAE/IES Standard 90.1-2016, Directive on Energy Performance of Buildings of the European Union, etc.

[6] Useful energy is the energy needed to increase the building's temperature. Primary energy (also called end energy) is the energy embodied in the energy carrier used for this purpose. The primary energy demand is always higher. The difference stems from the inefficiencies in the transformation and distribution process.

[7] Norms normally do not prescribe a specific energy demand in kWh/m2 for retrofitted buildings, but only u-values for retrofitted components.

[8] In Switzerland, for example, new building's walls or windows shall have a u-value of 0.17 or 1.0, respectively.

[9] For example, triple glass windows may have better u-values than norm windows. And they lower the useful energy demand.

[10] U-values would thus rather be useful for a deemed savings approach.

Benchmarks that comprise **the building envelope and the heating system** typically use the primary energy demand per energy reference area [kWh/m2] and thus can be directly related to statistics of the housing stock. In addition to the hull, greenhouse gas reductions may also stem from a better heating system (heating boilers, storage and distribution system, ventilation system as well as control system). Calculations of greenhouse gas reductions are more straightforward. They are the difference between benchmarked and measured specific final energy demand multiplied with the emission factor of the building's energy carrier.

Benchmarks that comprise **the building envelope, the heating system and the energy carrier** may directly be related to national greenhouse gas target of the building stock. Beyond the possibilities of the hull and the heating system elements, greenhouse gas savings may also stem from a fuel switch. This may be a switch to a building level heating system that is less greenhouse gas intensive (e. g. from coal or oil to gas, to wood or to a heat pump) or through a connection to a district heating system (with a lower greenhouse gas emission factor).[11]

Benchmarks may differentiate between several characteristics related to heating of a building stock. Typically, the differentiation between newly constructed buildings and retrofits of existing buildings as well as different climatic conditions is made. The latter can roughly be accounted for using heating or cooling degree days. In addition, one may consider construction vintages, types (e. g. single-family houses, multi-family house, schools, administration building, restaurants, hospitals, industry, ware-houses, gyms, indoor-pools), etc. The more elaborate such differentiations are, the more accurate the calculated greenhouse gas reductions will be. On the other side, the effort for both benchmark development and MRV increases.

Cooling

For cooling the approach would be quite similar. The energy needed for cooling decreases if the building's envelope and the energy efficiency of the air conditioning unit are being improved. The actual power consumption may then be compared with a benchmark. Using the local grid emission factors allows to calculate the greenhouse gas reductions. An additional difficulty as compared to heating is that the energy consumption also depends on the air's humidity levels.

2.4 Transport

Greenhouse gas reductions in the transport sector can be achieved by less driving, modal shifts, more efficient vehicles, and the usage of fuels that have a lower GHG intensity (e. g. assumed for biofuels). We will focus on more efficient vehicles.

Benchmarks for vehicle efficiency may be related to:

- Fuel efficiency standards (mileage, specific fuel consumption) which exist in various jurisdiction (e. g. EU regulation on CO_2-Emissions, CAFE and GHG in the US) and may depend on a wide variety of vehicle characteristics, such as engine displacement, fuel type, model year, weight or type of the vehicle,
- Handbooks of emission factors for several vehicle categories and a wide variety of traffic situations (e. g. HBEFA),
- data of comparable traffic situations.

Using fuel efficiency standards, greenhouse gas reductions can be determined as the difference of the vehicles' specific fuel consumption as compared to the applicable fuel efficiency standard. In addition, the average or a vehicle's specific vehicle-kilometres travelled, and the fuel's emission factor are needed.

---

[11] In the draft of the revised CO_2-law, Switzerland plans to introduce an upper limit of the six kilograms of CO_2-emissions per square meter energy reference area for existing residential and service buildings whose energy system is replaced. For existing commercial buildings, the upper limit is 4 kilograms, New buildings may not emit CO_2 at all.

It is important to avoid double counting with respect to the jurisdiction's regulations.[12] Also, emissions standards normally have certain features that facilitate manufactures to meet the requirements. In the European Union, for example, the electric cars count more, eco-innovations can be accounted for, etc. Therefore, an Art. 6 system based on emissions standards has to be designed carefully, in order not to infringe environmental integrity.

Handbooks of emission factors allow to design benchmarks related to a clearly containable purpose (e. g. a public transport system, a business car fleet), incorporating the standard emission factor(s) that apply for this purpose. Such benchmarks can then be compared with the actual fuel use.

For clearly containable purposes, one may also use fuel data derived from business-as-usual situations as benchmark. For example, it is possible to collect data of a business-as-usual diesel bus fleet and derive a benchmark. This benchmark can be compared with fuel data from hybrid buses that serve comparable routes and have a comparable size.

2.5 Waste water treatment

There are two major types of waste water treatment plants (WWTP): anaerobic (where mainly methane emissions occur) and aerobic (where methane and nitrous oxide emissions occur) WWTP. These must be distinguished as emission sources, levels and types differ. There also exist various design types for aerobic and anaerobic systems. Emissions depend in addition on a range of influencing factors, which vary substantially among the WWTP. Those factors are e. g. design and type of the WWTP, loading and type of pollutants (organics, nutrients, solids, toxic material, etc.), pH, dissolved oxygen, retention time, temperature or amount and type of bacteria. Controlling for those factors and comparability between WWTP is complex. As a corollary, deriving a suitable benchmark is very challenging. For further details on WWTP see chapter 4.3.

---

[12] For example, emission reductions may be credited for specific vehicles that are more efficient than an emission standard. In the inventory of the jurisdiction's accounting those vehicles must accordingly be considered as if they would exactly meet the emission standard (and shall not be accounted for with their real efficiency).

3 Criteria for good benchmarks to be used under Article 6.4 mechanism

For the development of benchmarks under the EU ETS, 11 guiding principles were developed in advance (Ecofys and ISI 2008). They characterise what – from a theoretical point of view – good benchmarks should respect. However, while at first glance the EU ETS benchmarks and hence the underlying guiding principles seem to provide a good starting point for the discussion, in detail one faces difficulties in the application of benchmarks in the context of international crediting (rather than linking of ETS). First, not all installations, but only some, either new or significantly improved installations will apply the benchmark on a voluntary basis (self-selection of installations). Second, applying similar benchmarks in all countries may not necessarily be the best solution in contrast to the harmonization required under the EU ETS where it was a clear prerequisite that the same benchmark value is applied to installations in all countries regulated under the EU ETS. Certainly, for products competing in the global market, a level playing field seems appropriate. However, in the international context, differences in the countries' national circumstances and the UNFCCC's principle of common but differentiated responsibilities and respective capabilities (CBDRC) may call for factoring in the national context in the definition of the benchmark.

3.1 General criteria for good benchmarks

Instead of building on the criteria for benchmark development from the EU ETS, the following criteria for good benchmarks are based on the process necessary to apply a benchmarking approach in the international context (DEHSt 2013):

- Clear definition of **system boundaries**: The system boundaries define the activities/production phases and/or products for which a benchmark value is defined. The definition of a (final) product or the whole production process increases the emission coverage and the emission reduction potential. However, high heterogeneity of the production processes (maximize emission/energy coverage of the process) may reduce the applicability of a large system boundary. In addition, benchmarks for smaller systems may be easier to be applied in different installations and countries, while large systems may be very plant-specific and general definition of a benchmark may be more difficult. Independent of the size of the system it is important to define the system boundaries in a clear and transparent manner that does not allow for shifting emissions from within and outside the system boundaries by installations.
- Adequate **definition** of key performance indicator/ benchmark: The performance that shall be measured is normally calculated as impact (e. g. in terms of GHG emissions, CO_2 emissions or energy use) divided by the activity (e. g. tonnes of product, electricity produced, heating, …). Both, the definition of the impact (inclusion of all GHG emissions or CO_2 only, inclusion of indirect emissions from electricity and heat, definition of benchmark based on energy use) as well as the definition of the activity (e. g. production of tonnes of paper in general, production of tonnes of specific quality types of paper, production of tonnes of specific paper types or even paper types further differentiated by certain parameters such as thickness of paper or colour of paper) determine the benchmark but also the demand for data and the emission reduction options. An adequate definition reflects functionality along with a certain pragmatism for data availability and collection.
- **Availability of data** for determination of benchmark level: Different approaches exist for the definition of the benchmark level. All of them have in common that data is needed. If performance benchmarks shall be defined based on actual existing installations (e. g. best achieved level, top percentile based, average level), a significant level of statistical data on existing plants (either within a country or group of countries or larger region or globally) are required. Definitions based on the best available levels in the stricter definition requires less real-world data, but the calculation of emissions from a virtual best plant by theoretical modelling can also be very arguable. In both cases, theoretical modelling as well as statistical data, it needs to be ensured that data collected are accurate and robust. If available data was subject to high uncertainties, the definition of a benchmark based on that data is not recommended. That applies to both activity level data as well as impact data.
  A clear connection exists between the definition of the system boundaries and the availability of data for quantification of benchmark levels. In some cases, official statistics may exist (such as the national energy balance) that contain – in parts – data that can be used for calculation of the benchmark level while in other cases – such as under the EU ETS – all data needs to be newly collected.

