How certain can we be about NGFS Climate Scenarios?

Working Paper

Authors: Alex Pui, Sebastian EV Werner, Gireesh Shrimali, Motoshi Tomita

Paper is also available on SSRN

July 2026

Abstract

This paper examines uncertainty in NGFS climate scenario projections across three dimensions—scenario, model, and vintage uncertainty—while further decomposing vintage uncertainty into two components: revisions to initial conditions and recalibration across successive scenario releases. Using 254 NGFS variables across four vintages, five climate pathways, and three integrated assessment models, we construct comparable uncertainty metrics for both near-term (2020-2025) and long-term (2020-2050) horizons.

The results show that revisions to initial conditions and cross-vintage recalibration are first-order features of NGFS outputs. At the near-term horizon most relevant for financial risk analysis, uncertainty arising from initial conditions exceeds scenario uncertainty for nearly 80% of variables and exceeds model uncertainty for over 40%. Recalibration across vintages frequently rivals or exceeds dispersion across scenarios and models even at longer horizons, implying that successive NGFS releases can alter projected pathways as much as changes in policy narratives or model structure.

These findings indicate that NGFS scenario outputs should be interpreted as evolving projections rather than fixed reference trajectories. For climate stress testing and financial supervision, the results highlight the importance of explicitly accounting for initial-condition and recalibration uncertainty, particularly over short horizons where financial decisions are most sensitive.

Acknowledgement

The authors are grateful to Chris Cormack, Daniel Dao, and working group participants at the Oxford Sustainable Finance Group Lab for helpful discussions and comments. We also thank participants at the European Geosciences Union (EGU) General Assembly 2024 for their valuable questions and feedback following the presentation of the conceptual framework underlying this study.

How certain can we be about NGFS climate scenarios?*

Alex Pui[1,2,3] Sebastian E V Werner[4,5] Gireesh Shrimali[6]
Motoshi Tomita[7]

[1]Marsh, Japan
[2]Climate Change Research Center, University of New South Wales
[3]Kyushu University Institute for Asian and Oceanic Studies (Q-AOS)
[4]Citigroup Inc.†
[5]University of Bristol
[6]Oxford Sustainable Finance Group, University of Oxford‡
[7]CRIEPI

9 July 2026

Abstract

This paper examines uncertainty in NGFS climate scenario projections across three dimensions—scenario, model, and vintage uncertainty—while further decomposing vintage uncertainty into two components: revisions to initial conditions and recalibration across successive scenario releases. Using 254 NGFS variables across four vintages, five climate pathways, and three integrated assessment models, we construct comparable uncertainty metrics for both near-term (2020–2025) and long-term (2020–2050) horizons.

The results show that revisions to initial conditions and cross-vintage recalibration are first-order features of NGFS outputs. At the near-term horizon most relevant for financial risk analysis, uncertainty arising from initial conditions exceeds scenario uncertainty for nearly 80% of variables and exceeds model uncertainty for over 40%. Recalibration across vintages frequently rivals or exceeds dispersion across scenarios and models even at longer horizons, implying that successive NGFS releases can alter projected pathways as much as changes in policy narratives or model structure.

These findings indicate that NGFS scenario outputs should be interpreted as evolving projections rather than fixed reference trajectories. For climate stress testing and financial supervision, the results highlight the importance of explicitly accounting for initial-condition and recalibration uncertainty, particularly over short horizons where financial decisions are most sensitive.

Keywords: NGFS scenarios; climate risk; integrated assessment models; uncertainty; stress testing

**Acknowledgements:* The authors are grateful to Chris Cormack, Daniel Dao, and working group participants at the Oxford Sustainable Finance Group Lab for helpful discussions and comments. We also thank participants at the European Geosciences Union (EGU) General Assembly 2024 for their valuable questions and feedback following the presentation of the conceptual framework underlying this study.

†*Disclaimer:* The views expressed in this document are solely those of the authors and do not reflect the views of Citigroup Inc. or any of its subsidiaries.

‡*Disclaimer:* The views expressed in this document represent those of the authors and do not necessarily represent those of the Oxford Sustainable Finance Group, or other institutions or funders.

1 Introduction

1.1 Background

At a time when climate risk assessment is becoming mainstream—extending well beyond the confines of academia—and arguably under the null hypothesis that climate change may pose a systemic risk to the global financial system, climate scenarios have been developed to provide a common starting point for analyzing both physical and transition risks to the economy. Led by a consortium of central banks, policymakers, and academic institutions, the Network for Greening the Financial System (NGFS) scenarios were designed to simulate complex dynamics across energy, economic, and climate systems. These scenarios provide climate pathways alongside a wide array of macroeconomic and sectoral outputs at multiple time horizons and increasing levels of geographical and sectoral granularity.

Although NGFS scenarios were initially developed to support prudential climate risk analysis by central banks and supervisors ([ACPR, 2020]; [Bank of England, 2021]; [European Central Bank , 2022]; [Hong Kong Monetary Authority , 2023]), they have since become a common reference framework for climate scenario analysis across supervisory exercises, financial institutions, and corporate climate-risk assessments ([Acharya et al., 2023]; [Baudino and Svoronos, 2021]). Importantly, these scenarios are generally understood by practitioners as exploratory and illustrative rather than predictive forecasts, with outputs depending on underlying policy assumptions, carbon-price paths, and model-specific optimisation frameworks. Their widespread use therefore reflects their value as a common benchmark for comparative analysis, rather than an assumption of point precision or deterministic interpretation. This is also consistent with climate disclosure frameworks such as the [Task Force on Climate-related Financial Disclosures, 2020] and [ISSB , 2023], which encourage scenario analysis using externally developed, publicly available climate scenarios. Their rapid adoption by a broad set of users—including government departments beyond prudential authorities, non-governmental organizations, academics, and corporations—has been so extensive that the NGFS itself has raised concerns that some applications may exceed the technical precision of the scenarios, given their simplifying assumptions and limited regional detail. In response, the NGFS has issued guidance on appropriate and inappropriate uses of the scenarios and emphasized the importance of complementary analysis, particularly when scenarios are applied at the level of individual institutions ([NGFS , 2023]).

NGFS scenarios provide a common analytical framework for exploring how climate policy pathways and transition narratives may affect economic and financial variables relevant to climate risk analysis. Their widespread use reflects both their policy relevance and the practical value of a shared reference framework for comparative scenario analysis. To this end, NGFS scenarios seek to provide a coherent framework for assessing climate-related risks and opportunities within the financial sector ([Allen et al., 2020]). However, the inherent limitations and uncertainties embedded in these scenarios—especially those related to data, models, and methodological choices—necessitate careful interpretation when scenarios are used for applied financial analysis ([Baer et al., 2023]). The focus of this paper is therefore not on evaluating the predictive accuracy of NGFS scenarios, but on examining their internal evolution across successive releases and the implications this has for interpretation over time.

NGFS scenarios are derived from process-based Integrated Assessment Models (IAMs), which play a central role at the climate science–policy interface ([van Beek et al., 2020]). IAMs are crucial for constructing and evaluating long-term mitigation pathways and informing policy debates, including contributions to Intergovernmental Panel on Climate Change (IPCC) assessments ([Keppo et al., 2021]). They translate emissions and land-use scenarios into inputs for Earth system models and generate projections of how climate policies affect energy systems and economic activity ([Tebaldi et al., 2022]; [Dekker et al., 2023]). Since 2020, the NGFS has published mitigation scenarios annually using three IAMs: MESSAGEix–GLOBIOM, REMIND–MAgPIE, and GCAM (NGFS, 2020–2024)[1].

1.2 Key results

The objective of this study is to examine the evolution of NGFS scenario outputs across four scenario vintages, five climate scenarios, and three integrated assessment models (IAMs), and to assess uncertainty across four empirically measurable sources. These include dispersion in reported initial conditions across vintages, variation across successive NGFS releases within the same model–scenario pair (vintage uncertainty), and the more commonly studied dimensions of scenario uncertainty and model uncertainty. We compare these sources using newly developed

[1]Table A1 in Appendix A provides an overview of the NGFS IAMs used in this study.

uncertainty ratios and characterize their distribution using empirical cumulative distribution functions (CDFs).

The analysis yields several key findings. First, dispersion in reported initial conditions is frequently as large as, and often larger than, uncertainty arising from modeled dynamics. Over the near-term horizon most relevant for financial risk analysis (2020–2025), initial-condition uncertainty exceeds recalibration uncertainty for around 40% of all variables, exceeds scenario uncertainty for nearly 80%, and exceeds model uncertainty for over 40%. Even over the full projection horizon to 2050, initial-condition uncertainty remains dominant for a non-trivial share of variables, exceeding recalibration, scenario, and model uncertainty in 14%, 23%, and 13% of cases, respectively. Importantly, the CDFs reveal that this dominance is often not marginal: for around 10% of variables over the 2020–2050 horizon, and close to 20% over the 2020–2025 horizon, initial-condition uncertainty exceeds scenario uncertainty by at least a factor of twenty.

Second, recalibration uncertainty across successive NGFS scenario vintages emerges as a first-order source of dispersion. Recalibration uncertainty exceeds scenario uncertainty for 85% of variables at the five-year horizon and for nearly half of variables by 2050, and exceeds model uncertainty for roughly half of variables in the near term and about one third over the full projection horizon. This implies that changes across annual NGFS releases within the same model can rival or exceed differences induced by alternative scenario narratives or by inter-model structure. Beyond dominance at parity, the cumulative distributions show that both initial-condition and recalibration uncertainty can exceed competing sources by large margins. In particular, for each of these dimensions, uncertainty is at least twenty times larger than scenario uncertainty for around 20% of variables at the five-year horizon and for roughly 10% of variables over the full projection horizon.

Taken together, these findings indicate that dispersion in NGFS outputs cannot be attributed primarily to either scenario or model differences alone. While differences across transition scenarios are often expected to dominate given their distinct policy pathways ([Guivarch et al., 2017, Rogelj et al., 2018]), and the IAM literature frequently emphasizes inter-model variation as a key source of uncertainty ([Kriegler et al., 2014, Krey et al., 2019, Riahi et al., 2021]), our results show that revisions across scenario vintages and differences in reported initial conditions constitute quantitatively important drivers of dispersion. The objective of the paper is not to provide a formal statistical identification of climate-transition uncertainty or to assess the validity of any individual NGFS pathway. Rather, the analysis documents how NGFS scenario outputs evolve across vintages and compares the magnitude of this variation with more commonly studied differences across scenarios and models. In this sense, the contribution is methodological and empirical: it provides a structured framework for analysing cross-vintage variation and assessing its implications for scenario interpretation, comparability, and practical use.

