HAL
open science

The accountability imperative for quantifiying the uncertainty of emission forecasts : evidence from Mexico

Daniel Puig, Oswaldo Morales-Nápoles, Fatemeh Bakhtiari, Gissela Landa

To cite this version:

Daniel Puig, Oswaldo Morales-Nápoles, Fatemeh Bakhtiari, Gissela Landa. The accountability imperative for quantifiying the uncertainty of emission forecasts : evidence from Mexico. 2017. hal-03389325

HAL Id: hal-03389325

https://sciencespo.hal.science/hal-03389325v1

Preprint submitted on 20 Oct 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The accountability imperative for quantifying the uncertainty of emission forecasts: evidence from Mexico

Daniel Puig, Oswaldo Morales-Nápoles, Fatemeh Bakhtiari,
Gissela Landa Rivera

SCIENCES PO OFCE WORKING PAPER n° 17, 2017/09/15

EDITORIAL BOARD

Chair: Xavier Ragot (Sciences Po, OFCE)

Members: Jérôme Creel (Sciences Po, OFCE), **Eric Heyer (**Sciences Po, OFCE), **Lionel Nesta** (Université Nice Sophia Antipolis), **Xavier Timbeau (**Sciences Po, OFCE)

CONTACT US

OFCE
10 place de Catalogne | 75014 Paris | France
Tél. +33 1 44 18 54 87

www.ofce.fr

WORKING PAPER CITATION

This Working Paper:
Daniel Puig, Oswaldo Morales-Nápoles, Fatemeh Bakhtiari, Gissela Landa Rivera,
Labour Force Participation and Job Polarization:
Evidence from Europe during the Great Recession, *Sciences Po OFCE Working Paper*, n°17, 2017-09-15.
Downloaded from URL : www.ofce.sciences-po.fr/pdf/dtravail/WP2017-17.pdf
DOI - ISSN

© 2017 OFCE

ABOUT THE AUTHORS

Daniel Puig*, UNEP DTU Partnership, Copenhagen, Denmark
Email Address: dapu@dtu.dk

Oswaldo Morales-Nápoles, Delft University of Technology, Netherlands
Email Address: o.moralesnapoles@tudelft.nl

Fatemeh Bakhtiari, UNEP DTU Partnership, Copenhagen, Denmark
Email Address: fatebak@dtu.dk

Gissela Landa Rivera, OFCE, Sciences Po Paris, France
Email Address: gissela.landa@sciencespo.fr

ABSTRACT

Governmental climate change mitigation targets are typically developed with the aid of forecasts of greenhouse-gas emissions. The robustness and credibility of such forecasts depends, among other issues, on the extent to which forecasting approaches can reflect prevailing uncertainties. We apply a transparent and replicable method to quantify the uncertainty associated with projections of gross domestic product growth rates for Mexico, a key driver of greenhouse-gas emissions in the country. We use those projections to produce probabilistic forecasts of greenhouse-gas emissions for Mexico. We contrast our probabilistic forecasts with Mexico's governmental deterministic forecasts. We show that, because they fail to reflect such key uncertainty, deterministic forecasts are ill-suited for use in target-setting processes. We argue that (i) guidelines should be agreed upon, to ensure that governmental forecasts meet certain minimum transparency and quality standards, and (ii) governments should be held accountable for the appropriateness of the forecasting approach applied to prepare governmental forecasts, especially when those forecasts are used to derive climate change mitigation targets.

KEY WORDS

Uncertainty, projections, structured expert judgement, accountability, emission-reduction targets, gross domestic product growth rates

JEL

(*) corresponding author.

POLICY INSIGHTS

- No minimum transparency and quality standards exist to guide the development of greenhouse-gas emission scenario forecasts, not even when these forecasts are used to set national climate change mitigation targets.
- No accountability mechanisms appear to be in place at the national level to ensure that national governments rely on scientifically sound processes to develop greenhouse-gas emission scenarios.
- Using deterministic forecasts to underpin emission reduction targets is defective from a scientific standpoint, in that those forecasts are subject to large uncertainties, which calls into question the validity of the target.
- Setting up minimum transparency and quality standards, and holding governments accountable for their choice of forecasting methods could lead to more robust emission reduction targets nationally and, by extension, internationally.

ACKNOWLEDGEMENTS

We thank Roger Cooke and Jorge Gutiérrez for comments on an early draft of the manuscript.

We thank three anonymous reviewers and the editor for helpful comments.

The work of Oswaldo Morales-Nápoles and Gissela Landa Rivera was partly supported by the French Development Agency.

The accountability imperative for quantifying the uncertainty of emission forecasts: evidence from Mexico

Peer reviewed version of the work has now been published by Climate Policy
www.climatepolicy.com

1. Introduction

The vast majority of the parties to the United Nations Framework Convention on Climate Change (UNFCCC) have committed to specific greenhouse-gas emission-reduction targets by 2030 (INDCs 2017). The level of ambition of each party's individual commitment, and the extent to which commitments are actually implemented, will determine whether the UNFCCC goal can be met (UNEP 2016, UNFCCC 2016).

Several of the largest developing-country parties to the UNFCCC have defined their greenhouse-gas emission-reduction targets relative to a forecast 'baseline' level in 2030 – that is, the emission level that would be reached in 2030 if the policies and measures directed at meeting the target were not implemented (INDCs 2017).[1] For these parties, two factors determine the level of ambition of their individual targets (Puig *et al.* 2013): the stringency of the target, measured as the extent to which the target represents a departure from those 'baseline' emission levels; and the robustness and credibility of the forecasts that describe the 'baseline' emission levels, measured in terms of the appropriateness of the forecasting approach used.[2]

While both determinants matter with regard to measuring the level ambition of the individual targets, the former receives far more scrutiny than the latter (UNEP 2016, UNFCCC 2016). Nonetheless, the latter is by no means negligible. Consider a percentagewise emission reduction target that applies to a forecast of 'baseline' greenhouse-gas emission levels in a certain future year: depending on whether that forecast is high or low, the emission reductions target will represent, respectively, a low or a high level of actual emission reductions in that future year.

Therefore, eminently technical issues, such as the appropriateness of the assumptions made, and methods used, to calculate emission forecasts, have a major influence on the outcomes of the largely non-technical decisions made by parties to the UNFCCC at their annual conferences.[3] Yet, in spite of the stakes, no requirements or even generic (voluntary) approaches exist to inform greenhouse-gas emissions forecasting, although a limited amount of guidance can be drawn from two United Nations-sponsored initiatives (IPCC 2006, CGE 2016).

This raises two questions about forecasts of greenhouse-gas emissions. Firstly, to what extent do the assumptions made, and methods used, by governments to prepare official forecasts stand scientific scrutiny? Secondly, to what extent are governments held accountable for the appropriateness of the forecasting approaches on which they rely?

Both questions are under-researched. Limited evidence from both developed and developing countries indicates that (i) forecasting assumptions are undisclosed or poorly documented, and (ii) forecasting methods are undisclosed or rudimentary (Clapp *et al*. 2009, Clapp and Prag 2012, BURs 2017). We have been unable to identify any literature that documents accountability requirements on governments with regard to their choice of forecasting approaches, which suggests that such requirements are lacking.

We examine the methods used by the Mexican government to obtain forecasts of greenhouse-gas emissions.[4] This allows us to answer, for the case of Mexico, the two questions introduced above. In addition, we contrast Mexico's governmental deterministic forecasts of greenhouse-gas emissions with our probabilistic forecasts

[1] In addition to Mexico, several other countries have defined their NDC targets in terms of emission reductions against a future 'reference' level. In Latin America, Colombia, Costa Rica and Peru, among others, have followed the same approach. In Africa, the list includes Ethiopia, Ghana and Kenya, for example. In Asia, Indonesia, Thailand and Vietnam are the largest among the countries having followed this approach.

[2] NDCs are framed around different types of targets, notably reductions relative to a baseline level (as is the case in Mexico), reductions per unit of gross domestic product, or implementation of specific policies and measures. While uncertainty analysis is relevant to all types of NDCs, several approaches can be used to conduct such analysis. The approach presented in this paper, centred in quantifying the uncertainty around projections of gross domestic product, is directly applicable to NDCs structured around the emissions intensity of the economy (that is, NDCs whose targets are defined as emission reductions per unit of gross domestic product).

[3] While the choice of assumptions made, and projections used, is at least partly a political matter, the extent to which those choices are appropriate from a scientific standpoint is what we term an eminently technical issue. The latter (appropriateness) is the main subject of our analysis.

[4] The work presented in this paper was conducted as a part of two bilateral development cooperation programmes. Development priorities on the part of the donor countries account for the choice of beneficiary country (Mexico).

for the country.[5] This allows us to judge the robustness of the governmental forecasts.[6]
The Mexican government's efforts in this area deserve praise, in that they go beyond what appears to be standard practice (Puig 2015). Nonetheless, our findings highlight a number of shortcomings of the Mexican government's approach to greenhouse-gas emission forecasting.[7]
We conclude by hypothesizing about the implications that our findings may have with regard to assessing the collective effect of national-level greenhouse-gas emission reduction targets. This leads us to make two recommendations. Firstly, transparency standards should be introduced, to ensure a minimum level of quality with regard to the forecasting approaches used, and promote comparability among forecasts from different governments. Secondly, emission reduction targets should be derived from probabilistic forecasts which, unlike deterministic forecasts, reflect some of the uncertainty associated with future greenhouse-gas emission levels.
The remainder of the paper is structured in five sections. Section 2 introduces three greenhouse-gas emission scenarios prepared by the government of Mexico. The text focuses on the assumptions made in each scenario concerning future gross domestic product growth rates. Section 3 summarises the main findings of an elicitation of expert judgement aimed at obtaining estimates of short- and mid-term gross domestic product growth rates for Mexico. Section 4 presents greenhouse-gas emission forecasts for Mexico, obtained using an energy-economy model calibrated to follow economic pathways consistent with the estimates of gross domestic product growth rates described in section 3. Section 5 analyses the accountability implications of the government of Mexico's approach to greenhouse-gas emission forecasting. Section 6 discusses the impact on international climate change negotiations of improving the way national government forecasts of greenhouse-gas emissions are prepared.

2. Greenhouse-gas emission scenarios in Mexico

Limiting emissions of greenhouse gases has become a major public policy concern in most countries. With its 2012 General Law on climate change, Mexico has become a leader among developing countries with regard to planning for climate change mitigation (Nachmany *et al.* 2015).
Over the past decade government, academia and the private sector in Mexico, all have prepared greenhouse-gas emission scenarios in support of various planning processes (Puig 2015). Methods and assumptions vary across these scenarios and, as a result, forecasts of likely future levels of greenhouse-gas emissions differ.
For an industrialised country like Mexico, a key forecasting assumption is the choice of projections of gross domestic product growth rates (Puig *et al.* 2013).[8] In the period between 2012 and 2015 the government of Mexico issued three sets of forecasts of greenhouse-gas emissions. Each set relies on a different projection of gross domestic product annual growth rates:

- **Fifth 'National Communication' to the United Nations Framework Convention on Climate Change** (SEMARNAT 2012). The forecasts presented are based on the assumption that, in the period between 2006 and 2020, the average annual growth rate for gross domestic product will be 2.3 per cent. This rate is said to be based on "historical trends", with no further information given about it. No projection of gross domestic product growth rates is provided for the period 2021-2030, even though the document does include forecasts of greenhouse-gas emissions for that period.