- **Similar benchmarks for similar products** in similar countries: While it can be argued that national circumstances should be factored in in the definition of benchmark values (e. g. heating and cooling demand differs significantly for different countries in different areas of the world), similar benchmarks should be applied to products produced in countries with similar national circumstances. However, if benchmarks of different stringency are implemented in countries that participate in the same market for a specific product (e. g. steel), different benchmarks may lead to carbon leakage.
- Benchmark levels should **incentivise investment in no- and low-carbon technologies**: Another aspect that should be considered when defining the benchmark levels is that the benchmarks have to address two purposes: they need to be sufficiently ambitious to incentivise long term investment in low- and no-carbon technologies; in contrast, investment in technologies that lead to only small emission reductions compared to today's levels and will – in the long-run – lead to a lock-in of not-climate-friendly technologies (e. g. super-efficient coal-fired power plants without CCS) should not be incentivised. In addition, it is likely that the amount of money that can be generated with selling the credits needs to be sufficient for a firm to make the decision in favour of a cleaner production technology – in the absence of national climate policies such as carbon prices that would incentives such an investment. It should be noted that the voluntary nature of a benchmark-based crediting system leads only to positive incentives for low emitting installations, but does not provide negative incentives (such as higher compliance costs) for high emitting installations as under ETSs.
- **Improvement over time:** The benchmark level should not keep constant over time. Instead, it should decrease to reflect general technology improvements over time as well as the increasing level of ambition of overall emission reductions. That may require repeated collection of data.
- **Availability of data for performance evaluation:** Data need not only be available to define the benchmark level, but also – and at least as important – the performance needs to be monitored in the installation to check the own performance against the benchmark. While that can be easy if all production phases within the installation are included in the benchmark, it can be complicated if e. g. subinstallations are being defined as is the case under the EU ETS and certain processes needed to be included or excluded from the calculation of the performance value

3.2 Specific requirements for using benchmarks under Art. 6.4

With the Paris Agreement, all countries have committed themselves to NDCs and many developing countries have specific mitigation targets in their NDCs. From this, several additional requirements for the use of benchmarks for crediting baseline setting under Article 6.4 may be derived:

- **Consistency with the NDC target:** Crediting baselines for Article 6 mechanism do not only required to be below BAU emission but in addition need to be below an emissions trajectory that is in line with the host country reaching its NDC target (Schneider et al. 2017). This holds also for a benchmark derived baseline. In order to be suitable as benchmarks on a global level, benchmarks need therefore to be sufficiently stringent to assure with high certainty that it is more stringent than the NDC target.
- Consistency with long term goal of Paris Agreement: Article 4 states that the long term goal of reaching the 1.5/2°C target requires "global peaking of greenhouse gas emissions as soon as possible […] and to undertake rapid reductions thereafter in accordance with best available science, so as to achieve a balance between anthropogenic emissions by sources and removals by sinks of greenhouse gases in the second half of this century, on the basis of equity […]". This may be interpreted that benchmarks should be so stringent that they do not allow for the lock-in of technologies of medium to high carbon intensity. For instance, there should be no fuel specific benchmarks incentivizing the construction of efficient coal power plants, as they are not in line with the required emission reduction pathway to meet the Paris Agreement goal (even though such a technology wold be in line with the NDC target).
- **Applicability to a wide range of countries:** The use of benchmarks under Article 6.4 would mean that they are applicable to a wide range of countries, not only to developed and emerging economies, but also middle- and low-income countries. Or, a scale of international benchmarks would be defined for each of the main group of countries (e. g. high-, middle- or low-income countries). Such an approach however is strongly dependent on the categorization used and the best performers in each group

- **Agreed process for determining benchmarks:** A suitable institution such as the Article 6.4's supervisory body would need to provide for a transparent and science-based process on how to determine benchmarks for different sectors and mitigation actions.

What if a different benchmark is used for additionality and for the crediting baseline?

One might think of a benchmark-based approach in the Article 6.4 mechanism that uses a benchmark A to determine if the mitigation would not have happened in absence of the mechanism (additionality) and another (less stringent) benchmark B that is used to determine the crediting baseline which defines the amount of mitigation outcomes issued. Such an approach with two benchmarks may increase the profitability and therefore the attractiveness of the mitigation activities in particu-lar if the prices are low.

However, in such a situation also the (less stringent) benchmark (B) would need to fulfill consistency with both the NDC target, because of the need for corresponding adjustments. With this, in practice benchmark (B) would need to be quite stringent and there would be not much room for widely differing benchmarks.

4 Three examples

In order to further analyse the practical feasibility of a benchmark approach to define crediting baselines, the following three examples in the cement and in the iron and steel industry as well as in wastewater treatment are given.

4.1 Case study: Benchmarking for Cement (Clinker)

Cement production process

The standard production process of cement follows three main steps (Ecofys, Fraunhofer ISI, Öko-Institut 2009).

1. **Raw material preparation:** The preparation of the raw material includes the transport of the raw materials (limestone approx. 90% and other e. g. clay, iron ore), crushing and grinding of raw materials into homogenized powder called "raw meal".
2. **Clinker production:** Temperatures of > 900°C are required to transform limestone ($CaCO_3$) into lime (CaO), releasing CO_2 in the process. The calcinated raw meal is sintered and formed into clinker. It is the most energy-intensive and CO_2 emitting part of the cement production. The step is necessary, because it lends the cement its binding properties. The process emissions due to calcination are constant, not avoidable (only by substituting clinker) and determined by the chemical reaction (about 0.507 t CO_2 per tonne of clinker, depending on fraction of lime in clinker (IPCC 2001)).

On average, the following percentages of CO_2 emissions can be assumed in clinker production: 55% calcination, 35% thermal energy (22% energy required for endothermic calcination, 13% heat losses), 5% transportation, 5% electricity. (Ecofys, Fraunhofer ISI, Öko-Institut 2009)

3. **Cement grinding:** To produce cement, the clinker is finally grinded and mixed with a variety of different ingredients.

Mitigation options

The main drivers for reducing the carbon content of cement are the addition of clinker substitutes, the fuel, the efficiency of the kiln or other new clinker types.

Clinker substitutes: Some mineral components such as fly ash from coal combustion, slags from the steel industry or some natural volcanic materials (e. g. Pozzolana) have hydraulic properties. These properties are a prerequisite for the materials to be suitable as clinker substitutes, as these materials harden both in air and under water and are also resistant. For this reason, these components can be used to reduce the clinker content in cement. Such blended cements can be manufactured with up to 65% of slags or 35% of fly ash without quality problems in terms of the strength of the cement. Ordinary Portland Cement typically contains 95% clinker. However, the availability of clinker substitutes that are being produced as by-product of other production or combustion processes is limited and is likely to decrease, because improved facilities, less steel produced with the Basic Oxygen Furnace process and lower coal combustion in the future can restrict the supply (Ecofys, Fraunhofer ISI, Öko-Institut 2009).

Kiln efficiency: More and more alternatives to coal are being used for firing the kiln (e. g. biomass, used tires, waste, solvents or waste oils). These alternatives often have a lower carbon content and are partly of organic origin, which can reduce emissions. Again, availability and usage competition of the alternative fuels can restrict the use of low-carbon fuel alternatives. The available of raw materials limits the choice of kiln type.

Choice of kiln type: One of the biggest problems of the calcination process are the large heat losses. Modern plants try to minimize these heat losses by using the heat for preheating or to dry the raw materials. But the losses are still very high even in modern plants.

New clinker types: In the last few years, some companies also have been trying to develop new types of clinker that use less limestone and require less heat for calcination (e. g. sulfoaluminate clinkers). Although the tests with these new types of clinkers were promising, they are currently not used on a commercial scale due to limited availability or higher costs for raw materials (e. g. alumina) (LCTPI 2015).

Benchmarking aspects

System boundaries

Since the three steps mentioned above are the only ones that occur within the cement plant or at least in its immediate neighbourhood, only these are taken into account when defining the system boundaries. Further steps such as transporting the cement would significantly increase complexity and would lead to overlaps with the transport sector. Especially with regard to the feasibility of data collection, emission control and the fact that the plant operators do not carry out steps outside the plant, system limits that go beyond these limits do not seem practicable.

Activity choice

One of the most important questions in developing a benchmark for cement is whether clinker or cement should be the basis for the benchmarks. The arguments put forward in the development of the European Emissions Trading Scheme in relation to clinker or cement benchmarks do not apply in context of an international market-based mechanism. Examples are the scope of the EU ETS Directive or perverse incentives to switch from internal production to external production. In favor of cement benchmarking is the principle that benchmarks should be defined on products in order to maximise GHG emissions reduction and energy-efficiency savings throughout the complete production process. One argument against a cement benchmark is the trade in clinker between plants. Keeping track of the carbon-content of traded clinker is not as easy as benchmarking clinker directly at its production site. Unlike the EU ETS, in which participation is mandatory, participation in a market-based mechanism is voluntary. For this reason, project developers themselves would have to set up a monitoring system that tracks the clinker trade, which may be a high if not prohibitive burden. Another argument against a cement benchmark is that several benchmarks or correction factors would have to be developed, as clinker substitutes are not fully available in some regions. However, this argument also applies to the production of clinker, where raw materials and the availability of alternative fuels differ between regions as well. Against this background, it must be considered whether the additional emission mitigation potential of a cement benchmark outweighs the additional complexity due to the addition of further production steps. In order to achieve a maximum reduction in greenhouse gases, for the remainder of this section we focus on the cement benchmark.

A common basis for a benchmark is specific emissions, which means the amount of greenhouse gas emissions per cement output in unit of GHG/unit of cement. For cement, CO_2 is the major GHG.