1.3 Literature review

These findings have direct implications for policy design, financial risk management, and investment decisions. Evidence from Carbon Disclosure Project surveys and related meta-analyses suggests that a key barrier to the effective uptake of climate scenario analysis is the difficulty practitioners face in interpreting global scenario outputs—including NGFS pathways— particularly with respect to underlying data choices and modelling assumptions [Task Force on Climate-related Financial Disclosures, 2020, IFRS Foundation, 2024a][2]. By explicitly identifying and decomposing uncertainty across initial conditions, recalibrations, scenarios, and models, this study provides a structured framework that facilitates interpretation of scenario outputs and supports consistent updates to scenario-based analyses over time.

Beyond its practical relevance, this work contributes to the literature on the evaluation and coherence of IAMs. A related body of research evaluates IAMs by comparing simulated pathways with historical energy, emissions, and economic dynamics [Schwanitz, 2013, Wilson et al., 2021]. This literature emphasizes that long-horizon policy scenarios cannot be validated in a conventional forecasting sense, given deep structural uncertainty and the counterfactual nature of policy pathways. Instead, evaluation focuses on internal consistency, structural interpretability, and the ability of models to reproduce key empirical patterns.

Extending this perspective, we propose that coherence can also be assessed intertemporally. When the same model–scenario pair is updated across successive releases, projections should exhibit a coherent internal narrative across key variables, conditional on changes in data, assumptions, or model structure. From this viewpoint, cross-vintage coherence represents a complementary dimension of IAM evaluation that, to our knowledge, has not been systematically

[2]International Financial Reporting Standards

documented or formally assessed in the published NGFS literature or accompanying model documentation. This makes it particularly relevant for policy, supervisory, and financial risk applications where scenario comparability over time matters.

Related questions arise in the literature on revisions, updates, and evolving information sets. This work can be grouped into three related strands: research on real-time data vintages, studies of forecast revisions produced by international institutions and outlook providers, and evidence from the social sciences on the limits of expert forecasting in complex systems.

The classic real-time data literature demonstrates that economic datasets evolve over time as measurement procedures, classifications, and estimation methods are revised. Croushore and Stark [2001] introduce real-time macroeconomic datasets and show that econometric inference and forecast evaluation can differ substantially depending on the data vintage used. Croushore [2006] further emphasizes that data revisions represent a persistent source of uncertainty that is conceptually distinct from model misspecification or forecast error. These insights are directly relevant for climate scenario analysis, where revisions to historical starting values across scenario vintages can propagate through model projections.

A related strand examines revisions to forecasts and outlooks produced by international organizations and energy agencies. Grigoli et al. [2015] document large and persistent revisions in real-time output gap estimates and show that such revisions can materially affect monetary policy assessments. Clements et al. [2025] analyze the revision process of professional forecasts and find that changes in uncertainty assessments across forecast vintages follow systematic patterns rather than reflecting purely random news. In the energy domain, Wachtmeister et al. [2018] show that long-term projections in successive editions of the International Energy Agency (IEA) *World Energy Outlook* are repeatedly revised, with forecast errors and outlook adjustments persisting even as forecast horizons shorten. Taken together, this literature highlights that repeated forecast and scenario publications reflect evolving data, assumptions, and institutional judgments rather than convergence toward a single stable reference trajectory.

Importantly, the objectives of these studies differ from those of climate scenario analysis. Most analyses of macroeconomic forecasts focus on relatively short horizons—typically five years or less—and evaluate performance using forecast accuracy, bias, or efficiency metrics relative to observed outcomes. In contrast, NGFS scenarios do not aim to predict a single baseline trajectory. Rather, they are designed to explore a range of possible transition pathways under alternative policy and technological assumptions. Given the profound uncertainty surrounding future climate outcomes and societal responses, scenario dispersion is therefore an intended feature rather than a modelling failure. Our analysis accordingly does not evaluate predictive accuracy but instead assesses the internal consistency of scenario vintages and compares variation across vintages with variation across scenarios and models.

Insights from the social sciences further emphasize the inherent limits of expert forecasting in complex systems. Grossmann et al. [2024] show that expert predictions of societal and economic change often perform no better than simple statistical benchmarks, particularly at long horizons. Their findings suggest that dispersion and revision in expert projections are intrinsic features of complex, evolving systems rather than anomalies associated with specific institutions or forecasters.

A substantial literature has also examined uncertainty in IAMs arising from alternative scenario narratives and policy assumptions. Scenario uncertainty reflects fundamentally different socio-economic pathways, policy timing, and technological availability rather than probabilistic forecasts. Studies such as Hawkins and Sutton [2009, 2011] and Hsiang and Kopp [2018] show that scenario uncertainty tends to dominate at longer horizons as divergence in emissions pathways and policy choices compounds over time. Within IAM-based mitigation analysis, scenario design is therefore understood as a deliberate tool to explore a wide range of plausible futures.

A parallel literature focuses on model uncertainty arising from structural differences across IAMs in technology representation, sectoral detail, and economic assumptions. Multi-model comparison exercises summarized in Gillingham et al. [2018], Kriegler et al. [2014], and Krey et al. [2019] document substantial dispersion in outcomes even when models are calibrated to similar climate targets. Dekker et al. [2023] further show that while emissions outcomes are often tightly linked to policy targets, large variation persists across models in sectoral energy use, technology deployment, and investment pathways. This literature demonstrates that model uncertainty is an intrinsic feature of IAM-based analysis, reflecting structural heterogeneity rather than noise or error.

More recently, research has examined how uncertainty in climate scenarios propagates into financial risk assessments. Tang et al. [2024] show that differences across climate scenarios and model pathways can generate substantial dispersion in firm-level valuation and default risk estimates within climate stress-testing frameworks, even when temperature targets are held constant. In such applications, scenario pathways are typically treated as fixed inputs, and the focus lies on how scenario and model uncertainty affect downstream financial metrics.

Taken together, this literature establishes scenario and model uncertainty as central features of IAM projections. However, most existing studies treat model frameworks, scenario definitions, and baseline calibrations as fixed within a given exercise. As a result, relatively little attention has been devoted to the extent to which uncertainty also arises from revisions to the same model–scenario pair over time. From the perspective of scenario users, such revisions introduce uncertainty regarding the persistence and robustness of scenario-implied pathways, even when scenario narratives and temperature targets remain unchanged. Although [NGFS, 2024] documents methodological developments across successive scenario vintages, to our knowledge this study is the first to systematically quantify variation in NGFS IAM outputs across vintages, scenarios, and models, and the first to evaluate the evolution of economic variables across annually released scenario vintages over several decades using a unified empirical framework.

Section 2 introduces the conceptual framework and presents the decomposition of uncertainty across scenarios, models, and vintages, including the role of initial-condition revisions and recalibration effects within vintages. Section 3 describes the data and empirical methodology used to quantify uncertainty across NGFS variables. Section 4 presents the empirical results and compares the relative importance of vintage, scenario, and model uncertainty across variables and horizons. Section 5 discusses the implications of these findings for the interpretation and use of climate scenarios in financial risk analysis. Section 6 concludes.

2 A Conceptual Framework of Uncertainty in NGFS Scenarios

Climate scenario projections embed several distinct sources of uncertainty. In the context of the NGFS scenario framework, these sources arise from three dimensions: differences in policy pathways across scenarios, differences in structural representations across IAMs, and revisions to the information set used to construct successive scenario vintages. Understanding how these dimensions contribute to projection dispersion is essential when scenario outputs are used for financial risk assessment, where decisions are often sensitive to the near-term trajectory of economic variables.

The literature on uncertainty provides several conceptual frameworks for interpreting these sources. In the classical distinction introduced by [Knight, 1921], situations of measurable risk—where probability distributions are known—are distinguished from situations of true uncertainty, where both outcomes and probabilities are not fully specified. Climate transition scenarios exhibit many characteristics of this latter category. They rely on complex structural models of energy, technology, and economic systems whose parameters and behavioral relationships are not known with precision and may evolve over time.

In the climate modeling literature, uncertainty in long-run projections is often decomposed into several components. Projection studies commonly distinguish between forcing uncertainty, structural model uncertainty, and internal variability (e.g., [Hawkins and Sutton, 2009]). For IAMs used in climate scenario design, these concepts have direct analogues: differences across policy pathways correspond to forcing uncertainty, alternative IAMs represent structural uncertainty, and differences in initial conditions reflect uncertainty about the current state of the system.

More recent work in climate-risk assessment further classifies uncertainty into epistemic, aleatory, and normative components, reflecting uncertainty about knowledge, stochastic variability, and societal choices (e.g., [Meiler et al., 2025]). These distinctions closely relate to the notion of model uncertainty studied in macro-finance. In that literature, Knightian uncertainty is often formalized as uncertainty about the true data-generating model governing economic fundamentals. Rather than assigning a single probability distribution to outcomes, agents consider a set of plausible models and evaluate decisions in a manner that is robust to potential model misspecification ([Hansen and Sargent, 2008]; [Drechsler, 2013]).

From this perspective, the IAMs used in climate scenario design can be interpreted as alternative candidate models of the energy–economy–climate system. Differences across models therefore reflect uncertainty about the correct structural representation of the underlying system. Scenario differences capture uncertainty about future policy pathways, while vintage revisions reflect updates to the information set used to initialize and calibrate the models. Taken together, these dimensions generate a set of plausible transition trajectories rather than a single probabilistic forecast, consistent with the notion of Knightian uncertainty emphasized in the economic literature.

The NGFS scenario framework naturally embeds these different dimensions. Each projection is generated by an IAM that simulates the joint evolution of energy systems, macroeconomic activity, and emissions under a particular policy pathway. Over time, the NGFS releases updated "vintages" of scenarios that incorporate new historical data, improved calibration, and methodological revisions. These updates introduce an additional dimension of uncertainty reflecting the evolving knowledge base used to construct the scenarios.