- **'National Strategy on Climate Change'** (SEMARNAT 2013). At present this document constitutes the only official source of forecasts of greenhouse-gas emissions for Mexico. The forecasts presented are

[5] In the context of this paper, probabilistic forecast refers to a forecast that is expressed as a continuous variable, with a probability of occurrence attached to each of its possible values. This contrasts with the concept of deterministic forecast, which is defined as a forecast expressed as a single value, with an implicit certainty of occurrence.
[6] Our analysis focuses on one source of uncertainty: gross domestic product growth rates. For a country with the industrialisation level found in Mexico, and as highlighted in section 2, economic growth is a major determinant of uncertainty in greenhouse-gas emission forecasts.
[7] The literature on this topic is extremely limited. A recent similar analysis, conducted in South Africa, suggests that the shortcomings identified in this paper may be common to other countries (Durbach *et al*. 2017).
[8] Within the time frames considered, the growth rate for gross domestic product is the major source of uncertainty when it comes to forecasting future greenhouse-gas emission levels in a country with the industrialisation level found in Mexico. The impact of technical innovations that can reduce the greenhouse-gas emissions intensity of the economy would only be relevant if longer time frames – several decades – were involved. Such technical innovations would typically entail larger than currently experienced increases in (i) the share of renewable energy sources in the fuel mix, or (ii) in energy efficiency across the economy. Given the time frames of the analysis, this type of considerations are excluded.

based on the assumption that, in the period between 2010 and 2030, the average annual growth rate for gross domestic product will be 3.6 per cent. This is the rate used to calculate Mexico's official projections for electricity demand (SENER 2012). The document describing the electricity demand projections considers three growth scenarios for gross domestic product – low, medium and high – and states that the medium scenario, corresponding to an annual growth rate of 3.6 per cent, is believed to be more likely. No information is provided as to why this is so, or how the 3.6 per cent figure was calculated.

- **'Intended Nationally Determined Contribution', submitted to the United Nations Framework Convention on Climate Change** (INDC Portal 2017).[9] The forecasts presented are based on the assumption that, in the periods between 2014 and 2020, and 2021 and 2030, the annual growth rates for gross domestic product will be, respectively, 3.37 per cent and 3.85 per cent (Jorge Gutiérrez, Mexico's National Institute of Ecology and Climate Change, personal communication, 2015). These figures were obtained through an expert judgement elicitation conducted in 2014 (Supplementary Information SI.1). The elicitation was framed around three economic scenarios. In the 'neutral' (or 'medium economic growth') scenario, gross domestic product would grow by a median annual rate of 3.36 per cent in the period between 2014 and 2020, and by 3.88 per cent in the period between 2020 and 2030.

In addition to mitigation-scenario forecasts of greenhouse-gas emissions, the three documents referred to above present reference-scenario forecasts.[10] For a given year (2020 or 2030), these reference-scenario forecasts differ by between two and thirty percent, depending on the combination of scenarios considered (Table 1). Since greenhouse-gas emission forecasts across all three reference scenarios ought to be largely comparable, the variations observed respond to differences in forecasting assumptions (in particular, estimates of gross domestic product growth rates) and forecasting methods.

Table 1: Assumptions about gross domestic product growth rates, by Mexican government planning scenario

Policy process in support of which the scenario was prepared	Assumed annual growth rate for gross domestic product	Analytical method behind the assumption	Reference scenario forecasts ($MtCO_2e$)	
			2020	2030
Fifth National Communication (2012)	2.3 % (2006-2020)	Following "historical trends"	872	996
National Strategy on Climate Change (2013)	3.6 % (2010-2030)	Unspecified	960	1,276
Intended Nationally Determined Contribution (2015)	3.37 % (2014-2020) 3.85 % (2021-2030)	Expert judgement elicitation	792	973

Irrespective of whether they correspond to reference or mitigation scenarios, the robustness of (the Mexican) forecasts of greenhouse-gas emissions depends to a great extent on the robustness of the projections of gross

[9] Mexico's non-conditional NDC target represents a 25 percent reduction in greenhouse-gas emissions by 2030, compared to Mexico's official 'reference' emission levels in the same year.

[10] There are two main types of greenhouse-gas emission scenarios: reference scenarios and mitigation scenarios. In contraposition to mitigation scenarios, reference scenarios assume no additional policy efforts, compared to present-day emission-reduction policies.

For a given future year, the forecasts derived from different scenarios will be comparable if the scenarios consider the same sectors in the same regions, have the same (or analogous) base years, and share key assumptions. In most instances mitigation scenarios consider different sets of emission-reduction policies. Therefore, in general, full comparability across mitigation scenarios is rare. Reference scenarios, on the other hand, ought to be largely comparable if the above premises are met.

Contrasting forecasts from comparable scenarios makes it possible to assess the impact on forecasts of the forecasting assumptions and forecasting methods chosen. In this article, we compare reference scenario forecasts.

domestic product growth rates used to calculate those forecasts (Puig *et al.* 2013). Because projections of gross domestic product growth rates are surrounded with uncertainty, quantifying that uncertainty represents one of the best ways of increasing the robustness of the projections (Morgan *et al*. 2009). By extension, this makes it possible to increase the robustness of the associated forecasts of greenhouse-gas emissions.

3. Quantifying the uncertainty associated with gross domestic product growth rates

We use structured expert judgement to quantify the uncertainty associated with projections of gross domestic product growth rates in Mexico (Cooke 1991). To keep the elicitation process manageable, while taking due account of the strong dependencies between gross domestic product and a number of other macro-economic variables, we structure the elicitation around different scenarios of economic growth (Morales-Nápoles 2014).

3.1 Developing scenarios of economic growth

We use an econometric model for gross domestic product in Mexico, which was built specifically for this work (Loría 2013). The outputs of the model underpin three economic scenarios: 'pessimistic' (low economic growth), 'neutral' (medium economic growth) and 'optimistic' (high economic growth). Each scenario is defined by a plausible combination of values for variables such as interest rates, unemployment, and inflation in Mexico, and economic growth in the United States.[11]

The scenario approach is chosen because of the strong interdependencies among variables. By capturing those interdependencies through scenarios (as opposed to eliciting experts on all individual variables potentially influencing gross domestic product), it is possible to avoid a cumbersome post-hoc dependence analysis.

Table 2: Results of the expert judgement elicitation for gross domestic product growth rates (by scenario of economic growth and percentile), and corresponding ThreeME-model forecasts of greenhouse-gas emission levels

Percentile	Gross domestic product annual growth rates (percent)		Reference scenario forecasts ($MtCO_2e$)	
	2014-2020	2021-2030	2020	2030
'Pessimistic' scenario				
5th	1.23	1.60	582	583
50th	2.44	2.79	613	694
95th	3.20	3.69	633	781
'Neutral' scenario				
5th	1.79	2.85	596	663
50th	3.36	3.88	638	801
95th	4.10	4.50	658	883
'Optimistic' scenario				
5th	3.85	3.13	651	790
50th	4.58	4.84	671	937
95th	5.80	5.90	889	1,102

Notes: Projections of annual growth rate for gross domestic product are obtained through expert judgement elicitation. 'Reference scenario forecasts' refers to emissions of all greenhouse gases covered in the Kyoto protocol to the United Nations Framework Convention on Climate Change, and excludes the net impact of agriculture, land-use change and forestry, and emissions from waste management. Forecasts are obtained using the ThreeME general-equilibrium model.

[11] Economic growth in the United States is a key determinant of economic growth in Mexico.

3.2 Eliciting expert judgements on gross domestic product growth rates

We rely on the so-called Cooke method of expert judgement elicitation, which involves external validation of experts' opinions (Supplementary Information SI.1 summarises the main features of the model) (Cooke 1991).[12] We elicit probability distributions from nine experts, through a three-day workshop.[13] The variable elicited are gross domestic product growth rates (the 5th, 50th and 95th percentiles). Experts are asked to provide uncertainty estimates for each of the three scenarios referred to above.[14]

Through the scenarios, experts are explicitly confronted with a wide range of plausible economic growth pathways. This allows us to obtain a similarly wide range of estimates, thus characterising more fully the uncertainty around likely future trends in gross domestic product growth rates (Table 2). Rather than focusing on individual years, we elicit uncertainty around average rates in the time periods 2014-2020 and 2021-2030 (Supplementary Information SI.1).

3.3 Comparing our results with the projections used in Mexican government forecasts

The estimates of annual growth rate for gross domestic product used in the 2012 (SEMARNAT 2012) and 2013 (SEMARNAT 2013) Mexican government scenarios are consistent with the estimates associated with, respectively, our 'pessimistic' and 'neutral' scenarios (Table 1 and Table 2). We note that the methods used by the government of Mexico to obtain the projections of gross domestic product growth rates used in the 2012 and 2013 documents referred to above do not stand scientific scrutiny (Table 1) (Oppenheimer *et al*. 2016).

The estimates used in the 2015 Mexican government scenario (INDC Portal 2016) were derived from our 'neutral' scenario, and are nearly identical to those associated with that scenario (Table 3). Structured expert judgement, the method used to obtain the estimates in all our scenarios, is transparent and replicable (Aspinall 2010).

Table 3: Calculated annual growth rates in reference-scenario forecasts of greenhouse-gas emissions, and assumptions about gross domestic product growth rates, by scenario

Scenario	Reference scenario forecasts ($MtCO_2e$)		Annual growth rates (percent)		Ratio
	2020	2030	Emissions (2020-2030)	GDP (2021-2030)	
INDC (2015)	632	798	2.4	3.85	0.61
ThreeME pessimistic	613	694	1.2	2.79	0.45
ThreeME neutral	638	801	2.3	3.88	0.59
ThreeME optimistic	671	937	3.4	4.84	0.70

- Notes: 'Reference scenario forecasts' refers to emissions of all greenhouse gases covered in the Kyoto protocol to the United Nations Framework Convention on Climate Change, and excludes the net impact of agriculture, land-use change and forestry, and emissions from waste management. 'Ratio' refers to the quotient of greenhouse-gas emissions growth rates and gross domestic product growth rates. For each of the three scenarios considered, ThreeME forecasts correspond to estimate obtained using the 50th percentile for the projection of gross domestic product growth rates.
- Acronyms: INDC stands for 'Intended Nationally Determined Contribution', ThreeME refers to the forecasts obtained through the ThreeME model, and GDP stands for 'gross domestic product'.

[12] The Cooke method combines the various assessments from experts into one single probability distribution. It does this through differential weighting of experts' assessment, on the basis of measures of each expert's performance when responding to a set of so-called calibration questions (Supplementary Information SI.1).

[13] A total of 19 variables were assessed, of which 13 were calibration variables.

[14] The reader may notice in the second column of table S1 that the calibration scores of experts 6 and 9 are >0.05, which is typically the threshold level to render expert opinions as statistically accurate. Moreover, the equal weighting of expert opinions is also slightly above 0.05, while the performance-based combination has a higher calibration score (0.614) than the equal-weight combination, and every individual expert. One example of calibration variable and variable of interest is presented in SI.1. The format of other seed questions and variables of interest follows the format of those presented in the supplementary information. The format chosen for this research is not different from that used in other areas where the method has been applied in the past. See for example Cooke et al (2007) for an example of seed variables vs. variables of interest in epidemiology.