Data:

In order to determine the benchmark, a data set with production levels and emissions that is as representative as possible is required. There exists worldwide data on cement plants, for example, in the Global Cement Report. The data include information on the plant, the operator, the location, the type of cement or the capacity of the plants.

The Cement Sustainability Initiative (CSI) described in section 2.1, „Getting the numbers right“ (GNR), contains freely available data at country or regional level on production volume, kiln type, clinker type, emissions, energy consumption and fuels used (see Annex). Even if the data availability is not complete, the data provide a good basis for a first indication. A problem is that the data is only available at a regional level, but not at the plant level. A benchmark based on, for example, the average of the 20% most efficient plants cannot be calculated on the basis of this data without further information. Furthermore, detailed data, such as emissions by kiln type or production volume by kiln type, are only shown on a (grey) clinker basis and not for cement. A cement benchmark, as recommended above, cannot be developed on the basis of the data without further information.

Another possible source for benchmark development could be the document „Best Available Techniques (BAT) reference Document for the Production of Cement, Lime and Magnesium Oxide“ of the European Commission. The document provides a detailed overview of the processes and techniques used in cement production, of raw material consumption and emissions, as well as an overview of best available techniques and emerging new techniques. In the document an average value of specific emissions is available. Specific emissions of technologies or certain regions are not shown. A benchmark developed on the basis of this technical information will require a lot of effort to verify the compatibility of existing plants and the raw materials available in the regions.

Furthermore, the reported average value for specific emissions seems too high to represent an ambitious benchmark. The document estimates average emissions for one ton of cement at around 0.672 tons of CO_2 (p. 44), which seems very high compared to the global average based on GNR data for 2016 of 0.646 tons of CO_2/ton of cement (To the best of our knowledge, both values are based on the same calculation basis). Due to the lack of information on specific emissions in general and the rather high and therefore unambitious value for average specific emissions, the document appears unsuitable for benchmark development based on specific emissions.

More details on cement data can be found in the Appendix.

Monitoring:

In the market-based mechanism it is necessary to monitor emissions and the amount of cement produced by the plants. While monitoring cement quantities is relatively simple, a series of measurements and calculations are required for emissions. The process emissions from calcination for cement clinker are also simple to measure (about 0.507 t CO_2 per tonne of clinker). The emissions caused by the heating process are more complicated to calculate. At least information on the emission factors of the fuel and the amount of fuel used are needed. CSI has published the „CO_2 and Energy Accounting and Reporting Standard for the Cement Industry“ for this purpose, which can provide a good starting point.

Benchmark adjustment:

In order to establish an efficient benchmark system, it is necessary to continuously adjust the benchmarks. Changes in the availability of alternative fuels or – more importantly – developments in the sector of alternative low-CO_2 clinker types must be observed in particular. Other developments such as the improvement of kilns or research breakthroughs in the field of clinker substitutes may also be relevant. These improvements may cause specific emissions to decrease, making it necessary to adjust the benchmark.

Regional correction factors:

One of the main questions in developing a benchmark for the cement industry is whether or which correction factors should be introduced. A distinction could be made between

- the kiln types used
- the size of the plant
- the age of the plant
- the raw materials available
- the fuels available
- availability of clinker substitutes

According to Ecofys, Fraunhofer ISI (2009), no different benchmarks should be set for a homogenous end-product to avoid carbon leakage. This argument is less important in this context due to the voluntary nature of the market-based mechanism. A benchmark that is chosen too ambitiously or that makes it impossible for some plant operators to meet will fail to achieve its goal. It is important to note that in some regions there are only certain types of raw materials for which certain kilns are required in the calcination process. Since the transport of cement has not proven to be economically viable and would in turn cause large quantities of emissions, it could make sense to factor in regional characteristics such as different raw materials when developing benchmarks. Furthermore, the availability of alternative fuels and their compatibility with the required kiln types as well as the availability of clinker substitutes must be investigated and, if necessary, correction factors must be applied for certain regions. It is important that correction factors are only introduced if alternative fuels or substitutes are not available or cannot be used due to political regulations (e. g. mandatory clinker content in cement or waste must only be used in waste combustion plants).

4.2 Case study: Benchmarking for Steel

Steel production process

Steel can be produced from iron ores in integrated steel sites (primary steelmaking) or from steel scrap (secondary or recycling steelmaking). The production of primary steel in integrated steel sites, which is globally the dominant steelmaking process, follows three steps using different plants. The first step includes raw material preparation, i. e. is cokemaking from coking coal and sintering of iron ore fines. The second step includes the blast furnace (BF) and the basic oxygen furnace (BOF) where coke and sinter are converted to hot metal and finally crude steel. In a third step, crude steel is processed to finished steel by passing a set process steps including secondary metallurgy, casting, rolling and finishing. The off-gases from primary steelmaking are used throughout the process steps, e. g. in sinter plants, blast furnaces or rolling and parts are also used in power plants for electricity generation.

In secondary steelmaking, the process chain typically consists of the and secondary metallurgy. Direct reduction of iron ore, a less widespread process, typically uses natural gas to produce direct reduced iron that is then fed to the electric arc furnace to produce crude steel. The process can also be run with hydrogen.

Globally primary steelmaking is the dominant steel making process. Developing countries, however, seldom possess these capital-intensive facilities (WSA 2018a). Developing countries with a relevant primary steelmaking capacity[13] are China, Brazil, Turkey, Mexico, Argentina, South Africa, Iran, India, and Vietnam (UN 2014). Secondary Steelmaking is vastly spread among developing countries (Turkey, Belorussia, Trinidad and Tobago, Argentina, Brazil, Ecuador, Venezuela, Egypt, Morocco, South Africa, Oman, Iran, Qatar, Saudi Arabia, United Arab Emirates, China, India, Indonesia, Malaysia, Pakistan, Philippines, Thailand, Vietnam exceeding a production 0.5 Mt in 2015) .

According to Ecofys and Fraunhofer ISI (2009), 88% of the sector‘s CO_2 emissions originate from the production of coke, sinter, BOF crude steel and EAF crude steel. The remaining 12% of the CO_2 emissions stem from a large variety of downstream processes, among these foundries, casting machines, hot rolling, cold rolling, surface treatment like tinning and galvanizing.

Mitigation options

Energy efficiency

The energy intensity of primary steel production has been more than halved in the past decades in Germany. This was achieved by several measures, including the introduction of secondary steelmaking, the basic oxygen furnace and continuous casting as well as of pulverized coal injection and the recovery of basic oxygen furnace gas (Arens 2017).

Heat recovery is another option to increase energy efficiency. Coke dry quenching, for instance, cools hot coke with nitrogen. The heated nitrogen is then used to generate steam and electricity. Energy can also be recovered from, e. g. sinter coolers, furnaces, blast furnace slag, or hot slabs.

Increase of steel production from recycling

Steel produced via the primary route consumes about three times more energy than steel recycled from steel scrap. Thus, increasing the use of scrap or the production of secondary steel helps to reduce the energy consumption and CO_2 emissions of the total steel industry. However, steel production from scrap is limited by scrap availability and by the quality of steel grades that can (currently) be produced via secondary steelmaking.

Material efficiency

Reducing material losses during steelmaking and steel consumption during steel use improves efficiency. Reducing the number of processing steps also improves efficiency. For instance, directly injecting coal into blast furnaces reduces the need for coke. New reactors are designed to completely omit coke (i. e. COREX-process) and sinter making (i. e. FINEX-process). Belt-casting technologies process coils directly from liquid steel and thus skip reheating processes.

---

[13] exceeding 1.0 Mt in 2015

Fossil-free steelmaking:

Hydrogen from renewable energies is a promising technology providing a low CO_2 energy to industry. It can be directly used in the steel industry by producing direct reduced iron (DRI) that can be processed to steel. Plants to produce DRI, so-called DR-plants, are commercially available for the use of natural gas. There is also some experience with DR-plants that were run with hydrogen from fossil fuels. The concept of using hydrogen from renewable energy sources to produce DRI is new, but is currently explored by several European steel companies.

The electrolysis of iron ore offers a second option to produce steel directly from electricity. Compared to steelmaking from renewable hydrogen, the electrolysis of iron ore omits the generation of hydrogen. Thus, this concept may be more efficient. However, in contrast to DR-plants, the electrolysis of iron ore is currently under development. For 2030, a demonstration plant is announced.

Benchmarking aspects

System boundaries

Setting the system boundaries for benchmarking steel making is of key importance, since the system boundaries strongly affect the resulting energy and CO_2 intensity. For instance, Tanaka (2008) found that energy intensity may vary from 16 to 21 GJ per ton of crude steel depending on the chosen boundaries.

An integrated steelmaking site encompasses a set of plants, including cokemaking, sinter plants, blast furnaces, basic oxygen furnaces, secondary metallurgy, rolling and finishing. Next to these plants that are closely related to the steelmaking process, there may also be auxiliary plants located on the steelmaking site like a power plant or oxygen plants. The CO_2-intenstiy of primary steel is affected by all these plants. Additionally, the CO_2-intensity is affected by the processing depth that is achieved in each steelmaking site. Additional cold rolling and finishing, for instance, increases the specific CO_2-intensity.