To formalize these ideas, let $m \in \mathcal{M}$ index integrated assessment models, $s \in \mathcal{S}$ index scenarios, and $v \in \mathcal{V}$ index NGFS vintages. Let $t \in \mathcal{T}$ denote projection years and let Y denote a given NGFS variable. The projected value of variable Y at horizon t, produced by model m, scenario s, and vintage v, can be written as

$$Y_t^{m,s,v} = F^m\big(x_0^v, \theta^{m,v}, s\big), \tag{1}$$

where x_0^v denotes the vintage-specific initial condition and $\theta^{m,v}$ denotes model parameters, calibration choices, and structural assumptions that may vary across both models and vintages. The notation F^m denotes the structural mapping implied by IAM m. It allows the mapping from assumptions to outputs to evolve across vintages, reflecting possible changes in data, calibration, parameterization, or model implementation. x_0^v represents the vector of baseline economic and technological conditions at the time of scenario release v, $\theta^{m,v}$ denotes the calibrated parameters of the model, and s represents the scenario-specific policy pathway. This representation highlights three distinct sources of uncertainty. Throughout the analysis, we isolate each source by holding all other dimensions fixed and varying only the dimension of interest. This conditional comparison allows us to identify the contribution of scenario, model, or vintage assumptions while keeping the remaining dimensions constant. The empirical implementation of this identification strategy is described in Section 3.

2.1 Scenario Uncertainty

Scenario uncertainty arises from differences in policy assumptions governing the transition pathway. Holding the model and scenario vintage fixed, projections may differ across alternative scenarios.

$$\Delta_s Y_t = F^m(x_0^v, \theta^{m,v}, s_1) - F^m(x_0^v, \theta^{m,v}, s_2). \tag{2}$$

Here s_1 and s_2 denote alternative transition scenarios within the NGFS framework, such as *Net Zero 2050*, *Delayed Transition*, or *Current Policies*, which differ in the timing and stringency of climate policy, technological deployment, and socioeconomic assumptions. This dimension captures uncertainty about the policy environment, including carbon pricing trajectories, mitigation timing, and technological deployment pathways. It corresponds closely to the notion of forcing uncertainty in climate science and reflects the inherently political and institutional nature of transition pathways.

2.2 Model Uncertainty

Model uncertainty reflects differences in the structural representation of the energy–economy system across IAMs. Holding the scenario and vintage fixed, projections may differ because different models embed different technological assumptions, behavioral elasticities, or sectoral structures:

$$\Delta_m Y_t = F^{m_1}(x_0^v, \theta^{m_1,v}, s) - F^{m_2}(x_0^v, \theta^{m_2,v}, s). \tag{3}$$

Here m_1 and m_2 denote different IAMs, such as REMIND–MAgPIE, MESSAGEix–GLOBIOM, or GCAM, which differ in model structure, sectoral representation, and technology assumptions. This dimension captures the structural uncertainty inherent in large-scale modeling exercises. Even when calibrated to the same baseline conditions and policy pathway, different modeling frameworks may generate divergent projections.

2.3 Vintage Uncertainty

A third source of uncertainty arises from revisions to the scenario framework itself. Successive NGFS releases incorporate updated historical data, revised assumptions about technology costs and resource availability, and methodological improvements to the models. These revisions generate differences across scenario vintages even when the same model and scenario are considered.

For a given model m and scenario s, each vintage $v \in \mathcal{V}$ is characterized by a vintage-specific information set $\mathcal{I}^v$, consisting of: (i) historical input data and initial conditions x_0^v; (ii) model parameters and calibration choices $\theta^{m,v}$; and (iii) scenario assumptions embedded in the published NGFS release. Vintage comparisons therefore capture changes in the underlying information set used to generate projections over time. Holding the model and scenario fixed, vintage uncertainty can be expressed as

$$\Delta_v Y_t = F^m(x_0^{v_1}, \theta^{m,v_1}, s) - F^m(x_0^{v_2}, \theta^{m,v_2}, s). \tag{4}$$

Vintage updates therefore capture the evolving knowledge base used in scenario construction. Unlike scenario or model uncertainty, which represent alternative contemporaneous assumptions, vintage uncertainty reflects the process of learning and revision over time.

Importantly, vintage updates simultaneously modify both the historic data used to initialize the models and the parameters used in model calibration. As a result, vintage uncertainty combines two conceptually distinct mechanisms. To clarify this distinction, vintage uncertainty can be decomposed as

$$\Delta_v Y_t = \underbrace{F^m(x_0^{v_1}, \theta^{m,v_1}, s) - F^m(x_0^{v_2}, \theta^{m,v_1}, s)}_{\text{Initial-condition uncertainty}} + \underbrace{F^m(x_0^{v_2}, \theta^{m,v_1}, s) - F^m(x_0^{v_2}, \theta^{m,v_2}, s)}_{\text{Recalibration uncertainty}}. \tag{5}$$

The first term captures differences in projections that arise solely from revisions to initial conditions x_0. These revisions reflect updated historical statistics, measurement improvements, or corrections to previously reported data. We refer to this component as *initial-condition uncertainty*. The second term captures the uncertainty of model recalibration and methodological updates that accompany new scenario releases.

Figure 1 summarizes the conceptual structure of uncertainty in NGFS projections implied by the framework above. At the highest level, projection differences can arise from three observable sources: differences across transition scenarios, differences across IAMs, and differences across NGFS scenario vintages. Vintage uncertainty can in turn be decomposed into two distinct mechanisms. The first is an initial-condition effect, reflecting revisions to the baseline values used to initialize model simulations. The second is a recalibration effect, reflecting updates to model parameters, assumptions, and calibration inputs within a given model across successive scenario releases. From the perspective of scenario users, these different sources of dispersion correspond to distinct decision problems. Dispersion across scenarios raises the question of which policy pathway might ultimately materialize. Dispersion across IAMs raises the question of which structural representation of the energy–economy system most accurately captures the underlying dynamics. Differences across scenario vintages introduce a different type of question: what if the information used to construct the scenario is itself incomplete or mismeasured?

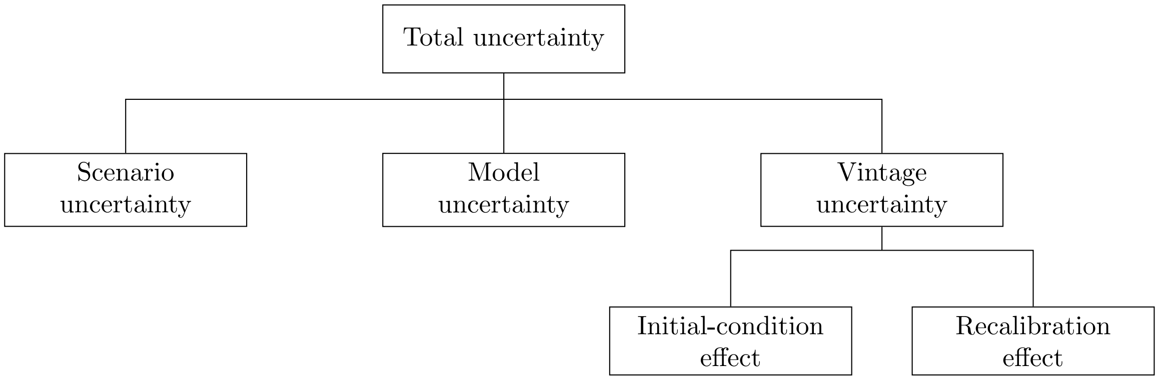
Successive NGFS releases incorporate revised historical data, updated technology cost assumptions, and model recalibrations that reflect improvements in the underlying knowledge base. As a result, vintage differences do not merely represent alternative contemporaneous assumptions, but instead capture revisions to the information set used to initialize and calibrate the models. For scenario users such as financial institutions or regulators conducting climate risk assessments, vintage dispersion therefore represents epistemic uncertainty about the baseline state of the economic and energy system rather than uncertainty about the transition pathway itself. Because near-term projections depend strongly on the initial conditions used to initialize the models, revisions to these baseline conditions can generate substantial differences in short-horizon projections even when the policy scenario and model structure are held fixed. This decomposition provides the conceptual basis for the empirical strategy developed in Section 3, which measures each source of uncertainty by holding the remaining dimensions fixed.

2.4 Implications for the Measurement of Uncertainty

The empirical analysis in this paper focuses on the observable dispersion across scenario outputs generated by different vintages, models, and scenarios. Within the framework described above, the dispersion across vintages reflects a combination of initial-condition and recalibration effects, both of which are components of vintage uncertainty.

Because revisions to initial conditions primarily affect the starting point of the projection paths, the contribution of the initial-condition component is expected to be most pronounced at short horizons. By contrast, differences in model structure and policy assumptions typically

Figure 1: Conceptual decomposition of uncertainty in NGFS projections



Note: Total uncertainty in NGFS scenario projections can arise from three main sources: scenario uncertainty (differences in policy and transition assumptions), model uncertainty (differences across IAMs), and vintage uncertainty (differences across NGFS scenario releases). Vintage uncertainty can be further decomposed into an initial-condition effect, reflecting revisions to reported baseline values used to initialize the models, and a recalibration effect, reflecting updates to model parameters, assumptions, or calibration inputs across vintages.

generate divergences that accumulate over longer horizons as projected pathways evolve.

The empirical sections that follow quantify these patterns by comparing the dispersion of projections across scenario vintages, models, and scenarios while holding the remaining dimensions fixed.

2.5 Initial Conditions and Horizon Dependence of Uncertainty

An important implication of the framework above concerns the relative importance of initial-condition uncertainty across projection horizons.

Let the projected value of a variable in vintage v be written as

$$Y_t^v = Y_0^v g_t^v, \tag{6}$$

where Y_0^v denotes the initial condition at the time the scenario is released and g_t^v represents the normalized transition pathway implied by the model, with $g_0^v = 1$. Variation across vintages can therefore arise from two distinct sources: differences in initial conditions Y_0^v and differences in projected transition dynamics g_t^v.

A useful variance decomposition follows directly from this multiplicative representation. Treating variation across vintages as variation in the random variables Y_0^v and g_t^v, the variance of projections at horizon t can be written as

$$\begin{aligned}\operatorname{Var}(Y_t^v) &= \mathbb{E}[g_t^v]^2 \operatorname{Var}(Y_0^v) + \mathbb{E}[Y_0^v]^2 \operatorname{Var}(g_t^v) \\ &\quad + 2\,\mathbb{E}[Y_0^v]\,\mathbb{E}[g_t^v] \operatorname{Cov}(Y_0^v, g_t^v) + \operatorname{Var}(Y_0^v) \operatorname{Var}(g_t^v).\end{aligned} \tag{7}$$

The first term captures dispersion arising from revisions to initial conditions, while the second term reflects dispersion arising from differences in normalized transition pathways. The remaining two terms capture interaction effects between the two sources of variation.