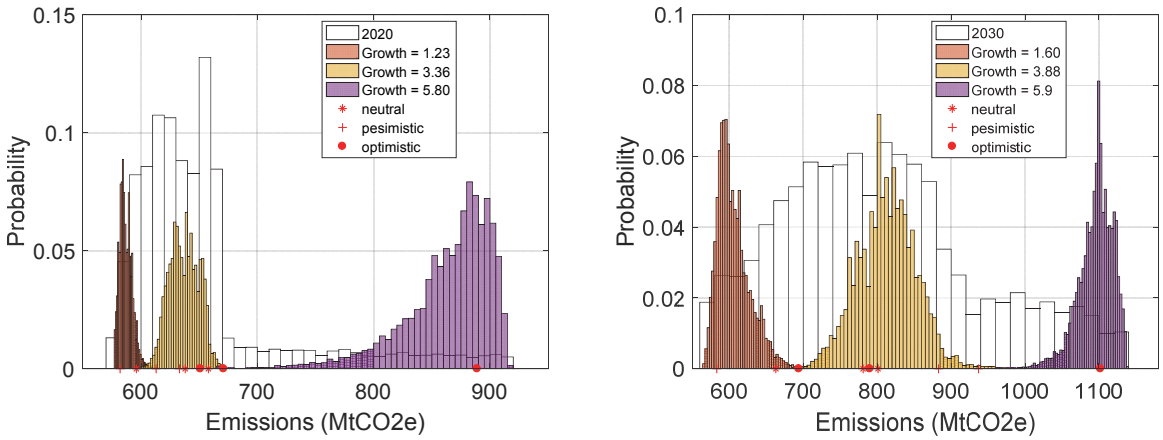
4. Calculating probabilistic reference-scenario forecasts of greenhouse-gas emissions

We use the Multi-sector Macroeconomic Model for the Evaluation of Environmental and Energy policy (hereinafter the ThreeME model) to obtain forecasts of greenhouse-gas emissions in Mexico for the years 2020 and 2030 (Landa *et al*. 2016, Callonec *et al*. 2013).[15] We run the model several times, each time using a different economic pathway. Each of the pathways used is consistent with one of the various estimates of gross domestic product growth rates that we obtained through expert judgement elicitation (Supplementary Information SI.1). This allows us to produce a range of forecasts of greenhouse-gas emissions that reflects the uncertainty associated with the full range of gross domestic product growth rate estimates obtained through structured expert judgement (Table 2). For 2030 the range of forecasts spans from 583 $MtCO_2e$ ('pessimistic' scenario, 5th percentile) to 1,102 $MtCO_2e$ ('optimistic' scenario, 95th percentile).

The ThreeME model excludes emissions from agriculture, land-use change, and waste management (Supplementary Information SI.2). Since the three Mexican government scenarios referred to above include emissions from these sectors, the ThreeME model forecasts are not directly comparable to either of the forecasts resulting from those scenarios. To make comparisons possible, emissions from agriculture, land-use change, and waste management have to be excluded from the governmental scenario forecasts. This can only be done for the 2015 scenario, as this is the only scenario that provides sector-specific forecasts (INECC 2016). We find that the 2015 Mexican government forecasts are fully consistent with the ThreeME forecasts corresponding to the median estimate in the 'neutral' scenario (Table 3).

To obtain probability distributions for greenhouse-gas emissions in Mexico, we analyse the statistical dependence of the ThreeME forecasts (Table 2). The rank correlation of the 18 estimates is approximately 0.96. In this case, a Gaussian copula is a valid model for statistical dependence between gross domestic product growth rates and greenhouse-gas emissions (Joe 2014). The marginal distributions of gross domestic product growth rates and greenhouse-gas emissions for 2020 and 2030 can be calculated on the basis of the combined distributions obtained through structured expert judgment, assigning equal probability to the three scenarios (section 3.2). Summary results are presented in the main text (Figure 1), while a fuller description of the approach is included in annexes (Supplementary Information SI.3).[16]

Figure 1: Unconditional and conditional uncertainty distributions associated with, greenhouse-gas emission forecasts obtained with the ThreeME model and structured expert judgment.



Source: Own elaboration

5. Accountability considerations of governments' choice of forecasting approach

Our results show that the range – and, therefore, the uncertainty – of plausible values for greenhouse-gas emission levels in Mexico in 2030 is large (Table 2, Figure 1). We conjecture that greenhouse-gas emission forecasts for most other industrialised countries are characterised by similarly large uncertainty ranges.

Had we analysed more sources of uncertainty, in all likelihood the resulting uncertainty range would have been even larger. Notwithstanding, experience shows that, when it comes to model assumptions, annual growth rates for gross domestic product are, with energy prices, the main sources of uncertainty in greenhouse-gas

[15] A dedicated two-year calibration and verification exercise was conducted, to ensure that the ThreeMe model could be meaningfully used to assess likely future levels of greenhouse-gas emissions in Mexico (Supplementary Information SI.2). This exercise benefited from input by a number of (mostly Mexican) forecasters, and users of emission forecasts in Mexico (mostly governmental agencies) (Landa *et al*. 2016).

[16] We chose a combination of scenario-based analysis and probabilistic methods, which makes quantification easier, and thus can be more relevant from a policy standpoint.

emissions scenarios in industrialised countries.[17]

Several tools are available to estimate plausible future values for some uncertain variables (Morgan *et al.* 2009). Expert judgement elicitation is one such tools. In addition to being transparent and replicable, the appeal of expert judgement elicitation lies in the probabilistic nature of its outputs, which facilitates the use of probabilistic forecasting methods.

Concerns have been voiced over the "abuse" of expert judgement elicitation in support of decision-making for public policy (Morgan 2014). Clearly, like any other analytical tool, expert judgement elicitation has its limitations. Notwithstanding, when 'best guesses' are the alternative, well-conducted expert judgement elicitation is a better choice.

The government of Mexico quantified the uncertainty associated with trends in gross domestic product growth rates, one key driver of greenhouse-gas emissions. However, this uncertainty was not incorporated into the official forecast of greenhouse-gas emissions. This is because deterministic (as opposed to probabilistic) forecasting methods were used.

Forecasts of greenhouse-gas emissions that fail to incorporate uncertainty are defective scientifically and, when they are used to calculate targets, they are inappropriate from a policy standpoint (Oppenheimer *et al*. 2016).[18] Yet, there appear to be no requirements on governments with regard to the need to incorporate uncertainty in governmental forecasts, as evidenced by the lack of literature on this topic. In fact, by taking the initiative to quantify the uncertainty in gross domestic product growth rates (and oil and gas prices), the government of Mexico went beyond what appears to be standard practice (Clapp *et al.* 2009, Clapp & Prag 2012).

We argue that governments should be held accountable for their choice of forecasting approach. This would entail closer scrutiny of both the assumptions made, and the methods used, to produce governmental forecasts. Specifically, this would call for science-based governmental forecasts, including a quantification of the uncertainty associated with the forecasts.

It is for scientists to select the most appropriate analytical methods, to prepare the best projections that science can offer, and to communicate them properly to decision makers (Fischhoff & Davis 2014). In turn, the role of decision-makers is to ensure that public policy relies on transparent and replicable analytical processes, reflecting the full implications of the findings emerging from those processes. Specifically, in the context of the work described in this paper, it is for decision-makers to ensure that uncertainty estimates are incorporated in a meaningful way in the target-setting process.

6. Discussion

In most instances greenhouse-gas emission-reduction commitments by parties to the UNFCCC are based on governmental forecasts of greenhouse-gas emissions. Therefore, indirectly, the integrity of the international climate change regime rests on the robustness and credibility of those forecasts. Yet, the extent to which forecasts are robust and credible is generally unknown: forecasting approaches are often undisclosed, and accountability requirements on governments appear to be lacking. Against this background the need for minimum transparency and quality standards for forecasting approaches appears self-evident.[19]

We argue that the UNFCCC should champion the development of such standards. They would foster comparability between emission-reduction commitments from different parties, which could help ensure that individual UNFCCC party commitments are "fair and ambitious". Fairness and ambition are two of the fundamental tenets of the international climate change regime: sacrificing the flexibility that parties currently enjoy with regard to the forecasting approaches that they use for target-setting seems a price worth paying to safeguard those premises.

---

[17] For industrialised countries where land-use change is not a major source (or sink) of greenhouse-gas emissions, and barring unprecedented demographic changes, variations in gross domestic product and energy prices are the main drivers of emissions. Estimating the impact of changes in energy prices in Mexico's emissions of greenhouse gases is challenging. Firstly, energy prices are typically volatile, which makes them notoriously difficult to forecast. Secondly, when international oil prices are high, domestic oil consumption in Mexico is replaced partly by natural gas, as oil is a major export commodity for the country. The extent to which that replacement takes place, and the impact of oil prices on the overall energy consumption level in Mexico are similarly difficult to forecast. We conducted some of the work required to estimate the impact of oil and gas prices in greenhouse-gas emissions in Mexico (Supplementary Information 1). However, the results are not robust enough to warrant inclusion in the paper.

[18] From a scientific point of view, uncertainty quantification should be an integral part of any forecasting exercise (Puig 2015, Clapp & Pragg 2012,). From a policy point of view, however, uncertainty quantification may be less important, if the choice of target is not informed by some (uncertain) forecast of future emission levels, as is the case for emission reduction targets expressed as absolute emission levels in a future year (as opposed to emission reductions relative to a 'reference' level in that future year).

[19] The same, or similar, standards could be applied to non-state actor actions relevant to climate change mitigation. Indeed, (limited) accountability is one of the main shortcomings associated with most non-state actor actions (Bakhtiari 2017).

We further argue that greenhouse-gas emission-reduction commitments by parties to the UNFCCC should reflect the uncertainty associated with future levels of greenhouse-gas emissions. This could mean that individual UNFCCC party emission-reduction targets are derived from probabilistic forecasts of greenhouse-gas emissions. For example, UNFCCC party targets could be expressed as a range of greenhouse-gas emission levels, given by pre-defined quantiles of their uncertainty distribution. This could be associated with a small number of socio-economic scenarios: should certain socio-economic developments come to pass (notably in terms of gross domestic product growth rates and energy prices), a higher or lower value of the emission reductions target would apply.

With the Paris Agreement, the international climate change regime has strengthened its accountability requirements. Specifically, the so-called transparency framework under the Paris Agreement calls for increased scrutiny and disclosure of national reporting to the UNFCCC. As the provisions of the transparency framework come into effect, the need for reflecting uncertainty in national greenhouse-gas emission reduction targets is likely to become more apparent. Characterising and quantifying uncertainty, and communicating uncertainty to decision makers are areas where additional research is required.

In the area of climate change, key issues on which additional research is needed include:

- Which methods are appropriate to forecast uncertainty distributions, for which types of variables.
- How different uncertainties should be aggregated into a single distribution.
- How the latter can meaningfully inform the policy process.

REFERENCES

Aspinall, W. (2010). A route to more tractable expert advice. *Nature, 463*(7279), 294-295.

Bakhtiari, F. (2017). *International Cooperative Initiatives and the United Nations Framework Convention on Climate Change*. Climate Policy, 1-9.

BURs (2017). *Submitted biennial update reports (BURs) from non-Annex I Parties*. (n.d.). Retrieved April 30, 2017 from http://unfccc.int/national_reports/non-annex_i_natcom/reporting_on_climate_change/items/8722.php

Callonec, G., Landa, G., Malliet, P., Reynès, F., Yeddir-Tamsamani, Y. (2013). *A full description of the Three-ME model: multi-sector macroeconomic model for the evaluation of environmental and energy policy*. French Economic Observatory. Paris, France.

CGE (2016). *Training Materials for the Preparation of National Communications from non-Annex I Parties*. (n.d.). Retrieved May 07, 2016 from http://unfccc.int/national_reports/non-annex_i_natcom/training_material/methodological_documents/items/349.php

Clapp, C., Karousakis, K., Buchner, B., & Chateau, J. (2009). *National and sectoral GHG mitigation potential: a comparison across models*. Organisation for Economic Cooperation and Development. Paris, France.

Clapp, C. & Prag, A. (2012). *Projecting emissions baselines for national climate policy: options for guidance to improve transparency*. Organisation for Economic Cooperation and Development. Paris, France.

Cooke, R.M. (1991). *Experts in Uncertainty: Opinion and Subjective Probability in Science* Oxford University Press. Oxford, United Kingdom.