There are two types of steelmaking benchmarking systems. The estimation of the CO_2-intensity of the entire steelmaking site is discussed in Tanaka (2008) and is also proposed by the ISO 14404[14]. This approach relies on input to and output from the steelmaking site, thus on values that should be easily available. However, a certain amount of adjustments would have to be made to benchmark different steelmaking sites. Worldsteel claims that a comparison against a reference plant is possible, but no further information on this reference plant could be provided (World Steel Association, 2018b). Second, the EU-ETS as well as the DIN EN 19694[15] set benchmarks on the product level. While this approach requires detailed process data on each plant included as well as rules how to deal with by-product gases, it allows a direct comparison of the CO_2-intensity of the same products from different sites.

Activity choice

The chosen activity should be crude steel if the benchmark is set for the entire steel site. If the benchmark is set for a single plant then the product of this plant is the reference (e. g. sinter, hot metal, hot rolled steel).

Ecofys and Fraunhofer ISI (2009) proposed five to six product benchmarks for the iron and steel industry under the EU Emission Trading System. These benchmarks are coke, sintered ore, hot metal, EAF carbon steel and EAF high alloy steel. Furthermore, they argued that an additional benchmark for foundries could be possible. Direct emissions from foundries amount to about 1.4% of the overall emissions of this sector between 2005 and 2008. Ecofys and Fraunhofer ISI (2009) argued that since there is a large variety of semifinished products in the iron and steel sector, product benchmarks should be established only for a smaller number of products, but these should cover to a large extent the direct emissions of this sector. Since coke and sintered ore are traded as intermediate products in primary steelmaking, they are proposed to receive own benchmarks to allow allocation to installations selling these intermediate products.

---

[14] Calculation method of carbon dioxide emission intensity from iron and steel production

[15] Emissionen aus stationären Quellen – Bestimmung von Treibhausgasen (THG) aus energieintensiven Industrien

Crude steel from primary steelmaking and crude steel from secondary steelmaking can be regarded as distinctly different products in respect of steel qualities. Therefore, separate benchmarks should be set up for the process routes. First, a benchmark for hot metal is proposed that also includes the basic oxygen furnace and casting machines. Second, two benchmarks for secondary steelmaking are suggested, one for EAF carbon steel and the other for EAF-high alloy steel. For waste gases, e. g. blast furnace gas, that are exported from the production process outside the system boundaries of the relevant product benchmark and combusted for the production of electricity, no additional allowances are allocated beyond the share of the carbon content of the waste gas accounted for in the relevant product benchmark.

The product benchmarks also take account of the historical emissions from flaring of waste gases related to the production of a given product and fuel used for safety flaring should be considered fuel used for the production of non-measurable heat in order to take account of the compulsory nature of these flares.

Data

Benchmarks for the entire steel site require data on input to and output from the steel site (e. g. coal, natural gas, limestone; crude steel produced, by-product gases sold). Benchmarks on single plants require detailed data on its inputs and outputs (e. g. consumption of by-product gases and onsite produced steam and/or electricity).

IS0 14404 is a standardized method to calculate the energy and CO_2 emission intensity from iron and steelmaking on the company or site level. Its key purpose is to compare energy consumption and CO_2 intensity of single companies over two or more years and the assessment of energy and CO_2 savings from the introduction of new technologies. Companies could also compare their results against a representative reference site. The standard also enables the user to compare their results against better performing sites and identify the areas for improvement. Different companies can be compared if they have comparable production facilities. Three subnorms are available that refer to steelmaking via the blast furnace (ISO 14404-1), the electric arc furnace (ISO 14404-2), or via electric arc furnaces and coal- or gas-based direct reduction iron (DRI) facilities (ISO-14404-3). ISO14404 refers only to the input and output of the site where steel is produced. Thus, it benchmarks the whole steelmaking site. A benchmark on the process level is not possible. This norm seems to be appropriate for internal comparison of energy and CO_2 emission intensity. Since it does not take into account different setups of sites, it should be analyzed which adjustments would be required.

In contrast to the ISO 14404, the DIN EN 19694-3[16] refers to the process level rather than to the entire steel site. This allows to benchmark production plants or a combination of these. However, a large amount of plant specific data is required. If DIN EN 19694 is to be applied for benchmarking steel, companies in non-Annex-I countries, effectively this could only be applied to selected plants, e. g. sinter plants, coke ovens, blast furnaces, electric arc furnaces.

Monitoring

Detailed measuring of input data at single plants may be an issue in less developed countries: instruments may not be available or may not be calibrated. Inaccuracy of measurements distorts monitoring. In the EU-ETS there are strict rules regarding the accuracy of measurement instruments. If the calibration is outdated, for instance, as a fine additional consumption is added to the measured one.

---

[16] Stationary source emissions – Greenhouse Gas (GHG) emissions in energy-intensive industries – Part 2: Iron and steel industry; the respective ISO norm is currently under development.

Benchmark adjustments

Input factors that the steel company cannot influence, should be normalized. For instance, the CO_2-emission factor of the national electricity production may vary between very low (electricity from hydropower) to very high (electricity from bituminous coal)[17]. It is proposed to use the global average value or the average values of the countries participating in the benchmarking. In addition, the CO_2 emission factor of fossil fuels should be normalized as well.

Steelmaking follows similar production processes notwithstanding global regions. Thus, regional correction factors seem not to be necessary. However, it should be considered that steel companies cannot influence the CO_2 grid emission factor and corrections in this respect may be necessary.

---

[17] However, steel companies could assure that electricity comes from renewable sources (e. g. contracting a hydropower plant, purchasing electricity that comes from renewable energies, use low-carbon energy carriers in on-site power plants).

4.3 Case study: Waste Water Treatment

Processes and Emission Sources

Biological waste water treatment plants (WWTP) use living microorganisms to decompose pollution that originates from residential and industrial sources. Pollutants are mainly organics, but also nutrients (nitrogen and phosphorus), pathogens, heavy metals, etc. The main aim of a WWTP is to meet requirements set by the local environmental agencies regulating what is released into the water (as well as the ground and the air). Greenhouse gas emissions are a secondary concern or no concern at all.

The organics are largely converted to carbon dioxide. Yet, those emissions are commonly not considered as additional greenhouse gas emissions as the organics are predominantly of biogenic origin. The biological processes during the wastewater and sewage sludge treatment also result in methane and nitrous oxide emissions.[18] There are two major types of plant designs: anaerobic and aerobic systems. These differ substantially with respect to greenhouse gas emissions processes as well as mitigation options.

Anaerobic WWTP use microbial action to reduce the pollutants in wastewaters in the absence of additional oxygen (there is however oxygen in the waste water and the pollutants). This process is called anaerobic digestion and is good at treating high input levels of organic matter. Its operation is low cost, as e. g. no energy is needed for aeration. Anaerobic digestion produces biogas (a mixture of methane and carbon dioxide), which – if captured – can be used for energy generation. If methane is not captured, this is a major emission source. The removal of organic matter is not complete and thus anaerobic WWTP are not suitable for direct discharge to surface waters (post aerobic treatment would be needed). They also do not remove other contaminations such as nitrogen and there can be odour issues. The main greenhouse gas emissions source is the methane from the anaerobic digestion if it is not captured and destroyed.

Aerobic WWTP use oxygen (or air) and microbial action to reduce the pollutants in wastewaters. There are many different systems such as aerated lagoons, activated sludge or trickling filters. In general, they achieve a better discharge quality as compared to anaerobic systems (i. e. they remove more types of contaminants and the final level of organics is lower). Yet, their operation is costlier (energy for aeration, operation more complex, sludge treatment, etc.). And they produce considerable amounts of sludge[19] which has to be disposed of. Methane emissions stem from anaerobic processes at several stages of the treatment (e. g. anaerobic digesters, sludge treatment and storage as well as methane leakage of a biogas power plant). There are also nitrous oxide emissions that mainly arise as a side product of the biological removal of nitrogen (nitrification and denitrification processes), if this is part of the waste water treatment process. Nitrous oxide emissions are much lower in WWTP that do not have such a system. Another source of nitrous oxide is the storage and — if done — incineration of the sludge.

Mitigation options

Due to the different emission processes, the mitigation options differ between anaerobic and aerobic WWTP.

For anaerobic WWTP, a rather comprehensive mitigation option is to replacement the plant (existing or planed) with an aerobic WWTP, where emissions are generally considered to be lower (see e. g. CDM Method AM0080). With an anaerobic WWTP remaining in place, the most effective mitigation option is to cover the anaerobic digestor and capture biogas, which contains 60% or more methane (the rest is mainly carbon dioxide). End-of-pipe the methane can either be used (generation of heat and electricity or injection into a natural gas distribution grid) or destroyed (flaring in a torch).[20] See e. g. the CDM methods ACM0014 or AMS-III.H. Finally, the separation and treatment of solids in the waste water also yields less methane emissions (CDM Method AMS-III.Y). Under the CDM there are only methods related to anaerobic WWTP, which have high methane emissions from uncaptured anaerobic digestion in the baseline.

---

[18] Methane and nitrous oxide are powerful greenhouse gases with a global warming potential at a 100-year time scale of 25 and 298, respectively

[19] Sludge consists of the particulate components of the wastewater from the primary treatment as well as the active bacteria biomass from the biological purification stage.

[20] A less common method of methane removal are biofilter systems.