To obtain a tractable first-order approximation, we abstract from the covariance term and the higher-order interaction term $\operatorname{Var}(Y_0^v) \operatorname{Var}(g_t^v)$. This yields

$$\operatorname{Var}(Y_t^v) \approx \mathbb{E}[g_t^v]^2 \operatorname{Var}(Y_0^v) + \mathbb{E}[Y_0^v]^2 \operatorname{Var}(g_t^v). \tag{8}$$

Evaluating this expression at representative values of Y_0^v and g_t^v gives

$$\operatorname{Var}(Y_t^v) \approx (g_t^v)^2 \operatorname{Var}(Y_0^v) + (Y_0^v)^2 \operatorname{Var}(g_t^v). \tag{9}$$

This approximation makes the horizon dependence of uncertainty transparent. At short horizons, g_t^v is close to unity and $\operatorname{Var}(g_t^v)$ remains small, implying that dispersion in initial conditions constitutes a large share of total uncertainty. As the projection horizon increases, differences in transition pathways accumulate and $\operatorname{Var}(g_t^v)$ grows, causing the relative contribution of initial-condition uncertainty to decline. Near-term projections are therefore particularly sensitive to revisions in the initial conditions used to initialize scenario models, while longer-horizon projections are increasingly driven by differences in modeled transition dynamics.

When long-run outcomes are anchored by terminal targets in absolute levels, however, revisions to initial conditions may also propagate into the dynamics of the normalized transition path itself. In such cases, changes in initial conditions do not merely shift the starting point of the trajectory, but can induce corresponding changes in projected pathways as the model adjusts to remain consistent with the terminal constraint.

A useful qualification is that the normalized transition path need not be independent of the initial condition. More generally, the indexed pathway can be written as

$$Y_t^v = G_t(Y_0^v; \theta^v, s)\,, \tag{10}$$

where $G_t(\cdot)$ represents the transition mapping implied by the model. In this formulation the normalized growth factor $g_t^v = Y_t^v / Y_0^v$ may depend on the initial condition, implying that revisions to Y_0 can also alter the projected transition path. As a result, dispersion in projections may reflect not only uncertainty in initial conditions and pathway dynamics individually, but also their covariance. In this formulation the normalized growth factor $g_t^v = Y_t^v / Y_0^v$ may depend on the initial condition, implying that revisions to Y_0 can also alter the projected transition path. The sign and magnitude of $\mathrm{Cov}(Y_0, g_t)$ depend on how models recalibrate the transition path in response to revisions in initial conditions. In some cases $\mathrm{Cov}(Y_0, g_t) < 0$, for example when models impose terminal constraints in absolute levels so that a higher initial condition requires a flatter transition path to reach the same endpoint. In other cases $\mathrm{Cov}(Y_0, g_t) > 0$ if higher initial conditions are associated with more persistent structural dynamics or slower adjustment. When revisions to initial conditions do not materially affect the transition mapping, $\mathrm{Cov}(Y_0, g_t) \approx 0$.

This interaction arises naturally when models impose terminal constraints in absolute levels, such as

$$Y_T^v = \bar{Y}_T. \tag{11}$$

In such cases, revisions to initial conditions mechanically require adjustments to intermediate dynamics to remain consistent with the terminal target. An upward revision to the starting level therefore implies a lower implied growth path toward the terminal target, generating negative covariance between the initial condition and the normalized transition pathway. Consequently, revisions to initial conditions do not merely shift the starting point of the trajectory but can propagate throughout the entire transition path.

3 Data and Methodology

For this comparative study, we analyze NGFS scenario data released between NGFS v2 and v5, covering four scenario vintages, five common climate scenarios, and three IAMs: REMIND–MAgPIE, MESSAGEix–GLOBIOM, and GCAM. NGFS vintage v1 is excluded because scenario narratives and labels differ substantially from those used in later vintages, scenario availability is not consistent across models, and several variable definitions change materially, preventing meaningful cross-vintage comparison. The analysis therefore focuses on NGFS vintages for which consistent, comparable outputs are available across models, scenarios, and variables. Both scenarios and variables are selected based on their consistent definition and availability across all NGFS vintages from v2 onwards, ensuring cross-vintage comparability and avoiding spurious variation arising from changes in coverage or variable definitions. This selection balances the need for temporal depth with the requirement of cross-vintage comparability. From these combinations, we extract 254 variables from the NGFS–IIASA Data Explorer.

We illustrate this empirical strategy using global steel production, which provides a transparent example of how the uncertainty dimensions introduced in Section 2 can be isolated in practice. Moreover, steel production is emissions-intensive, economically significant, and directly relevant for climate stress testing. In addition, it displays substantial variability across the three main uncertainty dimensions examined in this study—scenario, model, and vintage—while also exhibiting large revisions in initial condition levels across scenario vintages. These revisions generate a sizable initial-condition component of vintage uncertainty, making steel a useful case for assessing how multiple sources of uncertainty interact.

All steel production pathways span the period 2020–2050 in five-year steps. As discussed in Section 2, paths are indexed to a common base year in order to separate the initial-condition component of vintage uncertainty from differences in the subsequent evolution of projection pathways. This normalization removes dispersion arising from revisions to initial condition levels across scenario vintages, allowing path-based uncertainty to be examined independently. The distinction is particularly important for steel, where revisions to global production levels across NGFS vintages are large and non-monotonic.

As also discussed in Section 2, the conceptual framework distinguishes uncertainty arising from scenario assumptions, model structure, and scenario vintages. Operationalizing this framework empirically requires fixing all but one dimension at a time. If model, scenario, and vintage are allowed to vary simultaneously, observed dispersion cannot be attributed to a specific source. For steel production, failing to fix these dimensions would conflate revisions to initial conditions across scenario vintages with changes in decarbonization assumptions, technology deployment, or model structure. We therefore adopt a reference configuration consisting of NGFS v5, the Net Zero 2050 scenario, and the REMIND–MAgPIE model, and vary one dimension at a time around this benchmark.

NGFS v5 is chosen as the reference vintage because it represents the most recent publicly available release and therefore reflects the cumulative effects of model revisions, updates to initial conditions, and scenario refinements observed across earlier vintages. Using the latest vintage as the reference point reflects the information set currently available to scenario users and allows dispersion observed across earlier vintages to provide empirical evidence on the magnitude of epistemic uncertainty embedded in the scenario construction process. The choice of reference vintage affects only the normalization used to construct the common-base indexed paths for visual comparison and does not affect the uncertainty measures, which are computed directly from the dispersion across all vintages. The Net Zero 2050 scenario is selected as the reference scenario because it is one of the most widely discussed transition pathways in public and policy discourse and features prominently in regulatory climate stress-testing frameworks. REMIND–MAgPIE is chosen as the reference model because it is a core integrated assessment model underlying NGFS scenarios and is reflected, either directly or indirectly, in supervisory climate stress testing exercises such as those conducted by [ACPR, 2020] and the [Hong Kong Monetary Authority , 2023]. In addition, REMIND is extensively used in the integrated assessment literature and in multi-model comparison exercises (e.g. [Kriegler et al., 2014]; [Krey et al., 2019]; [Riahi et al., 2021]), making it a well-documented and representative benchmark for analysing scenario-based uncertainty. Together, this reference configuration reflects a commonly used and policy-relevant baseline against which alternative vintages, scenarios, and models can be compared[3].

Figure 2 illustrates these concepts. Raw steel production paths exhibit substantial dispersion

[3] Appendix B shows some robustness results for selected variables using as reference scenario Current Policies or as reference model MESSAGE or GCAM.

in reported 2020 values across NGFS vintages, highlighting large revisions to initial conditions. Notably, these revisions do not converge over time: reported 2020 production ranges from roughly 1.7 billion tonnes in NGFS v2 to nearly 2.0 billion tonnes in v4, before declining again in v5. This non-monotonic pattern illustrates that scenario vintages incorporate substantial revisions to initial conditions, which constitute the initial-condition component of vintage uncertainty discussed in Section 2.

After indexing paths to a common base year (2020), differences in the subsequent evolution of the projected pathways can be analysed independently of their original units and magnitudes. This transformation reveals that substantial divergence across vintages remains even after differences in starting values have been standardised. Unlike carbon emissions, steel production is not governed by a terminal target, which complicates the interpretation of projected pathways. For example, although NGFS v5 steel production is slightly lower than v4 in 2020, it exceeds v4 production in 2050 by more than 20%. Inter-model differences further amplify uncertainty, with MESSAGEix–GLOBIOM projecting materially higher production levels than REMIND and GCAM within the same vintage and scenario.

These patterns demonstrate that, even after removing the initial-condition component of vintage uncertainty, path-based divergence across scenario vintages can rival or exceed dispersion across scenarios and models for a systemically important industrial variable. The steel example therefore motivates the explicit separation of scenario, model, and vintage uncertainty—while distinguishing the initial-condition component within vintage updates—which we formalize and quantify in the following sections for the full set of variables.

3.1 Measurement of uncertainty

Because NGFS scenarios do not provide probabilistic forecasts, uncertainty cannot be summarized using conventional variance-based confidence intervals. Instead, we follow the common practice in the climate modelling literature and measure uncertainty using dispersion across projections. Ensemble-based analyses of climate projections often interpret the spread across model outputs as a proxy for uncertainty when probability distributions are not available ([Hawkins and Sutton, 2009]). In this framework, differences across forcing pathways correspond to scenario uncertainty, while differences across models capture structural model uncertainty. Similarly, dispersion across IAM outputs is commonly interpreted as ensemble uncertainty in multi-model assessments ([Gillingham et al., 2018]). We apply this logic to NGFS projections and extend it by incorporating dispersion across scenario vintages and revisions to initial conditions. Let each variable path Y_t span the period from 2020 to 2050 in five-year increments, $t \in \{2020, 2025, \ldots, 2050\}$, such that each path consists of $N = 7$ observations: $Y_{2020}, \ldots, Y_{2050}$. Throughout this section, we distinguish between *level-based* uncertainty associated with revisions to initial conditions and *path-based* uncertainty associated with divergence in projected transition pathways.