Cooke, R.M., Wilson, A.M. Tuomisto, J.T. Morales, O. Tainio, M. and Evans, J.S. (2007). A Probabilistic Characterization of the Relationship between Fine Particulate Matter and Mortality: Elicitation of European Experts. *Environ. Sci. Technol., 41*(18), pp 6598–6605

Durbach, I., Merven, B., & McCall, B. (2017). Expert elicitation of autocorrelated time series with application to e3 (energy-environment-economic) forecasting models. *Environmental Modelling & Software, 88*, 93-105.

Fischhoff, B., & Davis, A. L. (2014). Communicating scientific uncertainty. *Proceedings of the National Academy of Sciences, 111*(Supplement 4), 13664-13671.

INDC Portal (2016). (n.d.). *Intended Nationally Determined Contributions Portal*. (n.d.). Retrieved May 11, 2016 from http://unfccc.int/focus/indc_portal/items/8766.php

INDCs (2017). *Intended Nationally Determined Contributions (INDCs)*. (n.d.) Retrieved April 30, 2017 from http://unfccc.int/files/adaptation/application/pdf/all__parties_indc.pdf

INECC (2016). *Instituto Nacional de Ecología y Cambio Climático*. (n.d.) Retrieved June 01, 2016, from http://www.inecc.gob.mx/descargas/difusion/2015_mex_indc_presentacion.pdf

IPCC (2006). *IPCC Guidelines for national greenhouse gas inventories*. Intergovernmental Panel on Climate Change. Geneva, Switzerland.

Joe, H. (2014). *Dependence Modeling with Copulas*. Chapman & Hall/CRC. London, United Kingdom.

Landa, G., Reynès, F., Islas, I., Bellocq, F.X., Grazi, F. Towards a low carbon growth in Mexico: is a double dividend possible? A dynamic general equilibrium assessment. *Energy Policy 96*, 314-327 (2016).

Loría, E. (2013). *Determinantes económicos del PIB total y sectorial de la economía mexicana, 1980-2012*. National Institute of Ecology and Climate Change. Mexico D.F., Mexico.

Morales-Nápoles, O. (2014): *Estimaciones de incertidumbre de largo plazo sobre variables económicas a través de expertos*. Reporte de resultados. TNO. Delft, Netherlands.

Morgan, G. M., Henrion, M., Keith, D., Lempert, R., McBride, S., Small, M., & Wilbanks, T. (2009). *Best practice approaches for characterizing, communicating, and incorporating scientific uncertainty in climate decisions. Synthesis and Assessment Product 5.2*. US Climate Change Science Program. Lead Agency: National Oceanic and Atmospheric Administration. Wahington, DC., United States of America.

Morgan, M. G. (2014). Use (and abuse) of expert elicitation in support of decision making for public policy. *Proceedings of the National Academy of Sciences*, *111*(20), 7176-7184.

Nachmany, M., Fankhauser, S., Davidová, J., Kingsmill, N., Landesman, T., Roppongi, H., Schleifer, P., Setzer, J., Sharman, A., Singleton, C.S. & Sundaresan, J. (2015) *The 2015 global climate legislation study: a review of climate change legislation in 99 countries: summary for policy-makers*. GLOBE International and London School of Economics. London, United Kingdom.

Oppenheimer, M., Little, C. M., & Cooke, R. M. (2016). Expert judgement and uncertainty quantification for climate change. *Nature Climate Change 6*, 445-451.

Puig, D., Krog Søbygaard, J., Larsen, P., Rygner Holm, S., Blatt Bendtsen, U., & Prag, A. (2013). *National Greenhouse Gas Emissions Baseline Scenarios: Learning from Experiences in Developing Countries*. UNEP Risø Centre, Danish Energy Agency, and Organisation for Economic Co-operation and Development. Roskilde and Copenhagen, Denmark, and Paris, France.

Puig, D. (2015). *Uncertainty in greenhouse-gas emission scenario projections*. UNEP DTU Partnership.

Copenhagen, Denmark.
SEMARNAT (2012). *México: Quinta Comunicación Nacional ante la Convención Marco de las Naciones Unidas sobre el Cambio Climático*. Secretaría de Medio Ambiente y Recursos Naturales. Mexico DF, Mexico.
SEMARNAT (2013). *Estrategia Nacional de Cambio Climático: Visión 10-20-40*. Secretaría de Medio Ambiente y Recursos Naturales. Mexico DF, Mexico.
SENER (2012). *Prospectiva del sector eléctrico 2012-2026*. Secretaría de Energía. Mexico DF, Mexico.
UNEP (2016). *The emissions gap report 2016: A UNEP synthesis report.* United Nations Environment Programme. Nairobi, Kenya.
UNFCCC (2016). *Aggregate effect of the intended nationally determined contributions: an update. Synthesis report by the secretariat (FCCC/CP/2016/2)*. United Nations Framework Convention on Climate Change. Bonn, Germany.

SUPPLEMENTARY INFORMATION

SI.1 Expert judgement elicitation

We used expert judgement elicitation to quantify the uncertainty around estimates of future trends in key drivers of greenhouse-gas emissions (Morales-Nápoles 2014). The drivers analysed were (annual growth rates of) gross domestic product, as well as oil (West Texas intermediate) and gas (Henry Hub) prices. For this research two expert judgment elicitation sessions were conducted in Mexico, both in 2014. Two different groups of experienced Mexican macro-economists were engaged.

We relied on the so-called classical (or Cooke's) method for elicitation of structured expert judgment. A complete discussion of the classical method is out of the scope of this paper. Nonetheless, a summary of the method's main features is provided. For a complete overview the reader is referred to (Cooke 1991).

The classical method for structured expert judgement seeks to achieve rational consensus. Simply stated, experts are asked to assess their uncertainty concerning certain continuous quantities in the form of a number of percentiles (most commonly the 5th, 50th and 95th percentiles) for their uncertainty distribution. The percentiles are assessed for uncertain quantities (*variables of interest*) but also for quantities whose value is known to the analysts (or will be known within the time frame of the research), but is not known to the experts at the moment of the elicitation. These are called *seed* or *calibration variables* and are used to ensure empirical control of the experts' uncertainty assessments.

Examples of a seed variable and a variable of interest (concerning economic growth in Mexico) follow:

a) *Seed variable*: Quarterly growth rates of gross domestic product in Mexico have been below -5% in four instances between the first trimester of 1994 and the third trimester of 2013. What was the average value of the 28-day Mexican Federal Treasury Certificates (CETES) interest rate in these four trimesters? Indicate the 5th, 50th and 95th percentiles of your uncertainty distribution.

b) *Variable of interest*: Consider a scenario in which, at the end of 2020, the Mexican (commercial) interest rate is between 3.5 and 4.0 percent, the unemployment rate is between 5.4 and 5.6 percent, the inflation growth rate is between 3.0 and 3.3, and growth rates of gross domestic product in the USA are between 2.8 and 3.3 percent. Please provide your estimate of the mean gross domestic product growth rates in Mexico up to 2020.

Seed variables are used to compute two measures of performance: the *calibration* and *information* scores. Simply stated, the calibration score measures the degree to which experts are statistically accurate, while the information score measures the degree to which experts' uncertainty estimates are concentrated, relative to a background measure.

SI.1.a Calibration

Assume we have answers from $e = 1, \dots, k$ experts on $i = 1, \dots, N$ calibration variables. Assume further that we assess three quintiles: 5th, 50th and 95th for each uncertain quantity. For each quantity, each expert divides his/her belief range into four inter-quantile intervals, for which the corresponding probabilities of occurrence are: $p_1 = 0.05$ for a realization value $\leq$ 5th percentile, $p_2 = 0.45$ for a realization value $\in$ (5th, 50th] percentile, $p_3 = 0.45$ for a realization value $\in$ (50th, 95th] percentile, and $p_4 = 0.05$ for a realization value > 95th percentile. The empirical version of $\boldsymbol{p} = (p_1, \dots, p_4)$ for expert e, is denoted $\boldsymbol{s}(\boldsymbol{e}) = (s_1, \dots, s_4)$, where

$$s_1(e) = \frac{\#\text{ realization in seed variables } \leq 5^{\text{th}} \text{ percentiles assessed by expert } e}{N}$$

$$s_2(e) = \frac{\#\text{ realization in seed variables } \in (5^{\text{th}}, 50^{\text{th}}] \text{ percentiles assessed by expert } e}{N}$$

$$s_3(e) = \frac{\#\text{ realization in seed variables } \in (50^{\text{th}}, 95^{\text{th}}] \text{ percentiles assessed by expert } e}{N}$$

$$s_4(e) = \frac{\#\text{ realization in seed variables } > 95^{\text{th}} \text{ percentiles assessed by expert } e}{N}$$

One way to measure the difference between $\boldsymbol{p}$ and $\boldsymbol{s}(\boldsymbol{e})$ is through relative information or entropy, which is a measure of the average information content that one is missing when one does not know the value of the random variable (Shannon & Weaver 1959).

$$I(\boldsymbol{s}(e), \boldsymbol{p}) = \sum_{j=1}^{4} s_j \ln\left(\frac{s_j}{p_j}\right) \tag{1}$$

Experts' assessments are treated as statistical hypotheses. Consider for each expert the null hypothesis

H_0: the inter quintile interval containing the true value for each variable is drawn independently from the probability vector $\boldsymbol{p}$.

The quantity $2\,N\,I(\boldsymbol{s}(e),\boldsymbol{p})$ where $I(\boldsymbol{s}(e),\boldsymbol{p})$ is given in equation (1) is asymptotically χ_3^2 (the degrees of freedom are the number of inter quintile intervals minus 1). This quantity can be used to test H_0 and it defines the calibration score:

$$CS(e) = P\{2\,N\,I(\boldsymbol{s}(e),\boldsymbol{p}) \geq r\} \tag{2}$$

where r is the value for expert e computed as in equation (1). The probability can be evaluated by a χ_3^2 distribution. The calibration score $CS(e)$ is the probability that a deviation at least as large as r could be observed on N realizations if H_0 were true. Values of calibration close to zero mean that it is unlikely that the experts' probabilities are correct.

SI.1.b Information

Recall that the information score measures the degree to which a distribution is concentrated with respect to a background measure. In the classical model uniform or log uniform background measures are used. An intrinsic range is calculated for the background measure $[q_l, q_h]$, where $q_l = q_5 - k(q_{95} - q_5)$ and $q_l = q_5 + k(q_{95} - q_5)$. k is typically chosen as 0.10 (10% overshoot). q_5 is the lowest 5th percentile assessed across experts for a particular item and q_{95} the largest 95th percentile assessed across experts for the same item. The information score is then computed as

$$IS(e) = \frac{1}{N}\sum_{i=1}^{N} I(f_{e,i}, g_i) \tag{3}$$

where g_i is the background density for item i (usually uniform or log uniform) and $f_{e,i}$ the density for expert e on item i. $I(f_{e,i}, g_i)$ is the mutual entropy between the densities of interest.

SI.1.c Combination

In the classical model the combination of experts' assessments is called a *Decision Maker* (DM). This is a weighted average of individual estimates. When the weights are determined based on the performance of experts in the seed variables, we speak of performance-based DM. The DM distributions are thus:

$$DM_\alpha(i) = \frac{\sum_{e=1}^{k} w_\alpha(e)\, f_{e,i}}{\sum_{e=1}^{k} w_\alpha(e)} \tag{4}$$

where the weighs $w_\alpha(e) = 1_{\{CS(e)>\alpha\}} CS(e) IS(e)$. Values of $\alpha < 0.05$ would fail to confer the study the required level of confidence. Note that the DM can also be evaluated in terms of calibration and information. For this reason the DM is referred to as 'virtual expert'. In the performance based DM the value of α is chosen such that the calibration score of the DM is maximized.