For aerobic WWTP, optimizing the nitrification and denitrification process has the potential to lower nitrous oxide emissions. Yet, emission levels depend on a variety of factors, whose influence is not very well understood.[21] Optimization is thus difficult and plant specific. It is important to note that optimizations with respect to nitrous oxide emissions must not impair the nitrogen removal rates (as well as the rates of other pollutants). There are e. g. following options:

- Solid retention time[22] can be increased to maintain low ammonia and nitrite concentrations.
- Larger systems are better able to buffer the pH value and reduce the risk of transient oxygen depletion, both of which are considered to increase nitrous oxide emissions.
- The methane that forms in aerobic WWTP during the anaerobic sludge treatment can be captured and used.[23]
- The foul water which forms during that treatment contains large amounts of nitrogen and is usually fed back into the WWTP. Stripping that nitrogen before recycling the foul water (and converting it e. g. to fertilizer) can decrease the nitrous oxide emissions of the WWTP.
- The inert remainder of the sludge treatment is often burned, and nitrous oxide forms, if temperatures are not high enough. Another mitigation option is thus to increase incineration temperature or to use catalysts to reduce nitrous oxide emissions.
- Finally, in all current WWTP only parts of the energy contained in the raw wastewater (as organic compounds) is converted into methane during anaerobic digestion and thus available for energy generation. The remaining part is converted to carbon dioxide in the nitrification and denitrification processes to remove nitrogen and organic matter simultaneously. From an energy generation perspective this constitutes a waste. There are ways to remove organic matter and nitrogen compounds in separated processes and increase the methane production. Yet, such technologies are still at an experimental phase.

- Larger systems are better able to buffer the pH value and reduce the risk of transient oxygen depletion, both of which are considered to increase nitrous oxide emissions.
- The methane that forms in aerobic WWTP during the anaerobic sludge treatment can be captured and used.
- The foul water which forms during that treatment contains large amounts of nitrogen and is usually fed back into the WWTP. Stripping that nitrogen before recycling the foul water (and converting it e. g. to fertilizer) can decrease the nitrous oxide emissions of the WWTP.
- The inert remainder of the sludge treatment is often burned, and nitrous oxide forms, if temperatures are not high enough. Another mitigation option is thus to increase incineration temperature or to use catalysts to reduce nitrous oxide emissions.
- Finally, in all current WWTP only parts of the energy contained in the raw wastewater (as organic compounds) is converted into methane during anaerobic digestion and thus available for energy generation. The remaining part is converted to carbon dioxide in the nitrification and denitrification processes to remove nitrogen and organic matter simultaneously. From an energy generation perspective this constitutes a waste. There are ways to remove organic matter and nitrogen compounds in separated processes and increase the methane production. Yet, such technologies are still at an experimental phase.

Benchmarking aspects

System boundaries

This chapter concentrates on the benchmarking of methane and nitrous oxide emissions, which arise during the treatment process.

It shall be noted however, that apart from those direct emissions, there are also indirect emissions that could be considered as well. These are related to e. g. the electricity supply, transportation of chemicals and sludge or

---

[21] Some identified factors are dissolved oxygen concentration, nitrite concentrations in both low COD/N ratio in the denitrification stage, sudden shifts of pH and timing of the anoxic and aerobic conditions.

[22] The solid retention time is the average time that bacteria (solids) are in the anaerobic digester.

[23] Sludge contains a high proportion on biodegradable high-energy components. In a separate process this sludge is usually treated in an anaerobic reactor (digester) such that only inert components remain. This is called sludge stabilization.

even the construction of the WWTP (concrete, steel, etc.). When defining an appropriate benchmark, these indirect emissions may also be considered using e. g. a specific planttype as a benchmark.

Activity and Benchmark Choice

Possible activity choices differ for methane and nitrous oxide emissions. For methane emissions possible choices are:

- The Biological Oxygen Demand (BOD), which is a proxy for the organic loading of the waste water.[24] As the methane emissions depend on the organic loading, BOD is a viable activity choice. Yet, the correlation is rather low. The same levels of BOD can arise from different types of pollutants, some of which are, among other differences, more easily biodegraded (e. g. methanol and sugars) than others (e. g. turpentine and soaps). For related reasons, the same levels of BOD may give rise to varying levels of methane emissions.
- The chemical oxygen demand (COD)[25], which is less time consuming to measure than BOD but the relation to methane emission is probably even weaker.
- The population size (for residential sources) or the population equivalent (for industrial sources) in the catchment area of the WWTP.

For nitrous oxides emissions, the nitrogen load is the crucial factor and therefore the most suitable activity choice. Additionally, the nitrogen-to-COD-relation may be accounted for as well.

Benchmark have to be differentiated between anaerobic (mainly methane emissions) and aerobic (methane and nitrous oxide emissions) WWTP. Yet, there also exist several types of WWTP within these classes and it would have to be decided whether design-specific benchmarks are implemented. In addition, emissions depend on a range of influencing factors which may be different in every single WWTP. Those factors are e. g. loading and type of pollutants (organics, nutrients, solids, toxic material, etc.), pH, dissolved oxygen, retention time, depth of a lagoon, temperature or amount and type of bacteria. Controlling for those factors is complex and comparability between WWTP is low. As a corollary, benchmarking development and application is complex.

Data

Methane Emission and — to a lower extent — nitrous oxide emissions have been measured in several WWTP. Yet, given the above-mentioned differences among WWTP in several dimensions and the fact that the studies use different activities, comparability among the data is low and does not suffice as a basis for benchmarking.

Monitoring

The above proposed activity parameters are usually measured for operational reasons and thus additional monitoring demand is low. Emissions, on the contrary, usually cannot be monitored, as in most cases it is too costly to capture and measure emissions just for the purpose of an Article 6.4 mechanism. Measurements can only be taken in specific studies (measurement campaigns) for a limited amount of time. Expected emission reduction from an intervention may then be based on those results (this is called the deemed savings approach).

Temporal adjustments

Temporal adjustments are not needed as the chemical processes that determine emissions stay the same. Yet, if influencing factors (see above) change with time, those would have to be considered if possible (mostly via changes in the activity data).

Regional corrections

Different regions have different WWTP-designs. Anaerobic WWTP (which have higher greenhouse gas emission) are e. g. common in the developing world, but not any more in the developed world. Influencing factors have regional variability as well (e. g. temperature). Therefore, regional benchmarks are meaningful.

---

[24] BOD is the amount of oxygen needed by biological organisms under aerobic conditions to break down the organic matter present in a water sample during 5 days at 20 °C. BOD is thus a surrogate for biodegradable pollution.

[25] COD is the amount of oxygen needed for a complete (chemical) oxidation of all components of the water sample. BOD thus measures biodegradable and non- biodegradable organics as well as oxidizable inorganic material such as chlorine.

5 Analysis – which benchmark approaches work for Art. 6.4?

5.1 General requirements for benchmarks for the Article 6.4 mechanism

A basic requirement for the use of benchmark is the availability of data: Data for benchmark values (as discussed in section 2) but also the availability of activity data information. Adequate levels of data availability are crucial, be it for benchmarks that are determined purely rule based (by formula like e. g. as average performance certain percentile of the market) or by expert judgement (considering ambition and technological leaps). In general, data availability is much more limited in developing countries, in more informal industries and the residential sector.

Besides the criteria for benchmarks in general (section 3.1) the criteria that are specific for the Article 6.4 mechanism (section 3.2) seem to leave little room under Article 6.4 for benchmarks with lower stringency levels, e. g. for fuel specific benchmarks do not allow to tap into the potential of fuel switching or technology specific benchmarks (e. g. for single cycle vs. combined cycle power plants).

The Paris Agreement requires also the regular updating of benchmark values, for instance in sync with the 5-year NDC cycle in order to prevent lock-ins into technologies with high to medium technologies that may not be in line with the longterm goal of the Paris Agreement.

5.2 In which (sub-)sectors is there the best potential for the use of benchmarks?

We consider the following four factors when determining the potential use of benchmarks under Article 6.4. The analysis builds on the evaluation of data availability for benchmarks in section 2.

Table 2: Evaluation of criteria that define the suitability of benchmarks for selected sectors

(Sub-) sector	Activity data availability	Benchmark data availability	Availability of global benchmarks	Carbon market contribution to profitability
Industry energy use – product benchmarks	***	**	**	*
Industry energy use – other benchmarks	**	*	*	*
Industry process emissions	***	**	***	***
Energy generation	***	**	**	*
Housing	**	*	*	*
Transport – general	*	*	*	*
Transport – fuel efficiency standards	**	**	**	*
Waste water	**	*	*	***

key: * = low, ** = medium, *** = high
Source: O

Industry energy use – product benchmark

Where homogeneous products are found, product benchmarks can be defined. Important is in that context to find a clear definition of the product and the related system boundaries. Several examples for product benchmarks are available under the EU ETS.

An advantage of products which are homogenous enough to define product benchmarks is that activity data are more likely to be available (although not normally freely available on the installation level) and comparable between installations and countries. Examples are the cement sustainability initiative or the collection of information on refineries by Solomon. Compared to other industry sectors, availability or collection of activity data is likely to be better.

Similarly, examples exist for the definition and values of product benchmarks from the EU ETS. However, also other ETS systems are interested in benchmarks and in the process of defining product benchmarks (e.g. Korean ETS). These activities may provide a good starting point for defining benchmarks under Article 6.4, although they are not necessarily to be used one-to-one.

In case of similar product definitions and benchmark setting systems among different ETS systems in the future, a comparison of the values can also provide a good indication on the necessity of defining differentiated benchmarks vs. a global benchmark. It needs to be considered, that there may be a bias in the benchmark values as ETS systems are more likely to be implemented in industrialized countries. Whether a definition of a global benchmark is per se an adequate solution in case of product benchmarks, cannot be determined in general. On the one hand, homogenous products such as steel or aluminium or other metals are more likely to be traded between countries and hence competition on a global market is more likely which would be a reason in favour of a global benchmark to prevent distortions. On the other hand, availability of raw materials and the costs of trading certain goods – e.g. in case of cement clinker – may result in the conclusion that the use of a global benchmark value is not adequate.