3.1.1 Common normalization of variable paths

To ensure comparability across variables with different units and magnitudes, the analysis normalizes variable paths in two steps. First, each raw projection is converted into an indexed path relative to its own 2020 starting value:

$$y_t^{m,s,v} = \frac{Y_t^{m,s,v}}{Y_{2020}^{m,s,v}}. \tag{12}$$

Second, the indexed path is rescaled to a common starting value given by the latest available NGFS vintage in 2020. Let v^* denote the latest vintage. The common-base path is defined as

$$\widehat{Y}_t^{m,s,v} = y_t^{m,s,v}\, Y_{2020}^{m,s,v^*}. \tag{13}$$

By construction, all common-base paths start from the same 2020 value, Y_{2020}^{m,s,v^*}, while preserving the growth profile implied by each model–scenario–vintage path. All path-based distance measures are computed using $\widehat{Y}_t^{m,s,v}$. This normalization puts path-based dispersion on a common scale across variables.

3.1.2 Initial-condition uncertainty

Initial-condition uncertainty captures dispersion in reported initial-condition levels across NGFS scenario vintages. In the decomposition of projection updates introduced in Equation 5, this corresponds to the first component, which isolates the effect of revisions to initial conditions while holding model parameters and scenario assumptions fixed. For a fixed variable, model, and scenario, initial-condition uncertainty is therefore measured using raw (unindexed) levels and defined as the dispersion of reported 2020 values across vintages:

$$D_{\mathrm{I}} = \max_{v} \left(Y_{2020}^{(v)} \right) - \min_{v} \left(Y_{2020}^{(v)} \right), \tag{14}$$

where v indexes NGFS vintages.

Figure 2 (Panel A, left-hand side) illustrates initial-condition uncertainty for global steel production, where dispersion at $t = 2020$ is shown as the vertical distance between the maximum and minimum reported starting values across NGFS vintages for a fixed model and scenario.

3.1.3 Path-based uncertainty

Path-based uncertainty captures divergence in projected transition dynamics after controlling for revisions to initial conditions. In the framework of Section 2, this corresponds to changes in projections arising from the Δ operators defined in Equations (2) and (3) and from the recalibration effect in Equation 5. Specifically, scenario and model uncertainty reflect differences in projections induced by changes in scenario assumptions s or model structure m, while vintage path uncertainty corresponds to the recalibration effect that arises when model parameters or calibration inputs $\theta^{m,v}$ change across vintages while initial conditions are held fixed. Empirically, these effects are first identified using indexed variable paths, which remove differences in scale while preserving the projected dynamics. Vintage, scenario, and model uncertainty are measured using indexed paths and exploit the time dimension of each variable over the 2020–2050 horizon. Panel B of Figure 2 shows indexed steel production paths, for which dispersion at $t = 2020$ is zero by construction and remaining differences reflect divergence in projected transition dynamics rather than revisions to initial conditions. For a given uncertainty dimension $k \in \{R, S, M\}$, where R, S, and M denote recalibration, scenario, and model uncertainty, respectively, and for each time point t, we identify the contemporaneous maximum and minimum indexed values observed across pathways while holding the remaining dimensions fixed. Figure 2 (Panel C) illustrates this construction: at each time step, broken vertical line segments indicate the distance between the contemporaneous minimum and maximum indexed values across pathways. The sequence of these extrema defines a composite maximum indexed path $y_{\max,k,t}$ and a composite minimum indexed path $y_{\min,k,t}$. These paths may combine segments from multiple individual projections, as the identity of the pathway attaining the maximum or minimum value may vary over time.

All variable paths are indexed to a common value at the base year 2020. As a consequence, dispersion at $t = 2020$ is zero by construction and does not contribute to any path-based uncertainty measure. The 2020 observation is therefore algebraically redundant for both long- and short-horizon metrics, but is retained in the notation for completeness.

Aggregate path-based uncertainty over the full projection horizon is quantified using the Euclidean distance between the composite maximum and minimum common-base paths:

$$D_k^{2050} = \sqrt{\sum_t \left(\widehat{Y}_{\max,k,t} - \widehat{Y}_{\min,k,t}\right)^2}, \tag{15}$$

where $t \in \{2020, 2025, \ldots, 2050\}$ and $\widehat{Y}_{\max,k,t}$ and $\widehat{Y}_{\min,k,t}$ denote the contemporaneous maximum and minimum values of the common-base paths, respectively. Since all paths share the same starting value by construction, the contribution of the $t = 2020$ term is zero.

To capture short-horizon uncertainty relevant for near-term policy and financial risk analysis, we define a complementary five-year metric based on the first forward time step only:

$$D_k^{2025} = \left|\widehat{Y}_{\max,k,2025} - \widehat{Y}_{\min,k,2025}\right|. \tag{16}$$

The path-based distance measures introduced above are computed using the indexed variable paths defined in Section 3.1.1. The indexed paths defined in Section 3.1.1 are sufficient for constructing the path-based distance measures D_k^{2025} and D_k^{2050}. However, the relative uncertainty ratios introduced in the following subsection compare these path-based measures with initial-condition uncertainty, which is defined in terms of revisions to the reported 2020 values across vintages. To ensure that both quantities are expressed on a common scale, the indexed paths are re-based using the common starting value defined in Equation (13) before the ratios are computed.

3.1.4 Relative uncertainty ratios

While the absolute dispersion metrics introduced above quantify the magnitude of uncertainty along each dimension, they do not directly indicate how these magnitudes compare across variables with different units and scales. To facilitate such comparisons, we construct relative uncertainty ratios that normalize path-based uncertainty by the dispersion in initial conditions.

Having defined D_{I} as initial-condition uncertainty, and letting $k \in \{R, S, M\}$ denote recalibration, scenario, and model uncertainty, respectively, we define

$$\mathrm{ratio}_{k,I} = \frac{D_k^{2050}}{D_{\mathrm{I}}}, \tag{17}$$

where D_k^{2050} denotes the path-based uncertainty measure over the 2020–2050 projection horizon. These ratios normalize path-based dispersion by initial-condition dispersion, providing a scale-free measure that facilitates comparison across variables whose levels and units differ substantially. They therefore serve as descriptive indicators of how large dynamic pathway divergence is relative to revisions in initial conditions. A lower value of $\mathrm{ratio}_{k,I}$ indicates that dispersion in initial conditions is large relative to uncertainty associated with dimension k, whereas a higher value indicates that path-based uncertainty dominates.

For example, a low value of $\mathrm{ratio}_{R,I}$ implies that variation in reported 2020 starting values across NGFS vintages is large compared with variation in projected pathways across vintages, conditional on model and scenario. Conversely, a high value of $\mathrm{ratio}_{R,I}$ indicates that recalibration-related divergence in projected dynamics dominates revisions to initial conditions.

Analogous interpretations apply to $\mathrm{ratio}_{S,I}$ and $\mathrm{ratio}_{M,I}$, which compare scenario and model uncertainty, respectively, to initial-condition uncertainty. Importantly, these ratios are dimensionless and therefore comparable across variables with different units and scales.

To assess whether the relative importance of initial-condition uncertainty differs over shorter horizons, we also compute ratios based on the five-year path-based metric:

$$\mathrm{ratio}_{k,I}^{2025} = \frac{D_k^{2025}}{D_{\mathrm{I}}}, \tag{18}$$

These short-horizon ratios are particularly relevant for applications focusing on near-term transition risks, such as climate stress testing and supervisory exercises. In Section 4, we also report the complementary ratios

$$\text{ratio}_{k,R}, \qquad \text{ratio}_{k,R}^{2025}, \qquad k \in \{S, M\},$$

which normalize scenario and model uncertainty by recalibration uncertainty. Throughout the analysis, we treat the ratio metrics as descriptive indicators of the relative importance of different uncertainty sources. We do not interpret them as evidence of model accuracy or forecasting performance.

4 Results

4.1 Empirical patterns of uncertainty across all variables

Figures 3 and 4 summarize uncertainty patterns across the full set of NGFS variables examined in this study. Figure 3 reports distributions of uncertainty ratios ($\text{ratio}_{k,I}$) that compare path-based uncertainty—arising from recalibrations, scenarios, and models—to dispersion in initial conditions across NGFS vintages. Ratios are expressed in $\log_{10}$ space and shown separately for the full projection horizon (2020–2050, left column) and for the short horizon (2020–2025, right column). Bars report the share of observations within each dominance bin, while the overlaid empirical cumulative distribution function (CDF), plotted only up to parity, summarizes the cumulative share of observations for which initial-condition uncertainty exceeds path-based uncertainty. Figure 4 reports analogous ratio distributions that compare recalibration uncertainty directly with scenario and model uncertainty at both horizons ($\text{ratio}_{k,R}$).

A first robust empirical finding is that initial-condition uncertainty frequently dominates path-based uncertainty, particularly over the initial five-year horizon. At the 2020–2025 horizon (right column of Figure 3), the CDF evaluated at parity indicates that dispersion in initial conditions exceeds recalibration uncertainty for approximately 40% of variables, scenario uncertainty for about 80% of variables, and model uncertainty for roughly 40% of variables. Because indexed paths coincide at 2020 by construction, these results imply that near-term uncertainty in NGFS outputs is often driven primarily by revisions to initial conditions rather than by modeled dynamics, scenario differentiation, or inter-model divergence. Beyond dominance at parity, initial-condition uncertainty can also exceed path-based uncertainty by very large margins: at the 2020–2050 horizon, dispersion in initial conditions is at least 20 times larger than scenario uncertainty for around 10% of variables, while at the 2020–2025 horizon this share rises to approximately 20%.

At the longer 2020–2050 horizon (left column of Figure 3), the dominance of initial-condition uncertainty declines but remains non-negligible. Dispersion in initial conditions exceeds recalibration uncertainty for more than 10% of variables, scenario uncertainty for roughly 20% of variables, and model uncertainty for around 10% of variables. These results indicate that, even over multi-decade horizons, revisions to initial conditions continue to play a quantitatively relevant role for a non-trivial subset of variables, rather than being fully dominated by projected dynamics.

A second key result is that recalibration uncertainty emerges as a first-order source of dispersion relative to scenario and model uncertainty. As shown in Figure 4, at the short 2020–2025 horizon, the cumulative distributions indicate that recalibration uncertainty exceeds scenario uncertainty for approximately 90% of variables and exceeds model uncertainty for almost 50% of variables. This indicates that, even over very short horizons, changes across successive NGFS releases within the same model–scenario pair can dominate differences arising from alternative scenario narratives and, in many cases, differences across IAMs. Beyond dominance at parity, Figure 4 shows that recalibration uncertainty can exceed scenario uncertainty by very large margins: at the 2020–2050 horizon, recalibration uncertainty is at least 20 times larger than scenario uncertainty for around 10% of variables, while at the 2020–2025 horizon this share rises to approximately 20%.