SI.1.d Econometric analyses and scenarios

To estimate gross domestic product growth rates, probability distributions were elicited from nine experts through a three-day workshop. A total of 19 variables were assessed, of which 13 were calibration variables. The variables of interest all referred to estimates of gross domestic product growth rates in 2020 and 2030.

To estimate trends in oil and gas prices, probability distributions were elicited from eight experts. A total of 26 calibration variables and 24 variables of interest were assessed. These variables also referred to 2020 and 2030.

An econometric model was built for each variable of interest (that is, gross domestic product, and energy commodity prices) (Loría 2013). The results of the econometric models were used to develop several sets of scenarios. The scenario approach was chosen because of the strong interdependencies among variables. By capturing those interdependencies through scenarios (as opposed to eliciting experts on all individual variables potentially influencing gross domestic product, and energy commodity prices), it was possible to avoid a cumbersome post-hoc dependency analysis.

For gross domestic product, six scenarios were built. They targeted both 2014-2020 and 2021-2030, and were defined around three combinations of macro-economic conditions – 'pessimistic' (low economic growth),

‘neutral’ (medium economic growth) and ‘optimistic’ (high economic growth). The scenarios were structured around different plausible combinations of values for interest rates, unemployment, inflation and economic growth in the United States (the latter is a key determinant of growth in Mexico). The variable elicited was gross domestic product (GDP) growth rates (the 5th, 50th and 95th percentiles).
For energy commodity prices twelve scenarios were built – six for oil prices and six for gas prices. The scenarios targeted the periods 2014-2020 and 2021-2030, and were defined around three combinations of economic growth and international oil trade conditions. These were also labelled ‘pessimistic’, ‘neutral’ and ‘optimistic’. In this case the scenarios were structured around plausible combinations of values for economic growth in China, economic growth in India and oil imports in OECD countries. For the analysis of oil prices West Texas Intermediate was used as a reference, whereas the Henry Hub spot price was used for the analysis of natural gas prices. As for gross domestic product, experts were asked to provide the 5th, 50th and 95th percentiles.

SI.1.e Results of the expert judgment elicitation for gross domestic product growth rates

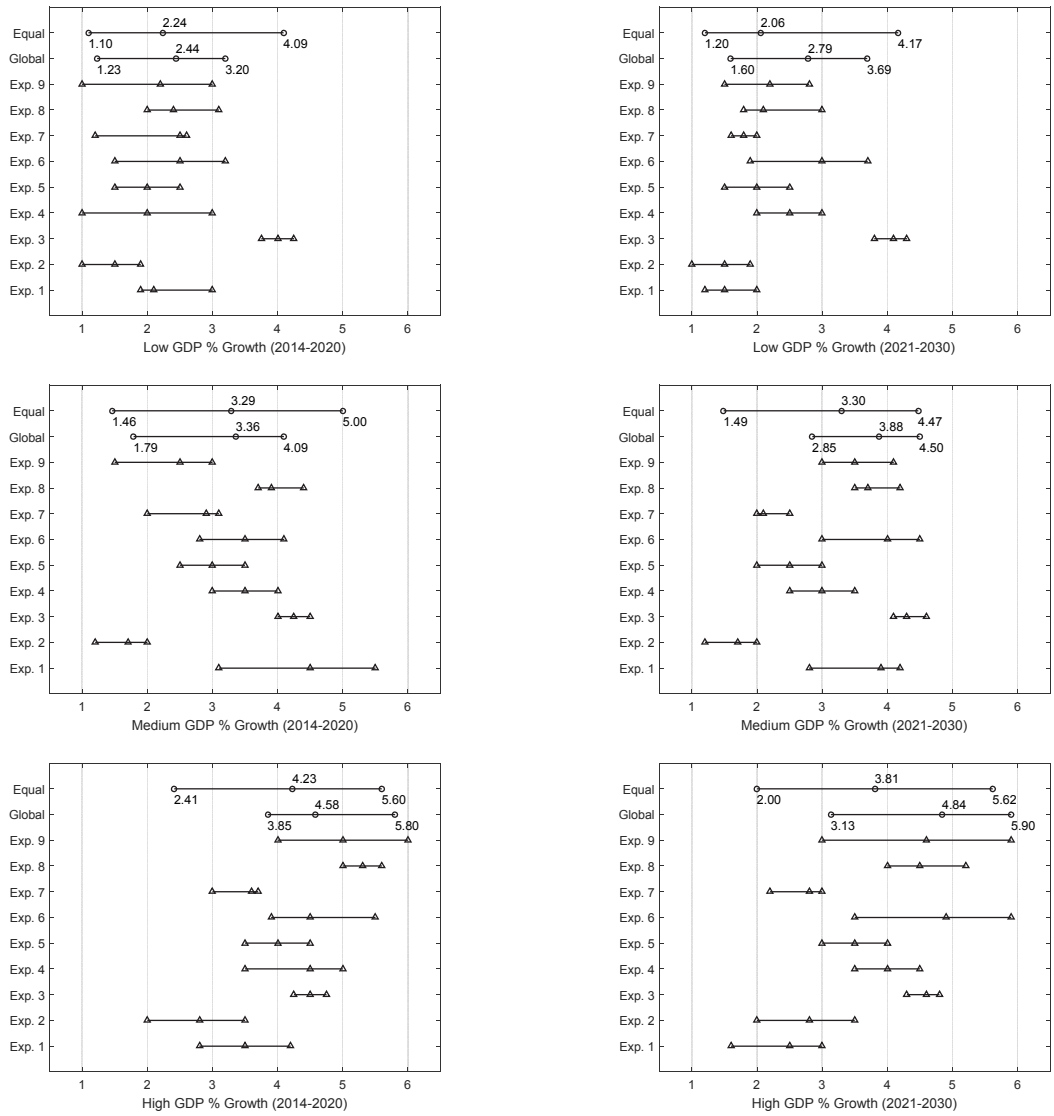
Supplementary Table 1 presents the results of the performance assessment of the 9 experts participating in a 2014 workshop for assessing the uncertainty associated with projections of economic growth in Mexico. The second column presents the calibration score (equation (2)) of each expert. Typically a value <0.05 would cast doubts regarding the credibility of the exercise. As shown in Supplementary Table 1, two individual experts (6 and 9) have calibration scores > 0.05. The calibration score of experts 3, 5 and 7 is marginal, while that of experts 1, 2, 4 and 8 is <0.0001. The “virtual experts”, that is, the combination of expert opinions according to equal weights (column 7 in Table S1) and global weights (column 8), both present a calibration score > 0.05. The equal weight combination shows a calibration score larger than all individual experts, except for expert 6, who is the individual expert with the highest calibration score. On the other hand, the global weight combination outperforms the calibration score of the highest calibrated individual expert by a factor of 2.2. It also outperforms the equal weight combination (by about an order of magnitude). Note that the equal weight combination has the lowest information score across all experts, including the performance-based combination. Columns 3 and 4 present the mean relative information (equation (3)) computed with all variables (column 3) and with calibration variables only (column 4). Across both columns individual experts have information scores that are in the interval (1.006, 1.944). The ratio of the information score per expert in calibration variables to all variables is within 28%, which shows that experts are consistent in expressing their uncertainty between seed variables and variables of interest. Column 5 presents the weights of individual experts, which correspond to the product of columns 2 and 4. Column 6 presents the ratio of the information score on calibration variables of individual experts to the equal-weight combination. For all experts, including the performance-based combination, this ratio is > 2. Based on the results presented in Supplementary Table 1, the performance-based combination is recommended as the best representation of the uncertainty regarding growth rates of gross domestic product in Mexico in 2020 and 2030. The results of expert opinions are shown in Supplementary Figure 1.

Supplementary Table S1: Results of the performance assessment for nine experts participating in a 2014 workshop to assess the uncertainty associated with projections of economic growth in Mexico

Expert ID	Calibration	Mean relative information		Weight	Information relative to Equal	Normalized weight	
		all variables	calibration variables			Equal	Global
Exp. 1	<0.0001	1.608	1.935	<0.0001	4.45	0.111	
Exp. 2	<0.0001	1.03	1.006	<0.0001	2.31	0.111	
Exp. 3	0.0001	1.666	1.651	0.0002	3.80	0.111	
Exp. 4	<0.0001	1.302	1.484	<0.0001	3.41	0.111	
Exp. 5	0.0015	1.223	1.301	0.0020	2.99	0.111	0.004601
Exp. 6	0.2766	1.081	1.285	0.3554	2.96	0.111	0.8353
Exp. 7	0.0002	1.825	1.944	0.0003	4.47	0.111	
Exp. 8	<0.0001	1.698	1.913	<0.0001	4.40	0.111	
Exp. 9	0.0530	1.078	1.285	0.0682	2.96	0.111	0.1601
Equal	0.0632	0.3378	0.4348	0.0275	1.00		
Global	0.6140	0.7797	0.924	0.5673	2.13		

Supplementary Figure 1 shows that the equal-weight combination presents (as expected) largest uncertainties across experts. However, the performance-based combination presents uncertainty levels comparable to those of the individual experts expressing larger uncertainties.

Supplementary Figure 1: The 5th, 50th and 95th percentiles of expert opinions regarding economic growth in Mexico in 2020 and 2030



Sources: own elaboration, based on the results of the expert judgement elicitation

SI.1.f Results of the expert judgment elicitation for oil and gas prices

Supplementary Table 2 presents the results of the performance assessment of the 8 experts participating in the 2014 workshop for uncertainty analysis of energy commodity prices. This table is equivalent to Supplementary Table 1 above.