As the case studies for iron and steel as well as for cement show, the definition of the system boundaries is key. It determines the energy/ emission content of the product, the options for reducing emissions/energy use, but also the resources needed to collect data for monitoring the plant performance which relates directly to the credits that can be generated under Article 6.4. For further investigation into product benchmarks in context of Article 6.4 it is therefore important to understand in detail the advantages and disadvantages of the setting of the system boundaries. As different approaches already exist, iron and steel, but also cement/ cement clinker may provide good starting points for in-depth analyses of product benchmarks.

Industry energy use – other benchmarks

Product benchmarks are only suitable for a small group of industry products that are homogeneous enough. The larger group of industry products is likely to require other benchmark approaches such as fuel- or heat-benchmarks as the products and hence the heat or fuel required in the production process are too heterogeneous to define a product-specific benchmark value.

The heterogeneity of the product also makes collection of activity data more difficulty. Due to the missing standardization — even when available —, production data may not always follow similar definition and are therefore difficult to compare. For similar reasons, the availability of benchmark data is limited. While there are definitions of heat and fuel benchmarks under the EU ETS, they are limited to emission-intensive products regulated under the EU ETS.

The availability of a global benchmark depends heavily on the product/ sector. While for some products it may be possible to define a global benchmark value, it is certainly not possible for all products. The specific production circumstances need to be evaluated to identify the possibilities for global benchmark values.

Again, the benchmarks defined under the EU ETS may provide a starting point for definition of other benchmarks in industry. It should be checked whether the approach for fuel and heat benchmark definitions from the EU ETS (i.e. definition of a standard fuel input in combination with an emission factor for that fuel; see chapter 2.1) can be applied in case of Article 6.4, even if the benchmark values chosen under the EU ETS may not be adequate in many cases.

Industry process emissions

The best starting point for benchmark development in industry present the process emissions. In contrast to energy-related emissions they follow clearly defined chemical processes and can easily be calculated. Examples are CO_2 emissions from calcination of cement clinker or N_2O emissions from adipic or nitric acid production.

The number of products with process-related emissions is limited. At the same time, the products with process-related emissions are often standardized and therefore collection of activity data is easier compared to other product groups. Like the product benchmarks, most process-related emissions are regulated under the EU ETS, hence the methodology applied under the EU ETS can be used in the context of Article 6.4. As the process-related emissions follow clearly defined chemical processes and products are standardized, the use of global benchmark values should in most cases be possible.

What may present a problem in some cases, is the choice of the benchmark value. It depends heavily on the availability of abatement options. In some cases, such as adipic or nitric acid abatement, options are available that allow for an almost complete reduction of process-related emissions (e. g. by a catalyst). In other sectors, e. g. calcination processes for cement and lime production, reduction of emissions is not that easy. It may either require the use of CCS technology, meaning that the emissions are not reduced, but captured and stored, or completely new production processes or product substitutes that are still under development. In these cases, it is unclear whether the definition of a purely process-related benchmark is sufficient to provide incentives to reduce process emissions in a meaningful way or if the definition of process-related benchmarks may in contrast result in a lock-in in old production technologies as the system boundaries do not take into account alternative production processes/product substitutes.

Energy generation

Another good starting point for the development and application of benchmarks is the area of energy generation. In particular the availability of data for power plants is very good for many regions of the world. An example is the global PLATTS database which contains information on capacity and efficiency as well as input fuels for large, but also for many small power plants around the world. This or similar data bases can serve as a starting point for calculating technical possibilities. Due to the relevance of electricity, national statistics normally provide significant information on the production of electricity from different sources which can complement the technical information from databases. Even for countries where emissions from power plants have not been monitored in detail in the past, the technical and production information can be used to calculate proxies for the emission intensity of electricity in a country.

The definition of global benchmarks for Article 6.4 may – at least in the short to medium term – be difficult for energy generation. While in general the same production technologies are available in all countries, national circumstances (availability of fossil fuels and renewable potentials) determine to a large attend the current energy mix. Measuring against a global benchmark would not make use of that benchmark beneficial, in particular in countries with a high share of fossil fuels, where mitigation action is particularly required. In contrast, countries with an already low emission intensity of the energy mix would be able to generate a significant amount of mitigation outcomes to be used under Article 6.4, which may neither be required nor provide a real environmental benefit. In the long-run, however, assuming that energy generation will get closer to being decarbonized in all countries, a global benchmark is more likely to being applicable.

Housing

Benchmark data availability depends on the scope of the benchmark (see section 2): benchmarks can be derived from norms related to the overall building's envelope or single components. Such data is available in many countries. To calculate emission reductions from such a benchmark or to directly derive a benchmark related to energy demand, energy consumption of houses clustered by energy categories is needed. Yet, whereas consumption data is quite often available, it is rarely clustered in such a way. In addition, there are many different types of houses (single family, multi family, office buildings, etc.). A further problem is that houses have a long livetime such that data of the existing housing stock do not serve as a benchmark for new houses, yet they may do so for retrofits. Therefore, benchmark data availability is low. It is therefore difficult to define a proper benchmark.

As housings' energy consumption depends heavily on the climate as well as the general standards of the region's building stock, benchmarks have to be defined locally. A global benchmark is not feasible.

The availability of activity data is medium, as for certain types of energy carriers (e. g. gas or district heating) energy consumption data is often readily available. For other fuels with decentralized distribution, data availability tends to be much lower.

Due to the problems in deriving a benchmark, lack of homogeneity in the building stock and the strong local differences, this sector is not suitable for benchmark-based baseline setting under Article 6.4.

Transport – general

General transport relates to measures in the transport sector, such as modal shifts or changes in the usage patterns (e. g. traveling less or car sharing). There are various mitigation options and deriving benchmarks is notoriously difficult. The baseline depends on many influencing factors such as user behaviour, city structure, culture, etc. Those also differ regionally, such that there is no comprehensive dataset that might be used for benchmarking and global benchmarks can usually not be applied.

Measuring activity is possible for public transport fleets, but difficult for private fleets, except on a very aggregate level.

Due to these problems, transport is in general not suitable for baseline setting under Article 6.4. An exception might be mitigation options of public transport fleets. For those, however, input-based finance[26] (instead of article 6.4 result-based finance) is most likely the more efficient mitigation instrument.

Transport – fuel efficiency standards

Data related to fuel efficiency standards of cars and light-duty vehicles are widely available. There are some regional differences as e. g. in the US cars are on average heavier.

Article 6.4 project could in principle set incentive to increase the fuel efficiency. In most countries, however, regulations on the fleet's fuel efficiency standards are already in place. Additional incentives using article 6.4 projects are thus not useful. On the contrary, they might cause double counting problems and create perverse incentives for the policy-makers not to strengthen the fuel efficiency standards.

Waste water

Activity data may relate to pollution loading or population size of the plant's catchment. Yet, deriving a suitable benchmark is challenging if not impossible, as there are various waste water treatment plant designs, the range of important influencing factors (which vary regionally) is wide and data availability is low (because measurements are costly).

Therefore, the waste water sector is not suitable for baseline setting under Article 6.4.

Carbon market contribution

The contribution of carbon markets to the profitability of low carbon investments differs strongly between different sectors. While e. g. for wind projects the impact of carbon revenues under the CDM added on average below three percentage points to its profitability (IRR), in landfill gas projects, where the high global warming potential of the avoided methane produces much more revenues from carbon markets, the impact on profitability is on average in the order of 14 to 15 percentage points, a much more significant contribution (Cames et al. 2016). This factor is not specific for the context of benchmarking. However, when identifying sectors that are most suitable for using benchmarks under Article 6.4, the economic attractiveness may be an important factor when it comes to recommending sectors with projects that are additional compared to what would otherwise occur and have the potential to best use benchmarks to scale *additional* mitigation action.

---

[26] With input-based finance we refer to instruments that finance mitigation measures up-front and do not measure the outcome (or measure only to evaluate the usefulness of the measure).

6 Findings and recommendations

Baseline setting with globally applicable and stringent benchmarks may be seen as an important instrument for scaling-up market mechanisms under Article 6. This builds on the rationale that benchmarks that are stringent enough lead to baselines that are automatically below both BAU and an emissions trajectory that is compliant with the host country achieving its NDC target. This promises to be an efficient and simple way to solve the challenging issue of defining crediting baselines on the basis of BAU and NDC targets with its uncertainties, the sometimes unclear and only partial scope of NDC targets etc.

The analysis identifies different sources for benchmark data in the industry, energy generation, housing, transport and waste water sectors. It defines criteria for good benchmarks and provides in-depth analysis of the feasibility of benchmarks in three case studies for cement clinker production, steel production and waste water treatment. It then identifies subsectors that appear particularly suitable for the use of benchmarks for baseline setting under Article 6.4. In this context, the analysis focusses on global benchmarks in the sense of simple reference values for specific products and services that are independent of a specific country and may be applied globally or on the level of groups of countries (e. g. low/middle/high income countries).