This pattern persists, albeit less strongly, over the full projection horizon. At the 2020–2050 horizon, recalibration uncertainty exceeds scenario uncertainty for roughly 50% of variables and exceeds model uncertainty for about 30% of variables. These findings demonstrate that annual updates to the same model–scenario pair can generate dispersion comparable to—or larger than—differences across scenario narratives or across IAMs, particularly for sectoral production, investment, and energy-demand variables.

Third, scenario uncertainty is systematic but not universally dominant. While scenario differentiation is clearly visible for variables tightly linked to mitigation ambition, Figures 3 and 4 show that scenario uncertainty is frequently dominated by either initial-condition or recalibration uncertainty, especially at the five-year horizon. This reflects both the gradual implementation of policy narratives and the limited scope for scenario-driven divergence in the near term, as many policy pathways initially share similar trajectories before diverging more strongly over longer horizons.

Finally, model uncertainty exhibits substantial heterogeneity across variables. Although inter-model differences dominate uncertainty for some variables, they do not do so uniformly. Across both figures, model uncertainty is often comparable in magnitude to recalibration uncertainty and, in a sizeable share of cases, is exceeded by it. This challenges the common presumption that inter-model differences are the primary source of uncertainty in NGFS scenario outputs.

Taken together, these results show that uncertainty in NGFS scenarios is fundamentally multi-dimensional. Revisions to initial conditions and annual model updates play a quantitatively important role alongside scenario narratives and inter-model differences, particularly over near-term horizons that are most relevant for climate stress testing and supervisory applications.

4.2 Empirical observations of uncertainty across selected variables

Figures 5–8 and Tables 1 – 2, together with Figure 2 which illustrates steel production pathways, examine a subset of NGFS variables chosen for two complementary reasons. First, these variables display particularly informative dispersion patterns across NGFS vintages, allowing the mechanisms and implications of recalibration uncertainty documented in Section 4.1 to be illustrated in concrete sectoral and technological contexts. Second, they are variables that feature prominently in climate stress testing, macro-financial analysis, and supervisory exercises, and are therefore representative of the quantities most commonly used in applied risk assessment.

Within this subset, carbon prices and CO_2 emissions serve a distinct benchmarking role. They are included not because they exhibit unusually large dispersion across vintages, but because they are tightly anchored by policy design. Carbon prices correspond to an explicit policy instrument embedded in NGFS scenarios, while CO_2 emissions capture the resulting aggregate mitigation outcome. As shown in Figure 5, both variables display comparatively stable pathway structures across recalibration, scenarios, and models. Their relative stability provides an important reference point: while headline policy targets and aggregate emissions outcomes remain broadly consistent across NGFS releases, this stability does not extend to many economically and financially relevant variables downstream.[4]

Across the remaining variables, heterogeneity in the dominant sources of uncertainty is pronounced. Figures 3 and 4 present distributions of uncertainty ratios in $\log_{10}$ space to facilitate comparison across variables with heavy-tailed dispersion. Tables 1–2 summarize the same information in multiplicative form, but reorganized to foreground qualitative dominance patterns. In each table cell, the top line reports which source of uncertainty dominates at the relevant horizon (initial-condition, recalibration, scenario, or model), while the second line reports the corresponding dominance factors for the near-term (2020–2025) and long-term (2020–2050) horizons. "Mixed" classifications indicate a reversal in dominance between the two horizons. As documented in the table notes, the reported multiplicative factors are obtained by exponentiating the corresponding $\log_{10}$ ratios shown in Figures 3 and 4.

Table 1 compares path-based uncertainty to initial-condition uncertainty by reporting dominance outcomes for recalibration/initial-condition, scenario/initial-condition, and model/initial-condition comparisons. It therefore addresses whether dispersion in a given variable is driven primarily by revisions to initial conditions or by divergence in projected dynamics. For several investment variables—most notably global nuclear electricity investment and global solar electricity investment—Table 1 reveals strong horizon dependence. Initial-condition uncertainty dominates in the near term, while path-based uncertainty (particularly recalibration or model uncertainty) dominates by 2050. This qualitative shift mirrors the movement across parity observed in the corresponding $\log_{10}$ ratio distributions in Figures 3 and 4.

Industrial production variables exhibit particularly strong sensitivity to model vintages. For global steel and cement production, Table 2 shows that recalibration uncertainty frequently dominates scenario and model uncertainty, especially at the near-term horizon. Figures 5 to 8 further indicate that recalibration-specific trajectories differ not only in levels but also in timing and curvature, even after all paths are aligned to a common starting point as illustrated in Figure 2. This pattern suggests that recalibration uncertainty primarily reflects changes

[4]Financial applications typically rely on downstream sectoral variables derived from climate scenarios, rather than solely on headline indicators such as emissions or temperature outcomes. Examples include industrial production, energy demand, investment, and sector-specific output pathways; see [Acharya et al., 2023, Noels et al., 2023, Wilkinson et al., 2025]

in modeled transition dynamics—such as technology deployment, substitution, or constraint structures—rather than revisions to initial conditions.

Energy investment variables are characterized by especially pronounced dispersion across recalibrations. Figures 6 and 7 reveal wide variation in both the timing and magnitude of investment peaks for nuclear, solar, and gas technologies across NGFS releases, even within a fixed scenario. Consistent with these visual patterns, Table 1 indicates that recalibration or model uncertainty often dominates scenario uncertainty for investment variables, underscoring the sensitivity of projected capital allocation paths to model updates, recalibration choices, and evolving internal assumptions.

Demand-side variables, including final energy demand in industry as well as in residential and commercial sectors, exhibit more moderate path-based dispersion. Table 1 shows that dispersion in initial conditions frequently dominates in the near term for these variables, with scenario or model uncertainty becoming more relevant only at longer horizons. This pattern is consistent with Figure 8, where near-term dispersion is concentrated around parity while longer-horizon divergence increases.

Land-use variables, such as global cropland area, display mixed uncertainty patterns. Panel C in Figure 5 shows moderate dispersion across recalibration, scenarios, and models, with no single uncertainty dimension consistently dominating across horizons. Correspondingly, Table 1 classifies cropland projections as "Mixed", indicating that dominance shifts across horizons and that projections are shaped by multiple interacting sources of uncertainty rather than a single dominant driver.

Taken together, the selected variables demonstrate that substantial and policy-relevant uncertainty persists well beyond headline mitigation indicators such as emissions and carbon prices. Even when aggregate policy objectives and emissions outcomes appear stable across NGFS releases, economically and financially salient variables—particularly sectoral production, investment, and energy demand—can display large and persistent recalibration-related dispersion. This finding is especially consequential for applications that rely on detailed sectoral and investment pathways rather than on aggregate climate outcomes alone.

5 Discussion

The empirical results above show that dispersion across model vintages can be comparable to, or larger than, dispersion across scenarios and across IAMs. This finding holds across a wide range of variables and is particularly pronounced at short horizons (2020–2025), where scenario narratives and model structures have limited time to diverge. By contrast, revisions to initial conditions (differences in reported 2020 starting points) and subsequent recalibrations across annual NGFS releases can already be substantial.

Importantly, the figures alone do not identify the mechanisms that generate this dispersion. They document *what* varies, but not *why*. To structure interpretation without over-claiming explanatory power, we therefore propose the following hypotheses. These are stated explicitly as conjectures to be tested against model documentation, change logs, and additional model outputs. Where such evidence is absent, interpretations should be treated as speculative.

5.1 Hypotheses on the sources of vintage dispersion

We distinguish between *initial-condition* and *recalibration* mechanisms as potential sources of dispersion across NGFS vintages. On the initial-condition side, differences in reported initial condition levels across vintages may reflect recalibration and base-year harmonization, arising from changes in underlying data sources, sectoral mappings, or calibration routines. This mechanism is consistent with variables for which large dispersion in raw levels collapses after alignment to a common starting point. In addition, some observed divergence may stem from changes in reporting practices or variable definitions, including adjustments to sectoral boundaries, unit conventions, or aggregation schemes, rather than substantive differences in modeled pathways. These mechanisms imply that a non-trivial share of vintage dispersion may be attributable to revisions in initial conditions or reporting conventions rather than to changes in projected dynamics.

On the recalibration side, persistent dispersion after alignment may arise from structural updates within a model family, including changes in parameterization, technology representation, constraints, or aggregation structures introduced across annual NGFS releases. In addition, structural updates within a model family are often reflected in changes to the internal model version used for a given NGFS release; for example, REMIND is reported under different internal versions across NGFS vintages (e.g. REMIND 3.2, REMIND 4.1), even when appearing under the common REMIND label in the NGFS scenario database. Even when scenario labels remain unchanged, scenario implementation may also drift across vintages due to evolving assumptions regarding policy timing, technology availability, or constraint settings, generating large within-scenario vintage spreads. Finally, differences in numerical implementation—such as solver settings, convergence criteria, or regularization choices—may affect investment timing and technology switching, amplifying vintage sensitivity even when long-run outcomes are less affected. Together, these recalibration mechanisms suggest that vintage dispersion may reflect substantive changes in model behavior over time. As before, these mechanisms are stated explicitly as conjectures and are intended to guide a structured validation exercise; where model documentation does not support a given mechanism, any associated interpretation should be treated as speculative and clearly labeled as such.

5.2 Interpretability challenges for systemically important variables

Beyond the magnitude of uncertainty, the results highlight a second challenge: *interpretability*. For several systemically important variables, it is difficult to construct a coherent high-level narrative that explains co-movements among closely related quantities, even within a single model and scenario. Similar concerns regarding the interpretability and narrative coherence of climate–economic scenarios are emphasized in [Baer, 2024], who argues that scenario usefulness for policy and financial applications depends not only on internal consistency but also on the ability to articulate plausible and transparent mechanisms underlying key outcomes. Related arguments appear in the IAM evaluation literature, where [Schwanitz, 2013] and [Wilson et al., 2021] emphasize that long-horizon scenarios should be assessed in terms of structural consistency, internal coherence, and their usefulness for exploring alternative transition pathways, rather than conventional forecast accuracy.

This issue is particularly evident for industrial production variables such as global steel and cement production. Steel and cement are key industrial sectors with high emissions intensity, high decarbonization costs, and major relevance for macroeconomic activity through construction and infrastructure ([IEA, 2018] [IEA, 2020]). They are therefore central to climate stress testing, macro-financial analysis, and policy evaluation.