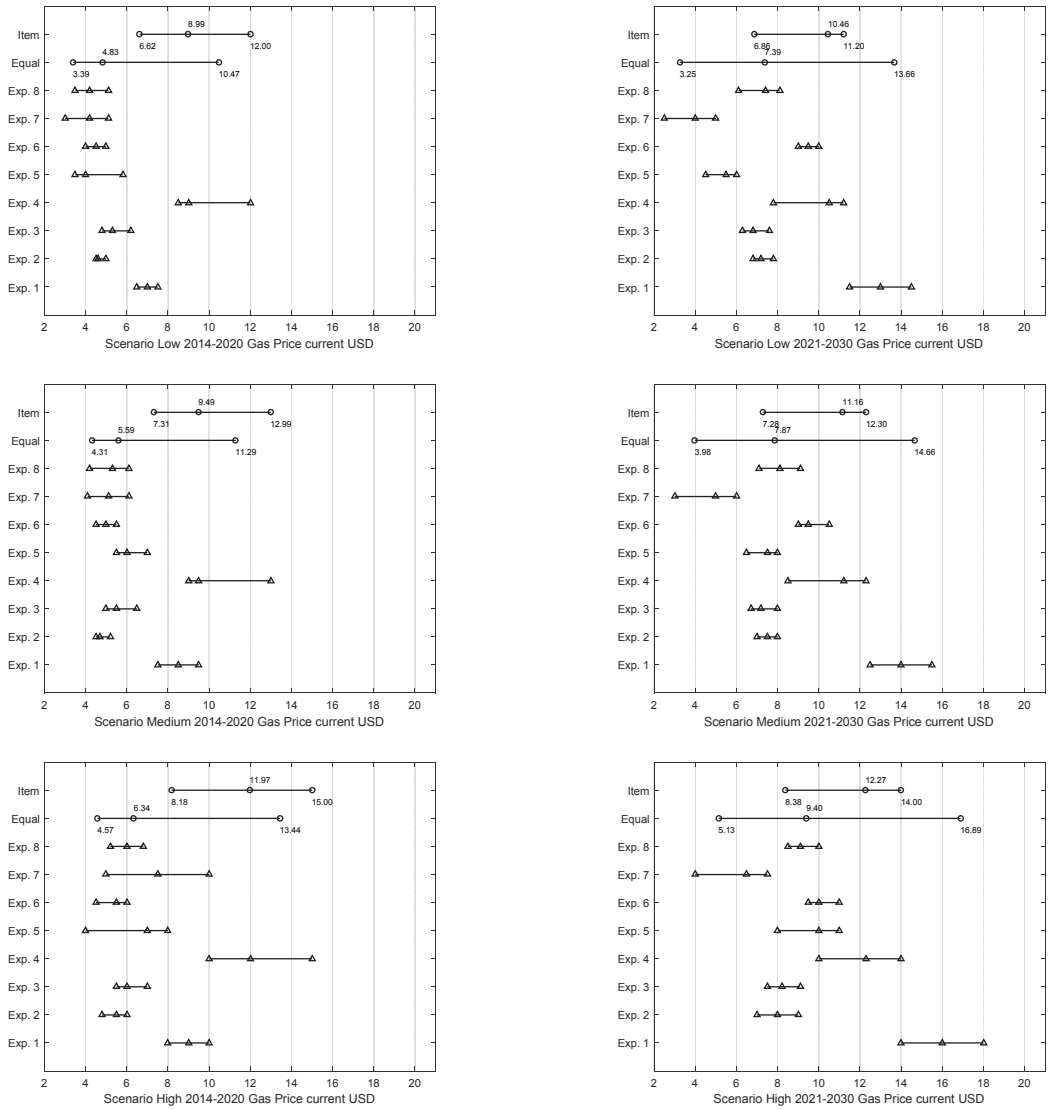
Column 2 shows the calibration scores per expert and decision maker. In this case only expert 4 has a calibration score > 0.05. While the equal-weight decision-maker is highly calibrated, the information scores of individual experts can be up to a factor of 4.5 larger than the one observed for this DM (column 7). The calibration score of the item weight combination is still > 0.05 and larger than that of each individual expert. The information score for the performance-based combination on all items is slightly smaller than that of individual experts (column 3). For calibration items alone, the information score of the performance based DM is bigger than that of expert 4. The performance-based combination is mostly that of expert 4 (98%) and experts 2 and 7 contributing the remaining 2%. In this case there is no evidence not to recommend the performance-based decision maker as the preferred choice. Results for variables of interest are presented in Supplementary Figure S2.

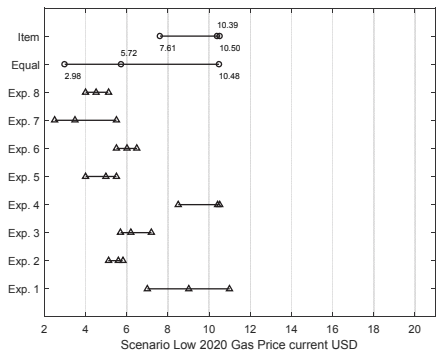
Supplementary Table 2: Results of the performance assessment for 8 experts participating in a

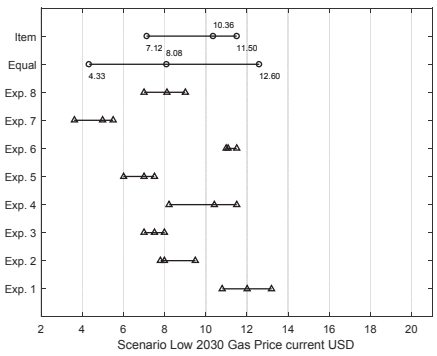
2014 workshop to assess the uncertainty associated with projections of international oil and gas prices

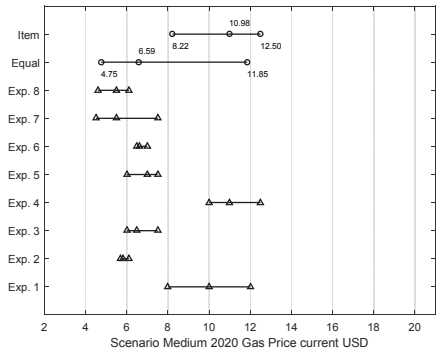
Expert ID	Calibration	Mean relative information		Weight	Information relative to Equal	Normalized weight	
		all variables	calibration variables			Equal	Global
Exp. 1	<0.0001	1.347	1.235	<0.0001	2.84	0.125	<0.0001
Exp. 2	0.0003	1.794	1.902	0.0005	4.37	0.125	0.0066
Exp. 3	<0.0001	1.603	1.576	<0.0001	3.62	0.125	<0.0001
Exp. 4	0.0721	1.045	1.004	0.0723	2.31	0.125	0.9852
Exp. 5	<0.0001	1.428	1.167	<0.0001	2.68	0.125	<0.0001
Exp. 6	<0.0001	2.016	1.955	<0.0001	4.50	0.125	<0.0001
Exp. 7	0.0005	1.075	1.262	0.0006	2.90	0.125	0.0082
Exp. 8	<0.0001	1.414	1.474	<0.0001	3.39	0.125	<0.0001
Equal	0.8256	0.2123	0.2566	0.2118	0.59		
Item	0.1893	0.9657	1.039	0.1967	2.39		

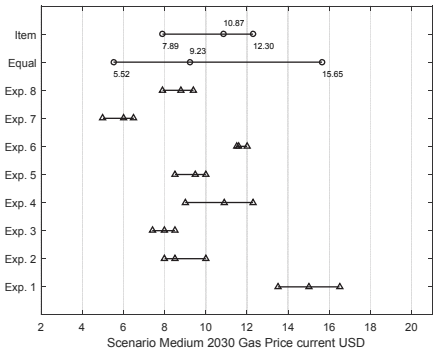
Supplementary Figure 2: **The 5th, 50th and 95th percentiles of expert opinions regarding oil and gas prices in Mexico in 2020 and 2030**

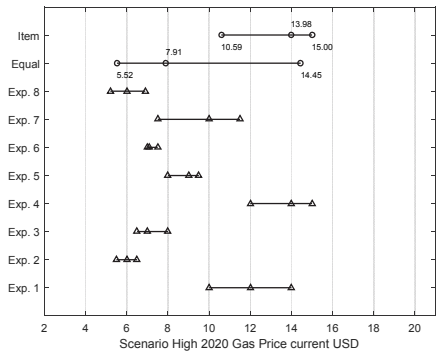


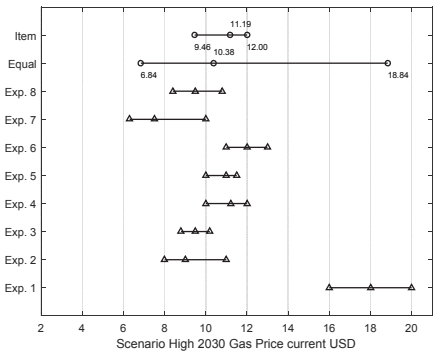

Item
Equal
Exp. 8
Exp. 7
Exp. 6
Exp. 5
Exp. 4
Exp. 3
Exp. 2
Exp. 1
10.39
7.61
10.50
5.72
2.98
10.48
2 4 6 8 10 12 14 16 18 20
Scenario Low 2020 Gas Price current USD

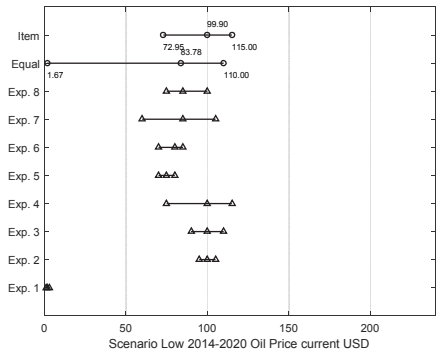

Item
Equal
Exp. 8
Exp. 7
Exp. 6
Exp. 5
Exp. 4
Exp. 3
Exp. 2
Exp. 1
10.36
7.12
11.50
8.08
4.33
12.60
2 4 6 8 10 12 14 16 18 20
Scenario Low 2030 Gas Price current USD

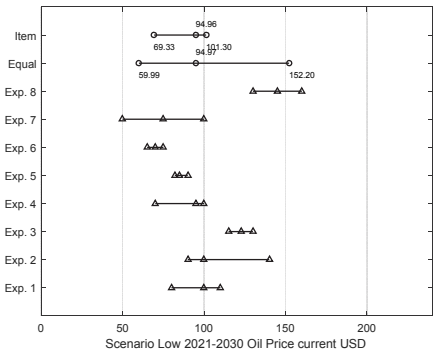

Item
Equal
Exp. 8
Exp. 7
Exp. 6
Exp. 5
Exp. 4
Exp. 3
Exp. 2
Exp. 1
10.98
8.22
12.50
6.59
4.75
11.85
2 4 6 8 10 12 14 16 18 20
Scenario Medium 2020 Gas Price current USD

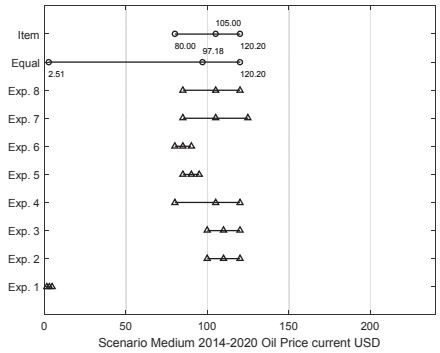

Item
Equal
Exp. 8
Exp. 7
Exp. 6
Exp. 5
Exp. 4
Exp. 3
Exp. 2
Exp. 1
10.87
7.89
12.30
9.23
5.52
15.65
2 4 6 8 10 12 14 16 18 20
Scenario Medium 2030 Gas Price current USD

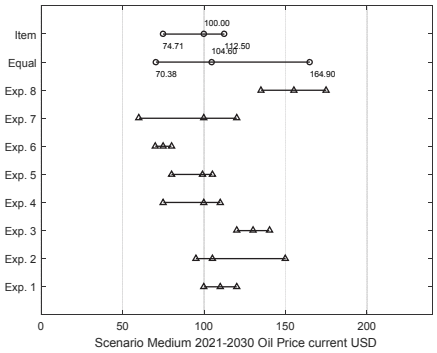

Item
Equal
Exp. 8
Exp. 7
Exp. 6
Exp. 5
Exp. 4
Exp. 3
Exp. 2
Exp. 1
13.98
10.59
15.00
7.91
5.52
14.45
2 4 6 8 10 12 14 16 18 20
Scenario High 2020 Gas Price current USD