The analysis indicates a limited potential for global benchmarks. There are some quick wins in the form of global benchmarks related to industry process emissions. Here, the CDM has established robust and stringent benchmarks for baseline setting e. g. in N_2O abatement in nitric acid or adipic acid production, or for abatement of HFC23 emissions in the production of refrigerants. it may also be assumed that with these high GWP gases, the revenues from the transfer of emission reductions may provide a significant contribution to overall profitability and therefore lead to mitigation action beyond BAU.

In addition, some other industries may be suitable for benchmarking, including cement or iron and steel. However, related emissions depend stronger on local factors (such as quality of raw materials) and are more difficult to implement on a purely global level. Here, baseline setting with approaches of intermediary complexity may be possible, building on proposed or approved CDM methodologies and EU-ETS guidance for product benchmarks. In practice, expected carbon prices may not be on a level that would trigger additional action in these sectors.

The process leading to the definition of benchmark values may be challenging to implement under an Article 6.4 mechanism. Providing such benchmarks may also open the door for loopholes and non-stringent values. A stringent and science-based process within the Article 6.4 supervisory body should define adequate global benchmarks but may also open loopholes. In settings of weak governmental oversight, using benchmarks may be less adequate than conventional methodologies of baseline setting, where baselines are set on the basis of project specific parameters that are validated by independent third parties.

Even though there are sub-sectors with large or medium potential for benchmarking, the majority of emission sources cannot be covered by global benchmarks, because the goods and services are heterogenous (e. g. “shoes”, “tonne-kilometers”) and emissions tend to depend also on exogenous local factors. Benchmarking is therefore barely the silver bullet to solve the issues with crediting baseline setting under the Paris Agreement.

However, if one moves from global benchmarks more towards standardized approaches to baseline setting, there is a large body of methodological approaches and reference values from ETS and the CDM that can be used to define crediting baselines in a more efficient and robust way. Their use under Article 6 requires their further development including comprehensive data collection exercises that would allow for standardized approaches taking into account at least some regional, local or project specific factors.

7 References

Arens, M., 2017: Technological change and industrial energy efficiency, exploring the low-carbon transformation of the German iron and steel industry, dissertation, available at: https://dspace.library.uu.nl/handle/1874/344174 [Accessed 12 September 2018].

Cames et al. 2016: Cames, Martin, O Harthan, Ralph, Füssler Jürg, Lazarus Michael, Lee Carrie, Erickson Peter and Spalding-Fecher Randall. (2016). How additional is the Clean Development Mechanism? Analysis of the application of current tools and proposed alternatives. Study prepared for DG CLIMA. 10.13140/RG.2.2.23258.54728.

CCAP (Center for Clean Air Policy, n.d.): CCAP Submission on Internationally Transferred Mitigation Outcomes.

CO$_2$ and Energy Accounting and Reporting Standard for the Cement Industry: www.wbcsdcement.org/index.php/key-issues/climate-protection/co-accounting-and-reporting-standard-for-the-cement-industry.

DEHSt, 2013: Carbon markets in transition. Bilateral Agreements as Basis Towards Piloting Sectoral Market Mechanisms.
www.bmu.de/fileadmin/Daten_BMU/Pools/Forschungsdatenbank/fkz_3712_41_507_carbon_market_mechanisms_bf.pdf

Ecofys, Fraunhofer ISI, 2009: developing benchmarking criteria for CO$_2$ emissions, by order of the European Commission, Environment Directorate-General, February 2009, www.ecofys.com/files/files/developingbenchmarkingcriteriaforco2emissions.pdf.

Ecofys, Fraunhofer ISI, Öko-Institut, 2009: Methodology for the free allocation of emission allowances in the EU ETS post 2012, Sector report for the cement industry, by order of the European Commission, November 2009.

Ecofys and Fraunhofer ISI, 2009: Methodology for the free allocation of emission allowances in the EU ETS post 2012, Sector report for the iron and steel industry, available at: https://ec.europa.eu/clima/sites/clima/files/ets/allowances/docs/bm_study-iron_and_steel_en.pdf (2018-09-12).

European Commission, 2006: Reference Document on Best Available Techniques, Waste Treatments Industries.

Getting the Numbers Right: www.wbcsdcement.org/index.php/key-issues/climate-protection/gnr-database.

Global CCS Institute, 2017: Projects database, Abu Dhabi CCS (Phase 1 being Emirates Steel Industries, available at www.globalccsinstitute.com/consultancy/our-expertise/case-studies/ [Accessed 12 September 2018].

IPCC, 2001: Good Practice Guidance and Uncertainty Management in National Greenhouse Gas Inventories, www.ipcc-nggip.iges.or.jp/public/gp/english/3_Industry.pdf.

Kanton Zürich, 2014: Energiekennzahl Wohnbauten. Baudirektion Kanton Zürich, März.

Lazarus, Michael, Peter Erickson and Lambert Schneider, 2013: Potential for International Offsets to Provide a Net Decrease of GHG Emissions. Working Paper 2013-06. Stockholm: SEI (Stockholm Environment Institute).

LCTPI, 2015: Low Carbon Technology Partnerships initiative, Cement, November 2015, www.wbcsd.org/Projects/Cement-Sustainability-Initiative/Resources/Low-Carbon-Technology-Partnership-initiative-LCTPi-Cement-full-report.

Obergassel, Wolfgang, 2017: Shaping the Paris Mechanisms Part II. An Update on Submissions on Article 6 of the Paris Agreement. JIKO Policy Paper 01/2017. Wuppertal: Wuppertal Institute for Climate, Environment and Energy.

Partnership for Market Readiness, pmr, 2017: A Guide to Greenhouse Gas Benchmarking for Climate Policy Instruments. Technical Note 14, April 2017. Policy Research Working Paper 7594. World Bank Group, Development Research Group, Environment and Energy Team.

Schorcht, F., Kourti, I., Scalet, B. M., Roudier, S., & Sancho, L. D., 2013: Best Available Techniques (BAT) Reference document for the production of cement, lime and magnesium oxide, European Commission Joint Research Centre Institute for Prospective Technological Studies (Report EUR 26129 EN). Luxembourg: Publications Office of the European Union.

Sectoral CDM Benchmarking Methodology by CSI and Ecofys,
www.wbcsd.org/Sector-Projects/Cement-Sustainability-Initiative.

Strand, Jon, 2016: Assessment of Net Mitigation in the Context of International Greenhouse Gas Emissions Control Mechanisms.

Tanaka, K., 2008, Assessment of energy efficiency performance measures in industry and their application for policy. Energy Policy, 36, Seiten 2887-2902.

The Cement Sustainability Initiative, 2012, 10 years of progress – moving on to the next decade,
http://csiprogress2012.org/CSI_ProgressReport_FullReport.pdf.

UNFCCC (United Nations Framework Convention on Climate Change, 2015): Draft Recommendation. Proposed further recommendations on the review of the joint implementation guidelines. Version 01.0. JI-JISC36-AA-A03.

United Nations (2014), World Economic Prospects., available at:
www.un.org/en/development/desa/policy/wesp/wesp_current/2014wesp_country_classification.pdf
[Accessed 10 September 2018].

Vrolijk, Christiaan and Gareth Philips, 2013: Net Mitigation through the CDM. A report for the Swedish Energy Agency.

World Steel Association, 2018a, World Steel Statistics 2017, available at:
www.worldsteel.org/en/dam/jcr:3e275c73-6f11-4e7f-a5d8-23d9bc5c508f/Steel+Statistical+Yearbook+2017.pdf [Accessed 10 September 2018].

World Steel Association:2018b: Benchmarking Systems, available at:
www.worldsteel.org/steel-by-topic/technology/worldsteel-benchmarking-systems.html
[Accessed 12 September 2018].

8 Appendix

BREF/BATC documents

Table 3: BREF/BATC documents and date of publication (where available)

Sector	Date of BREF/ BATC publication	Review process started (yes/no)
Ceramic Manufacturing Industry	08.2007	2018/2019
Common Waste Water and Waste Gas Treatment/ Management Systems in the Chemical Sector	06.2016	
Common Waste Gas Treatment in the Chemical Sector	---	yes
Emissions from Storage	07.2006	
Energy Efficiency	02.2009	
Ferrous Metals Processing Industry	12.2001	yes
Food, Drink and Milk Industries	08.2006	Yes (D1)
Industrial Cooling Systems	12.2001	
Intensive Rearing of Poultry or Pigs	07.2017	
Iron and Steel Production	03.2012	
Large Combustion Plants	07.2017	
Large Volume Inorganic Chemicals – Ammonia, Acids and Fertilisers	08.2007	
Large Volume Inorganic Chemicals – Solids and Others Industry	08.2007	
Manufacture of Glass	03.2012	
Manufacture of Organic Fine Chemicals	08.2006	
Non-ferrous Metals Industries	06.2016	
Production of Cement, Lime and Magnesium Oxide	04.2013	
Production of Chlor-Alkali	12.2013	
Production of Large Volume Organic Chemicals	12.2017	
Production of Polymers	08.2007	
Production of Pulp, Paper and Board	09.2014	
Production of Specialty Inorganic Chemicals	08.2007	
Refining of Mineral Oil and Gas	10.2014	
Slaugtherhouses and Animals By-products Industries	05.2005	yes
Smitheries and Foundries Industry	05.2005	yes
Surface Treatment of Metals and Plastics	08.2006	
Surface Treatment Using Organic Solvents (including Wood and Wood Products Preservation with Chemicals)	08.2007	Yes (D1)
Tanning of Hides and Skins	02.2013	
Textiles Industry	07.2003	yes
Waste Incineration	08.2006	Yes (D1)
Waste Treatment		Yes (FD)
Wood-based Panels Production		

Source: Own compilations based on Lazarus et al. 2013, Obergassel 2017, Vrolijk and Philips 2013, UNFCCC 2015.