Apart from the pronounced initial-condition and recalibration uncertainty observed for these variables, their joint evolution raises questions about cross-sector consistency within transition narratives. For example, in the NGFS v5 REMIND Net Zero scenario, global steel production declines only modestly, while cement production contracts substantially. Given the close linkage between these materials in construction, such divergence is difficult to reconcile without assuming significant changes in material substitution, construction practices, or technology adoption. This interpretation is consistent with the industrial decarbonisation literature, which highlights that reductions in cement demand rely heavily on material efficiency and substitution, while steel demand is typically more closely tied to structural investment dynamics ([Allwood et al., 2010] [Habert et al., 2020]).

This raises several questions. Does the sharp drop in cement production implicitly signal lower-than-expected growth in the construction sector, and if so, how is this consistent with the relatively modest macroeconomic impacts typically reported for NGFS scenarios? Conversely, steel production remaining relatively stable under Net Zero suggests that either high carbon costs are borne, carbon capture and storage is deployed at scale, or major technological change (e.g. widespread adoption of electric arc furnaces) occurs successfully. In that case, are the costs associated with these radical changes consistent with the implied macroeconomic outcomes?

These interpretability challenges are directly related to the quantitative dominance patterns documented in Tables 1 and 2. For steel and cement, initial-condition uncertainty and recalibration uncertainty frequently dominate scenario and model uncertainty (i.e., ratios below unity at one or both horizons). This implies that a substantial share of variation in projected sectoral paths arises from revisions to initial conditions and from changes across annual NGFS releases, rather than from alternative scenario narratives or stable inter-model differences. As a result, large shifts in sectoral trajectories across vintages cannot be straightforwardly interpreted as reflecting changes in policy ambition alone. Similar cautionary notes on the interpretation of NGFS sectoral outputs have recently been raised within the NGFS research community itself [Gidden et al., 2023].

5.3 Implications for scenario use and stress testing

Taken together, these findings underscore that NGFS scenarios should be interpreted as exploratory pathways characterized by multiple, interacting layers of uncertainty, including uncertainty arising from annual updates to the underlying modelling frameworks themselves. For downstream users— particularly in financial stability analysis, supervisory exercises, and climate stress testing—this implies that robustness assessments should not be confined to comparisons across scenarios or across models alone, but should also explicitly account for revisions to initial conditions and sensitivity to scenario vintages.

This distinction is especially important given that most financial decision-making, risk management, and supervisory stress testing operates over relatively short horizons. Our empirical results show that precisely over these near-term horizons (2020–2025), initial-condition uncertainty and recalibration uncertainty frequently dominate scenario and model uncertainty across a large share of variables. Because indexed paths coincide at the base year by construction, this dominance implies that near-term dispersion in NGFS outputs is driven primarily by revisions to initial conditions and by changes across annual NGFS releases, rather than by divergence in scenario narratives or inter-model structure. In contrast, scenario and model uncertainty become more prominent only at longer horizons, which are typically less relevant for balance-sheet risk assessment, capital planning, and supervisory exercises.

These implications can be interpreted through the lens of investment under uncertainty, as formalized by [Dixit and Pindyck, 1994] and extended in the macroeconomic literature by [Bloom, 2009]. In this framework, uncertainty— particularly when concentrated in the near term—creates an option value of waiting and dampens irreversible investment. Applied to NGFS scenarios, dominance of initial-condition and recalibration uncertainty at short horizons suggests that modeled near-term investment signals may be especially sensitive to scenario vintage. If revisions to initial conditions or model updates materially alter projected investment paths from one vintage to the next, these signals are unlikely to represent stable expectations about transition dynamics and should therefore be interpreted with caution in financial applications.

A stylized example helps illustrate this mechanism. Consider an investor or financial institution that, based on an earlier NGFS release, interprets the Net Zero scenario as implying stable or modestly growing global steel production over the transition horizon, and therefore assumes continued capacity utilization and investment demand in steel-related assets. In a subsequent NGFS vintage, however, the same model–scenario pair may imply a material decline in steel production, driven by revised starting values or updated model assumptions rather than by an explicit change in policy ambition. From the perspective of the investor, this revision constitutes a shift in the information set rather than a realized transition outcome, yet it can materially alter inferred transition risks, valuations, and capital allocation decisions.

Where initial-condition or recalibration uncertainty dominates, scenario labels such as *Net Zero* or *Current Policies* provide an incomplete guide to the underlying economic, technological, and sectoral assumptions embedded in the projections. In such cases, changes in outcomes across NGFS releases may reflect revisions to initial conditions or model updates rather than shifts in policy ambition or transition narratives. This highlights the need for cautious interpretation of NGFS outputs in applied settings and motivates the use of complementary analyses—such as sensitivity checks across vintages, initial condition harmonization, and cross-variable consistency checks—when scenarios are used for decision-making, risk assessment, or disclosure.

Communicating uncertainty through ensembles of scenarios is standard practice in climate science and policy-oriented decision-making under deep uncertainty, where projections are understood as exploratory representations of plausible futures rather than probabilistic forecasts ([van Asselt & Rotmans, 2002]; [Walker et al., 2003]; [Kwakkel et al., 2016]). By contrast, economic and sectoral outputs from IAMs used in climate stress testing are often communicated as point paths from a single scenario vintage, despite substantial dispersion across vintages, models, and scenarios. Incorporating simple uncertainty bands that summarize empirically observed dispersion would therefore provide a transparent complement to scenario narratives, helping users assess robustness without recasting IAM projections as forecasts.

6 Conclusion

This paper examines uncertainty in NGFS climate scenario projections across three main dimensions—vintage, scenario, and model uncertainty—while further decomposing vintage uncertainty into two components: uncertainty arising from revisions to initial conditions and uncertainty arising from recalibration across successive model releases. Using 254 variables across four NGFS vintages, five climate scenarios, and three IAMs, we construct comparable point-based and path-based uncertainty metrics over both near-term (2020–2025) and long-term (2020–2050) horizons. By distinguishing revisions to initial conditions and model recalibration from uncertainty embedded in scenario narratives and model structure, the analysis provides a more granular perspective on how NGFS scenario outputs evolve across releases.

The results show that revisions to initial conditions and updates across scenario vintages are first-order features of NGFS outputs. Initial-condition uncertainty frequently dominates path-based uncertainty at short horizons, while recalibration across vintages often rivals or exceeds dispersion arising from alternative scenarios or different IAM structures, including at longer horizons. A distributional perspective further reveals that these differences are often substantial rather than marginal: for a non-trivial subset of variables, uncertainty arising from initial conditions or recalibration exceeds competing sources by large margins, with order-of-magnitude differences observed for roughly 10–20% of variables depending on the horizon. These patterns appear across a broad range of variables—including sectoral production, investment, energy demand, and land-use indicators—and persist even when conditioning on a fixed model–scenario pair. Consequently, dispersion in NGFS outputs cannot be attributed solely to alternative policy narratives or structural differences across IAMs.

These findings have important implications for the interpretation and use of NGFS scenarios in financial risk analysis, stress testing, and policy evaluation. Because most financial decision-making operates over relatively short horizons, the prominence—and in some cases strong dominance—of initial-condition and recalibration-related uncertainty in the near term suggests that scenario-based signals for investment, production, and energy demand should be interpreted with caution. Changes in projected outcomes across successive NGFS releases may reflect revisions to starting values or model recalibration rather than shifts in policy ambition or transition narratives.

Beyond its applied implications, this study contributes to the broader literature on IAMs by highlighting cross-vintage coherence as a relevant and previously underexplored dimension of model evaluation. While existing IAM uncertainty analyses primarily focus on scenario and model differences, the results presented here indicate that temporal consistency across successive releases of the same model–scenario pair is equally important for interpretability and credibility, particularly in policy and financial contexts.

The contribution is therefore primarily methodological and interpretive: rather than providing a formal statistical identification of climate-transition uncertainty, the paper develops a structured framework for documenting how NGFS scenario outputs evolve across vintages and for assessing the implications of this evolution for comparability and interpretation. Several limitations should be acknowledged. The analysis is confined to NGFS-published outputs and does not attribute vintage uncertainty to specific underlying mechanisms such as data revisions, parameter recalibration, structural model updates, or changes in exogenous assumptions. Moreover, the proposed uncertainty metrics quantify dispersion but do not assess the plausibility or desirability of individual pathways. Future research could combine the framework developed here with detailed model documentation, version histories, and additional diagnostics to better identify the sources of cross-vintage variation, and could extend the analysis to institution-specific stress tests and firm-level financial outcomes. A fuller econometric identification of these mechanisms would be a valuable extension, but remains beyond the scope of the present analysis and likely requires additional model-level documentation, version histories, and methodological transparency beyond what is currently publicly available.

Overall, the results underscore that NGFS scenarios should be interpreted as evolving exploratory pathways rather than fixed reference trajectories. Explicit recognition of initial-condition and recalibration-related uncertainty—and of their potentially large magnitude for a meaningful subset of variables—is essential for robust interpretation, prudent risk assessment, and informed decision-making in the context of climate-related financial risks.

References

Acharya, V. V., Berner, R., Engle, R., Jung, H., Stroebel, J., Zeng, X., & Zhao, Y. (2023). Climate stress testing. *Annual Review of Financial Economics*, 15, 291–326.

Autorité de Contrôle Prudentiel et de Résolution (ACPR). (2020). *Pilot exercise on climate-related risks: Main lessons learnt*. Banque de France.

Allen, T., Dees, S., Boissinot, J., Caicedo Graciano, C. M., Chouard, V., Clerc, L., de Gaye, A., Devulder, A., Diot, S., Lisack, N., Pegoraro, F., Rabaté, S., Svartzman, R., & Vernet, A. (2020). Climate-related scenarios for financial stability assessment: An application to France. *Banque de France Working Paper*, No. 774.

Allwood, J. M., Cullen, J. M., and Milford, R. L. (2010). Options for achieving a 50% cut in industrial carbon emissions by 2050. *Environmental Science & Technology*, 44(6), 1888–1894.

Baer, M. (2024). *Climate risk scenario analysis and the role of central banks and supervisors in safeguarding financial stability in the net-zero transition.* DPhil thesis, University of Oxford.

Baer, M., Gasparini, M., Lancaster, R., & Ranger, N. (2023). "All scenarios are wrong, but some are useful": Toward a framework for assessing and using climate risk scenarios within financial decisions. *Frontiers in Climate*, 5, 1146402.