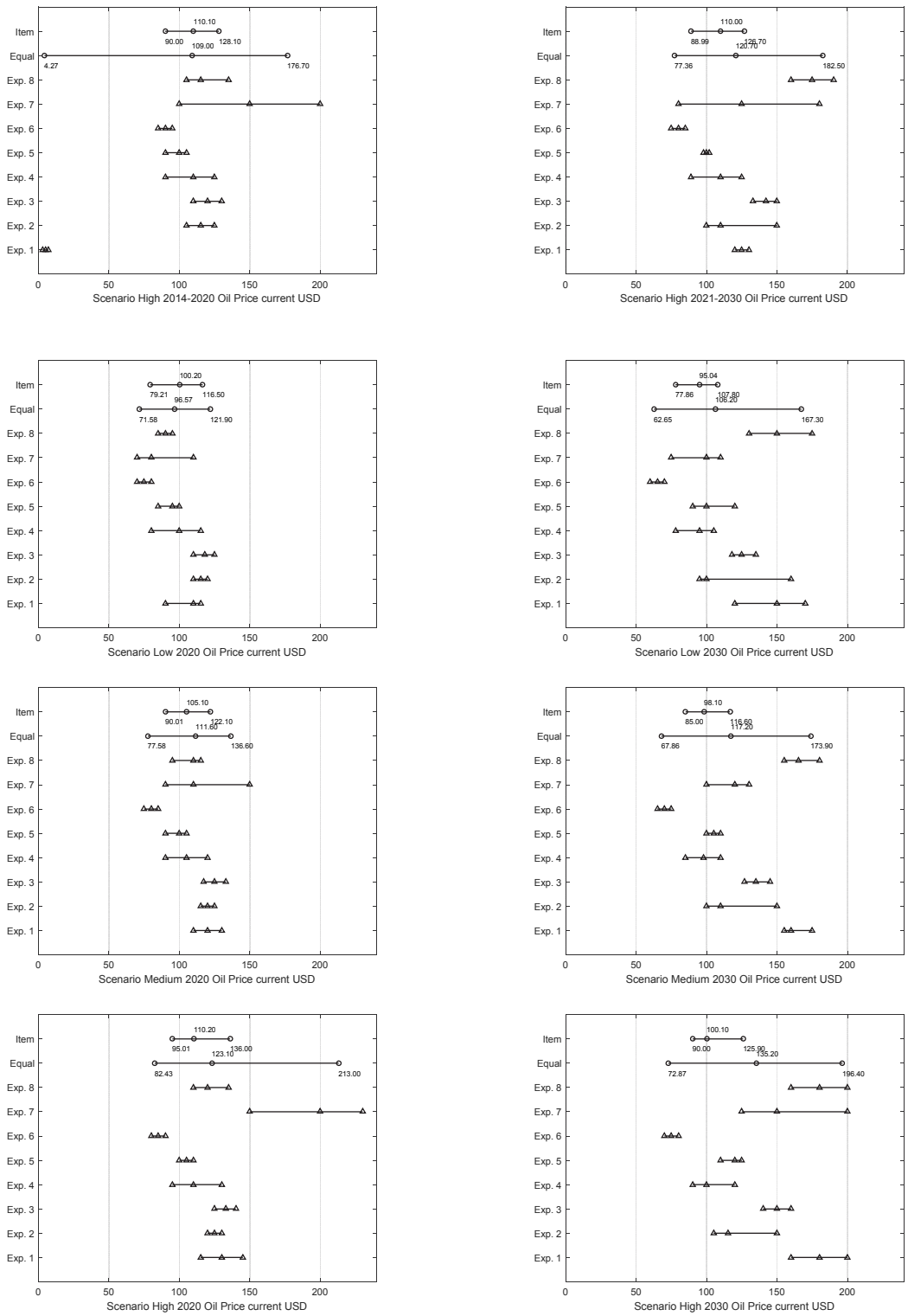

Item
Equal
Exp. 8
Exp. 7
Exp. 6
Exp. 5
Exp. 4
Exp. 3
Exp. 2
Exp. 1
11.19
9.46
12.00
10.38
6.84
18.84
2 4 6 8 10 12 14 16 18 20
Scenario High 2030 Gas Price current USD


Item
Equal
Exp. 8
Exp. 7
Exp. 6
Exp. 5
Exp. 4
Exp. 3
Exp. 2
Exp. 1
99.90
72.95
115.00
83.78
1.67
110.00
0 50 100 150 200
Scenario Low 2014-2020 Oil Price current USD


Item
Equal
Exp. 8
Exp. 7
Exp. 6
Exp. 5
Exp. 4
Exp. 3
Exp. 2
Exp. 1
94.96
69.33
101.30
94.97
59.99
152.20
0 50 100 150 200
Scenario Low 2021-2030 Oil Price current USD


Item
Equal
Exp. 8
Exp. 7
Exp. 6
Exp. 5
Exp. 4
Exp. 3
Exp. 2
Exp. 1
105.00
80.00
120.20
97.18
2.51
120.20
0 50 100 150 200
Scenario Medium 2014-2020 Oil Price current USD


Item
Equal
Exp. 8
Exp. 7
Exp. 6
Exp. 5
Exp. 4
Exp. 3
Exp. 2
Exp. 1
100.00
74.71
112.50
104.60
70.38
164.90
0 50 100 150 200
Scenario Medium 2021-2030 Oil Price current USD



Sources: own elaboration, based on the results of the expert judgment elicitation.

SI.1.g List of participating experts

Alejandro Villagomez Amezcua (PhD, University of Washington in St. Louis). Dr. Villagomez specialises in savings, pensions and Mexico's monetary policy. He works in Mexico's Centro de Investigación y Docencia Económicas.

Alberto Moritz Cruz Blanco (PhD, University of Manchester). Dr. Moritz specialises in prospective macro-econometric analysis. He works in Mexico's Universidad Nacional Autónoma de México.

Eliseo Díaz González (PhD, Universidad Nacional Autónoma de México). Dr. Díaz specialises in remittances, labour market, fiscal fraud, foreign direct investment flowing to Mexico and economic growth in Mexico. He works in Mexico's Colegio de la Frontera Norte.

Ignacio Perrotini Hernández (PhD, New School for Social Research). Dr. Perrotini specialises in monetary policy, inflation and economic growth in Mexico. He works in Mexico's Universidad Nacional Autónoma de México.
Pablo Ruiz Nápoles (PhD, New School for Social Research). Dr. Ruiz specialises in input-outputs analysis applied to economic growth in Mexico. He works in Mexico's Universidad Nacional Autónoma de México.
Juan Carlos Rivas Valdivia (MSc, Colegio de México). Mr. Rivas specialises in econometric analysis applied to economic growth in Mexico and Central America. He works in BBVA Bancomer, an international bank.
Junior Alfredo Martínez Hernández (MSc, Universidad de Guadalajara). Mr. Martínez specialises in econometric analysis applied to economic growth in Mexico. He works in Mexico's Chamber of Deputies.
Ramón Padilla Perez (PhD, University of Sussex). Dr. Padilla specialises in analyses of exports and competitiveness in Mexico's manufacturing sector. He works in Mexico's delegation of the United Nations Economic Commission for Latin America and the Caribbean.
José Manuel Iraheta Bonilla (MSc, Instituto Tecnológico Autónomo de México). Mr. Iraheta specialises in econometric modelling, inflation and monetary policy. He works in Mexico's delegation of the United Nations Economic Commission for Latin America and the Caribbean.

SI.2 General-equilibrium modelling

Three-ME (Multi-sector Macroeconomic Model for the Evaluation of Environmental and Energy policy) is a model especially designed to evaluate the medium and long term impact of environmental and energy policies at the macroeconomic and sector levels. To do so Three-ME combines two important features. Firstly, it has the main characteristics of neo-Keynesian models by assuming a slow adjustment of effective quantities and prices to their notional level, an endogenous money supply, a Taylor rule and a Philips curve. Compared to standard multi-sector Computable General Equilibrium Model (CGEM), this has the advantage to allow for the existence of under-optimum equilibria such as the presence of involuntary unemployment. Secondly, in an extended version, Three-ME is a hybrid model in the sense that it combines the top-down approach of general equilibrium macroeconomic models with elements of bottom-up models of energy models developed by engineers. As in bottom-up models, the amount of energy consumed is related to their use, that is the number of buildings or cars, and the energy class to which they belong. This hypothesis is more realistic compared to the assumption made in the majority of top-down models where energy consumption is usually directly related to income through a nested structure of utility function (Callonec *et al.* 2013).

Three-ME is a country-generic and open source model developed since 2008 by the ADEME (French Environment and Energy Management Agency), the OFCE (French Economic Observatory) and TNO (Netherlands Organisation for Applied Scientific Research). Initially developed to support the energy/environment/climate debate in France, Three-Me has then been applied to other regional contexts.

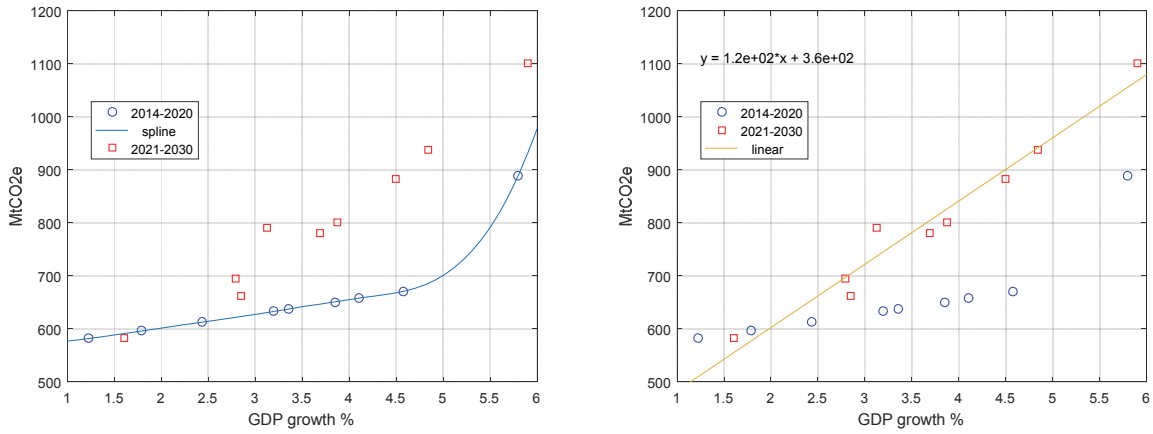
In the period between 2013 and 2015 a team of French and Mexican researchers adapted the ThreeME model for use by government agencies in Mexico. The main objective of the modelling project was to evaluate the impact of climate and energy policies. Through this work, the model was used to produce projections of greenhouse-gas emissions for Mexico for the years 2020, 2030 and 2050. To do this, the model relied on the estimates of growth rates in gross domestic product obtained through the expert judgement elicitation outlined above.

In the ThreeME model gross domestic product is not an exogenous variable that is entered as input to the model runs. Instead, gross domestic product is calculated by the model. For this reason, to produce estimates of greenhouse-gas emissions that are consistent with any one estimate of growth rates in gross domestic product, the model had to be recalibrated each time.

SI.3 Probability distributions of greenhouse-gas emissions for Mexico in 2020 and 2030

We take the estimates from Table 2 and plot the reference scenario forecasts in $MtCO_2e$ as a function of economic growth. For 2014-2020, the relationship is linear up to approximately 4.5 in gross domestic product growth rates, and non-linear thereafter. For this period a spline interpolant would approximate perfectly the relationship between gross domestic product growth rates and emissions of greenhouse gases. In fact, the rank correlation for these nine observations equals 1. This implies a perfect functional relationship between gross domestic product growth rates and emissions of greenhouse gases, when calculated as described in SI.2 for the period of time under consideration. For the period between 2021 and 2030, the relationship is approximately linear throughout the entire period.

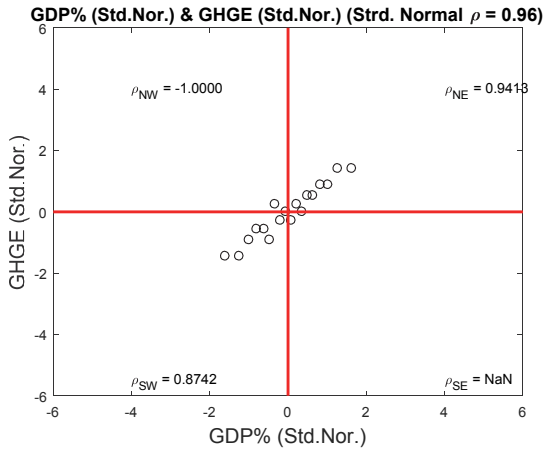
Supplementary Figure 3: Reference scenario forecasts in $MtCO_2e$ as a function of economic growth for Mexico in 2020 and 2030



Sources: own elaboration, based on Table 2.

If the 18 estimates from Table 2 are considered together, and transformed to a standard normal distribution, a product moment correlation of ≈ 0.96 would be observed for the standard normal variates. This is presented in Supplementary Figure 4. Different statistical hypothesis tests (Genest *et al*. 2009, Joe 2014) would indicate that a Gaussian copula is a good approximation for the data.