Cement: A closer look at the data

CSI has already developed a benchmarking method based on the GNR database in cooperation with Ecofys. The CSI website contains the following six principles that the concept follows:

1. It is intensity-based (measured in CO_2 per ton of clinker and/or CO_2 per ton of cement).
2. It is applicable to project activities combining a number of different greenhouse gas reduction measures and process improvements.
3. It uses statistical information for actual emissions of clinker and cement production in the region and world-wide, sourced from the GNR system.
4. Representative statistical information is used to define in a systematic and objective way the „Business-as-Usual" and baseline emissions, through the „baseline benchmark". The statistical information is also used to evaluate and demonstrate reductions that are additional, through the „additionality benchmark". The baseline benchmark level is based on the current practice, excluding the worst performing facilities. It is updated throughout the duration of the project. The additionality benchmark is representing the top 30% performing facilities and it is set at the starting date of the project activity.
5. The project participant can claim credits only when the emission intensity is below the additionality benchmark level, giving incentive only to projects with high environmental integrity.
6. Newly built facilities are included, but use different benchmarks compared to existing facilities. The approach ensures that the baseline for new cement production facilities is lower than that of existing facilities to acknowledge that a new cement production facility can take the greenhouse gas emission performance directly into account in the design of the facility.

Due to sparse information about the exact design of the benchmark method developed, we cannot finally assess the CSI benchmark. In general, the six principles do not conflict with the requirements developed in this text. However, according to the CSI Report 2012 (10 years of progress – moving on to the next decade), the benchmark method was rejected because no consensus could be found on environmental integrity and business incentives.

Brief overview of Getting the numbers right data

The "Getting the numbers right" data distinguish between six different kiln types, with the most inefficient type (wet / shaft kiln) having average specific emissions of about 1.03 t CO_2/t clinker and the most efficient type (dry with preheater and precalciner) 0.837 t CO_2/t clinker. As described above, a certain type of kiln is required for the production of clinker, depending on the raw material. For example, the semiwet process is used in places where the chalk deposits have a high pit moisture. On average, this type of kiln generates 0.860 CO_2 emissions per ton of clinker worldwide and is therefore less efficient than dry kilns. However, it is required if the raw materials have special properties.

Table 4: Specific emissions (CO_2 emissions/ton of clinker) 2016 by region and kiln type

Region	Dry with preheater and precalciner	Dry with preheater without precalciner	Dry without preheater (long dry kiln)	Mixed kiln type	Semiwet/ semi dry	Wet/ shaft kiln
EU28	0,808	0,831	0,761	0,848	0,843	0,976
Latin America	0,857	0,846		0,894		
World	0,837	0,843	0,840	0,851	0,860	1,030
Austria	0,794	0,832				
Brazil	0,882			0,852		
Canada	0,831	0,797		0,886		
China + Korea + Japan				0,834		
Czech				0,787		
Egypt	0,877					
France		0,833		0,820	0,849	
Germany	0,784			0,826		
India	0,823			0,839		
Italy	0,888	0,845		0,794		
Morocco + Algeria + Tunisia	0,802			0,839		
Philippines				0,864		
Poland	0,837	0,826		0,769		
Spain	0,748	0,833				
Thailand				0,850		
UK	0,845			0,850		
US	0,866	0,866		0,919		1,141

Source: Own illustration based on “Getting the numbers right”

Although a distinction is made between the different types of kilns in the data, there are large differences in specific emissions between countries and regions as can be seen in Table 5. For example in Germany, emissions are around 0.784 t CO_2/t clinker for the most efficient kiln type, while in Italy they are around 0.888 t CO_2/t clinker for the same type. It is striking that there are no clear regional differences in emissions for the two most commonly used kiln types (dryer with preheater and precalcinator, mixed kiln type), which indicate differences in specific emissions in developing, emerging and industrialised countries. For example, the specific emissions for these two kiln types are lower in India, Morocco/Algeria/Tunisia than in Great Britain, the USA or Canada. These observations can have several reasons, which are not necessarily due to inefficient plants. For example, the raw materials used may differ in particular with regard to moisture content of the limestone or the fuels used may also differ. The raw materials used are not clear from the available data, but the data provide information on the use of the fuels, as can be seen in Table 6. The table shows that the ratios of the fuels used in Italy and Germany are very different. The share of classic fossil fuels in Germany is 34%, whereas in Italy it is 86%. This difference can help to explain the above difference in specific emissions for the same kiln type. For a benchmark development it must be examined whether the differences shown in Table 6 are based on the availability of alternative fuels or whether the differences have other reasons such as economic considerations. It should also be examined whether the non-availability is due to political or technical reasons. A correction factor should only be introduced in regions where the availability of alternative fuels is limited. How political restrictions are dealt with must then be clarified on a case-by-case basis.

Table 5: Percentage of fuels used in clinker production in 2016 by region

Region	Alternative fossil and mixed wastes	Biomass	Fossil fuel
EU28	30%	15%	56%
Latin America	9%	5%	86%
World	11%	6%	83%
Austria	59%	17%	24%
Brazil	12%	10%	78%
Canada	9%	3%	87%
China + Korea + Japan	8%	2%	90%
Czech	45%	20%	35%
Egypt	3%	5%	92%
France	23%	16%	60%
Germany	45%	21%	34%
India	3%	1%	97%
Italy	10%	4%	86%
Morocco + Algeria + Tunisia	4%	2%	94%
Philippines	3%	8%	89%
Poland	40%	18%	42%
Spain	16%	11%	73%
Thailand	6%	4%	91%
UK	22%	18%	60%
US	13%	3%	84%

Source: Own illustration based on “Getting the numbers right”

Table 7 shows the average amount of clinker substitutes added to a unit of cement in the respective regions. India and Austria are the top countries with 31% and 30% admixture respectively, whereas in North America with 10% in the US and 12% in Canada there is very little admixture. In the case of cement benchmarking, it must also be examined whether the differences are due to poor availability of substitutes or incompatibilities with regard to the raw material used or whether they are for economic reasons. A correction factor should only be introduced if substitutes are not available or the raw material does not allow the substitutes, i. e. financial reasons must not lead to a correction factor. For the substitutes, too, it must be examined on a case-by-case basis whether or not a correction factor should be introduced due to political unavailability.

Table 6: Additions of clinker substitutes per unit of cement in percent by region

Region	Gypsum	Limestone	Slag	Fly ash	Puzzolana	Others
EU28	4%	6%	8%	2%	1%	2%
Latin America	4%	10%	5%	1%	6%	2%
World	4%	7%	4%	4%	2%	1%
Austria	5%	7%	15%	2%	0%	1%
Brazil	5%	7%	11%	3%	2%	1%
Canada	4%	4%	1%	1%	0%	0%
China + Korea + Japan	4%	11%	4%	1%	1%	0%
Czech	5%	6%	7%	1%	0%	1%
Egypt	7%	4%	2%	0%	0%	0%
France	3%	7%	8%	1%	0%	7%
Germany	5%	4%	12%	1%	0%	5%
India	4%	1%	6%	19%	0%	0%
Italy	4%	13%	1%	2%	2%	3%
Morocco + Algeria + Tunisia	5%	20%	1%	2%	1%	0%
Philippines	5%	5%	2%	1%	9%	0%
Poland	5%	4%	9%	6%	0%	0%
Spain	4%	6%	3%	5%	2%	0%
Thailand	4%	11%	0%	0%	0%	0%
UK	4%	5%	0%	3%	0%	0%
US	5%	3%	1%	0%	0%	1%

Source: Own illustration based on “Getting the numbers right”

Table 8 shows the electricity consumption per ton of cement and the heat consumption per ton of clinker used in the production process. The figures for electricity consumption vary considerably between the regions. As no information is available on the raw material or on the basis of the assessment, we cannot judge the reliability of the data for electricity consumption either.

Table 7: Electricity and heat consumption per tonne of cement/clinker by region

Region	KWh/t Cement	MJ/t Clinker
EU 28	118	3.742
Latin America	103	3.602
World	103	3.536
Austria	115	3.886
Brazil	109	3.561
Canada	135	3.759
China + Korea + Japan	103	3.207
Czech Republic	116	3.850
Egypt	118	3.917
France	121	3.964
Germany	105	3.759
India	75	3.087
Italy	126	3.477
Morocco + Algeria + Tunisia	96	3.518
Philippines	93	3.665
Poland	107	3.791
Spain	166	3.706
Thailand	97	3.297
United Kingdom	117	3.801
United States	135	3.957

Comparing the average specific emissions in Table 5 with the EU ETS benchmark of 0.766 tons of CO_2 per ton of clinker, it becomes clear that the data from Getting the Numbers right can provide a good first indication. The EU ETS benchmark is significantly lower than the average values of the regions, but not too far away, so that the data from Getting the Numbers right appear plausible. Nevertheless, the data alone can only form a basis and rough indication. The above mentioned problem that no data are available at the plant level, as well as the fact that the data are mainly based on clinker and not on cement, makes it difficult to base the benchmark development on these data only. Perhaps the data can be used in combination with further data sets or expert assessments for benchmark development.

German Emissions Trading Authority (DEHSt) at the German Environment Agency
Bismarckplatz 1
D-14193 Berlin

www.dehst.de/EN | emissionstrading@dehst.de