Supplementary Figure 4: Standard normal transform of data in Supplementary Figure 3



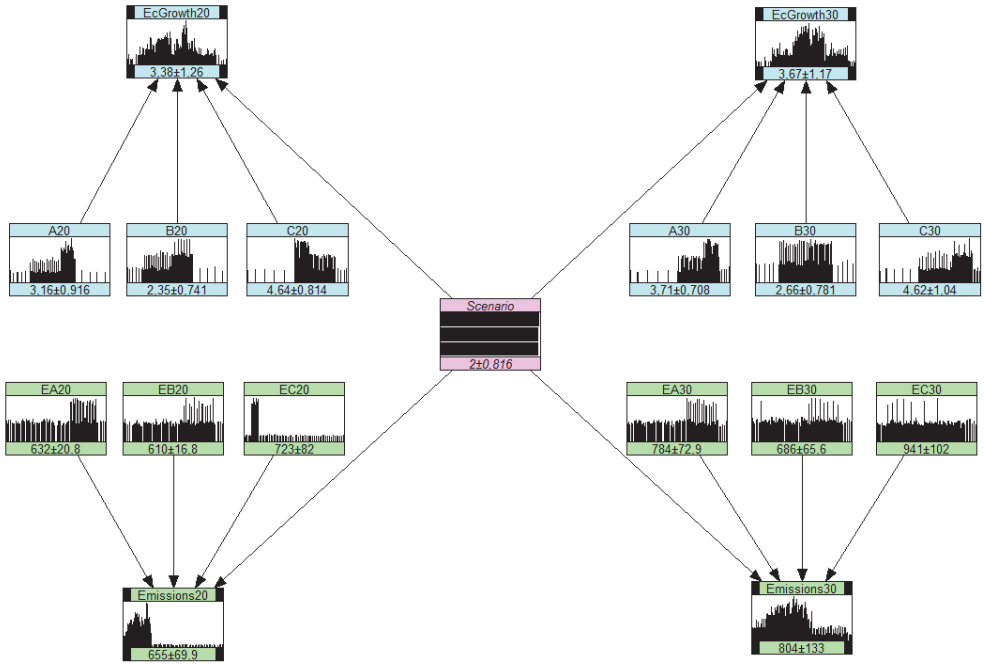
Sources: Own elaboration, based on Table 2.

We assume that the emissions given in Table 2, computed through the Three-ME model as described in SI.2, are indeed the 5th, 50th and 95th percentiles of the uncertainty distribution for each scenario. We determine the probability distribution of greenhouse-gas emissions as a mixture of the individual distributions, where the mixture coefficients are given equal weight in the mixed distribution. Stated differently, we assume equal probability for the three scenarios within each of the periods: until 2020 and until 2030, both for gross domestic product growth rates, and greenhouse-gas emissions. The graphical representation (as a Bayesian Network: see Hanea *et al.* 2015) is shown in Supplementary Figure 5.

The node scenario is a variable representing the equal probability for the three scenarios. Nodes A20, B20 and C20 represent the distribution of gross domestic product growth rates for the period 2014-2020, elicited through expert judgment, and corresponding to the neutral, pessimistic and optimistic scenarios, respectively (Supplementary Information SI.1). Nodes A30, B30 and C30 are the analogous distributions for the period 2021-2030.

Nodes EA20, EB20 and EC20 represent the distributions of emissions for 2020 obtained through the ThreeME model for the different percentiles of gross domestic product growth rates (Table 2 and SI.2.) also assuming neutral, pessimistic and optimistic scenarios. EA30, EB30 and EC30 are analogous to Nodes EA20, EB20 and EC20, but for 2030.

Supplementary Figure 5: Mixture distributions for Mexican gross domestic product growth rates and greenhouse-gas emissions for 2020 and 2030.

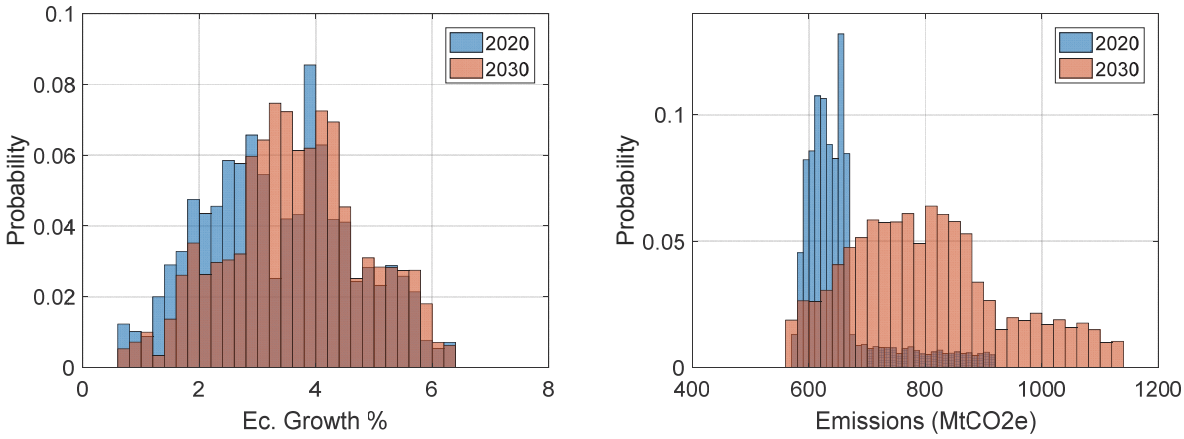


Sources: Own elaboration, based on estimates described in SI.1 and SI.2. The Bayesian Network is presented in UniNet (Hanea *et al.*, 2015)

For comparison, Supplementary Figure 6 shows in the same image the following nodes (from Supplementary Figure 5): EcGrowth20, EcGrowth30, Emissions20, and Emissions30. Figure 6 is constructed such that the height of each bar is equal to the probability of selecting an observation within that bin interval, and the height of all of the bars sums to 1. Notice the larger discrepancies in terms of the distributions of greenhouse-gas emissions, rather than in terms of the distributions of gross domestic product growth rates.

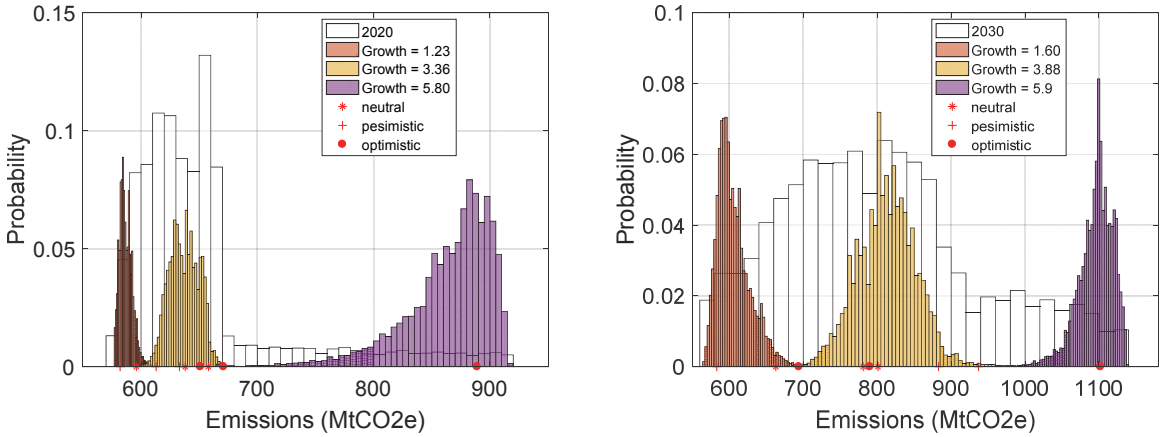
Finally, Supplementary Figure 7 presents the original distributions of greenhouse-gas emissions for Mexico in 2020 and 2030 alongside the conditional distributions given for gross domestic product growth rates = 1.23, 3.36, 5.8 for 2020 and given gross domestic product growth rates = 1.6, 3.88, 5.9 for 2030. The original neutral, pessimistic and optimistic estimates from Table 2 are also presented. All histograms in Figure 7 are constructed similarly as Figure 6 – that is, such that the height of each bar is equal to the probability of selecting an observation within that bin interval, and the height of all of the bars sums to 1,

Supplementary Figure 6: **Comparison of probability distributions of GDP growth and GHGE for Mexico in 2020 and 2030.**



Sources: Own elaboration.

Supplementary Figure 7: **Comparison of probability distributions of GHGE for Mexico in 2020 and 2030.**



Sources: Own elaboration.

Other approaches to estimate the probability distributions could have been used. For example, Oppenheimer and colleagues (2016) propose probabilistic inversion as a suitable alternative. We anticipate that probabilistic inversion would have yielded similar results. The method proposed here, the results of which are summarised in Supplementary Figure 7, is technically more accessible.

SI.References

Callonec, G., Landa, G., Malliet, P., Reynès, F., Yeddir-Tamsamani, Y. (2013). *A full description of the Three-ME model: multi-sector macroeconomic model for the evaluation of environmental and energy policy*. French Economic Observatory. Paris, France.

Cooke, R.M. (1991). *Experts in Uncertainty: Opinion and Subjective Probability in Science*. Oxford University Press. Oxford, United Kingdom.

Genest, C., Rémillard, B., & Beaudoin, D. (2009). Goodness-of-fit tests for copulas: A review and a power study. *Insurance: Mathematics and economics, 44*(2), 199-213.

Hanea, A., Napoles, O. M., & Ababei, D. (2015). Non-parametric Bayesian networks: Improving theory and reviewing applications. *Reliability Engineering & System Safety, 144*, 265-284.

Joe, H. (2014). *Dependence Modeling with Copulas*. Chapman & Hall/CRC. London, United Kingdom.

Loría, E. (2013). *Determinantes económicos del PIB total y sectorial de la economía mexicana, 1980-2012*. National Institute of Ecology and Climate Change. Mexico D.F., Mexico.

Morales-Nápoles, O. (2014). *Estimaciones de incertidumbre de largo plazo sobre variables económicas a través de expertos*. Reporte de resultados. TNO. Delft, Netherlands.

Shannon, C.E., & Weaver, W. (1959). *The mathematical theory of communication*. University of Illinois Press. Champaign, United States of America.

ABOUT OFCE

The Paris-based Observatoire français des conjonctures économiques (OFCE), or French Economic Observatory is an independent and publicly-funded centre whose activities focus on economic research, forecasting and the evaluation of public policy.

Its 1981 founding charter established it as part of the French Fondation nationale des sciences politiques (Sciences Po), and gave it the mission is to "ensure that the fruits of scientific rigour and academic independence serve the public debate about the economy". The OFCE fulfils this mission by conducting theoretical and empirical studies, taking part in international scientific networks, and assuring a regular presence in the media through close cooperation with the French and European public authorities. The work of the OFCE covers most fields of economic analysis, from macroeconomics, growth, social welfare programmes, taxation and employment policy to sustainable development, competition, innovation and regulatory affairs..

ABOUT SCIENCES PO

Sciences Po is an institution of higher education and research in the humanities and social sciences. Its work in law, economics, history, political science and sociology is pursued through ten research units and several crosscutting programmes.
Its research community includes over two hundred twenty members and three hundred fifty PhD candidates. Recognized internationally, their work covers a wide range of topics including education, democracies, urban development, globalization and public health.
One of Sciences Po's key objectives is to make a significant contribution to methodological, epistemological and theoretical advances in the humanities and social sciences. Sciences Po's mission is also to share the results of its research with the international research community, students, and more broadly, society as a whole.

PARTNERSHIP