Bank of England (2021). Key elements of the 2021 Biennial Exploratory Scenario: Financial risks from climate change. *Bank of England*.

Baudino, P., & Svoronos, J.-P. (2021). Stress-testing banks for climate change: A comparison of practices. *FSI Insights on Policy Implementation*, No. 34, Bank for International Settlements.

Bloom, N. (2009). The impact of uncertainty shocks. *Econometrica*, 77(3), 623–685.

Clements, M. P., Rich, R. W., & Tracy, J. (2025). An investigation into the uncertainty revision process of professional forecasters. *Journal of Economic Dynamics and Control*, 173, 105060.

Croushore, D., & Stark, T. (2001). A real-time data set for macroeconomists. *Journal of Econometrics*, 105(1), 111–130.

Croushore, Dean (2006). Forecasting with Real-Time Macroeconomic Data. In G. Elliott, C. W. J. Granger, and A. Timmermann (eds.), *Handbook of Economic Forecasting*, Vol. 1. Elsevier, pp. 961–982.

Dekker, M. M., Hof, A. F., van den Berg, M., Daioglou, V., van Heerden, R., van der Wijst, K.-I., and van Vuuren, D. P. (2023). Spread in climate policy scenarios unravelled. *Nature*, 624, 309–316.

Dixit, A. K., & Pindyck, R. S. (1994). *Investment under Uncertainty.* Princeton University Press.

Drechsler, I. (2013). Uncertainty, time-varying fear, and asset prices. *Journal of Finance*, 68(5), 1843–1889.

European Central Bank (2022). *Macro-financial scenarios for the 2022 climate risk stress test.* ECB Banking Supervision.

Gidden, M. J., Gasser, T., Grassi, G., Forsell, N., Janssens, I., Lamb, W. F., Minx, J., Nicholls, Z., Steinhauser, J., and Riahi, K. (2023). Aligning climate scenarios to emissions inventories shifts global benchmarks. *Nature*, 624, 102–108.

Gillingham, K., Nordhaus, W., Anthoff, D., Blanford, G., Bosetti, V., Christensen, P., McJeon, H., & Reilly, J. (2018). Modeling uncertainty in integrated assessment of climate change: A multimodel comparison. *Journal of the Association of Environmental and Resource Economics*, 5(4), 791–826.

Grigoli, F., Herman, A., Swiston, A., & Di Bella, G. (2015). Output Gap Uncertainty and Real-Time Monetary Policy. *IMF Working Paper* WP/15/14. International Monetary Fund.

Grossmann, I., Varnum, M. E. W., Hutcherson, C. A., and Mandel, D. R. (2024). When expert predictions fail. *Trends in Cognitive Sciences*, 28(2), 113–123.

Guivarch, C., Lempert, R., & Trutnevyte, E. (2017). Scenario techniques for energy and environmental research: An overview of recent developments to broaden the capacity to deal with complexity and uncertainty. *Environmental Modelling & Software*, 97, 201–210.

Habert, G., Miller, S. A., John, V. M., Provis, J. L., Favier, A., Horvath, A., and Scrivener, K. L. (2020). Environmental impacts and decarbonization strategies in the cement and concrete industries. *Nature Reviews Earth & Environment*, 1, 559–573.

Hansen, L. P., and Sargent, T. J. (2008). *Robustness.* Princeton University Press.

Hawkins, E., and Sutton, R. (2009). The potential to narrow uncertainty in regional climate predictions. *Bulletin of the American Meteorological Society*, 90(8), 1095–1107.

Hawkins, E., and Sutton, R. (2011). The potential to narrow uncertainty in projections of regional precipitation change. *Climate Dynamics*, 37, 407–418.

Hong Kong Monetary Authority (2023). *Climate Risk Stress Test for Authorized Institutions.* Hong Kong Monetary Authority.

Hsiang, S., & Kopp, R. E. (2018). An economist's guide to climate change science. *Journal of Economic Perspectives*, 32(4), 3–32.

International Energy Agency (2018). *Technology Roadmap: Low-Carbon Transition in the Cement Industry.* IEA, Paris.

International Energy Agency (2020). *Iron and Steel Technology Roadmap.* IEA, Paris.

IFRS Foundation. (2024). *Progress on Corporate Climate-Related Disclosures.*

International Sustainability Standards Board (ISSB) (2023). IFRS S2: Climate-related Disclosures. IFRS Foundation.

Keppo, I., Butnar, I., Bauer, N., Caspani, M., Edelenbosch, O., Emmerling, J., Fragkos, P., Guivarch, C., Harmsen, M., Lefèvre, J., Le Gallic, T., Leimbach, M., McDowall, W., Mercure, J.-F., Schaeffer, R., Trutnevyte, E., & Wagner, F. (2021). Exploring the possibility space: Taking stock of the diverse capabilities and gaps in integrated assessment models. *Environmental Research Letters*, 16(5), 053006.

Knight, F. H. (1921). *Risk, Uncertainty and Profit.* Boston: Houghton Mifflin.

Krey, V., Guo, F., Kolp, P., Zhou, W., Schaeffer, R., Awasthy, A., Bertram, C., de Boer, H.-S., Fragkos, P., Fujimori, S., He, C., Iyer, G., Keramidas, K., Koberle, A. C., Oshiro, K., Reis, L. A., Shoai-Tehrani, B., Vishwanathan, S., Capros, P., Drouet, L., Edmonds, J. E., Garg, A., Gernaat, D. E. H. J., Jiang, K., Kannavou, M., Kitous, A., Kriegler, E., Luderer, G., Mathur, R., Muratori, M., Sano, F., & van Vuuren, D. P. (2019). Looking under the hood: A comparison of techno-economic assumptions across national and global integrated assessment models. *Energy*, 172, 1254–1267.

Kriegler, E., Weyant, J. P., Blanford, G. J., Krey, V., Clarke, L., Edmonds, J., Fawcett, A., Luderer, G., Riahi, K., Richels, R., Rose, S. K., Tavoni, M., & van Vuuren, D. P. (2014). The role of technology for achieving climate policy objectives: Overview of the EMF 27 study on global technology and climate policy strategies. *Climatic Change*, 123(3–4), 353–367.

Kwakkel, J. H., Walker, W. E., & Haasnoot, M. (2016). Coping with the wickedness of public policy problems: Approaches for decision making under deep uncertainty. *Journal of Water Resources Planning and Management*, 142(3), 01816001.

Meiler, S., Kropf, C. M., McCaughey, J. W., Lee, C.-Y., Camargo, S. J., Sobel, A. H., Bloemendaal, N., Emanuel, K., & Bresch, D. N. (2025). Navigating and attributing uncertainty in future tropical cyclone risk estimates. *Science Advances*, 11, eadn4607.

Network for Greening the Financial System. (2023). *NGFS Climate Scenarios: Guidance on the Use of Scenarios*. NGFS.

Network for Greening the Financial System (NGFS). (2024). *NGFS Climate Scenarios Technical Documentation, Version 5.0*. November 2024. Paris: Network for Greening the Financial System.

Noels, J., Pouille, C., Jachnik, R., & Rocha, M. (2023). Climate change mitigation scenarios for financial sector target setting and alignment assessment: A stocktake and analysis of their Paris-consistency, practicality and assumptions. *OECD Environment Working Papers*, No. 223. Organisation for Economic Co-operation and Development.

Riahi, K., Bertram, C., Huppmann, D., Rogelj, J., Bosetti, V., Cabardos, A.-M., Deppermann, A., Drouet, L., Frank, S., Fricko, O., Fujimori, S., Harmsen, M., Hasegawa, T., Krey, V., Luderer, G., Paroussos, L., Schaeffer, R., Weitzel, M., van der Zwaan, B., Vrontisi, Z., Dalla Longa, F., Després, J., Fosse, F., Fragkiadakis, K., Gusti, M., Humpenöder, F., Keramidas, K., Kishimoto, P., Kriegler, E., Meinshausen, M., Nogueira, L. P., Oshiro, K., Popp, A., Rochedo, P. R. R., Ünlü, G., van Ruijven, B., Takakura, J., Tavoni, M., van Vuuren, D., and Zakeri, B. (2021). Cost and attainability of meeting stringent climate targets without overshoot. *Nature Climate Change*, 11, 1063–1069.

Rogelj, J., Shindell, D., Jiang, K., Fifita, S., Forster, P., Ginzburg, V., Handa, C., Kheshgi, H., Kobayashi, S., Kriegler, E., Mundaca, L., Séférian, R., & Vilariño, M. V. (2018). Mitigation pathways compatible with 1.5°C in the context of sustainable development. *IPCC Special Report on Global Warming of 1.5°C.*

Schwanitz, V. J. (2013). Evaluating integrated assessment models of global climate change. *Environmental Modelling & Software*, 50, 120–131.

Tang, K., Cervenka, J., Buller, A., Gallice, B., Jindal, A., Bär, M., and Shrimali, G. (2024). *Impact of Climate Scenario Choices on Climate Financial Risk Assessment: High Variance Company Performance Under Consistent Climate Stress Test.* Oxford Sustainable Finance Group, Smith School of Enterprise and the Environment, University of Oxford Working Paper.

Task Force on Climate-related Financial Disclosures. *Guidance on Scenario Analysis for Non-Financial Companies.* Financial Stability Board, 2020.

Tebaldi, C., Nicholls, Z. R. J., Friedlingstein, P., Gasser, T., and Knutti, R. (2022). STITCHES: creating new climate model projections by combining existing simulations. *Earth System Dynamics*, 13, 1557–1609.

van Asselt, M. B. A., & Rotmans, J. (2002). Uncertainty in integrated assessment modelling: From positivism to pluralism. *Climatic Change*, 54, 75–105.

van Beek, L., Hajer, M., Pelzer, P., van Vuuren, D., & Cassen, C. (2020). Anticipating futures through models: the rise of Integrated Assessment Modelling in the climate science-policy interface since 1970. *Global Environmental Change*, 65, 102191.

Wachtmeister, H., Henke, V., Höglund-Isaksson, L., Lundgren, J., & Lechtenböhmer, S. (2018). Revisiting the long-term projections of the IEA World Energy Outlook. *Energy Strategy Reviews*, 22, 195–204.

Walker, W. E., Harremoës, P., Rotmans, J., van der Sluijs, J. P., van Asselt, M. B. A., Janssen, P., & Krayer von Krauss, M. P. (2003). Defining uncertainty: A conceptual basis for uncertainty management in model-based decision support. *Integrated Assessment*, 4(1), 5–17.

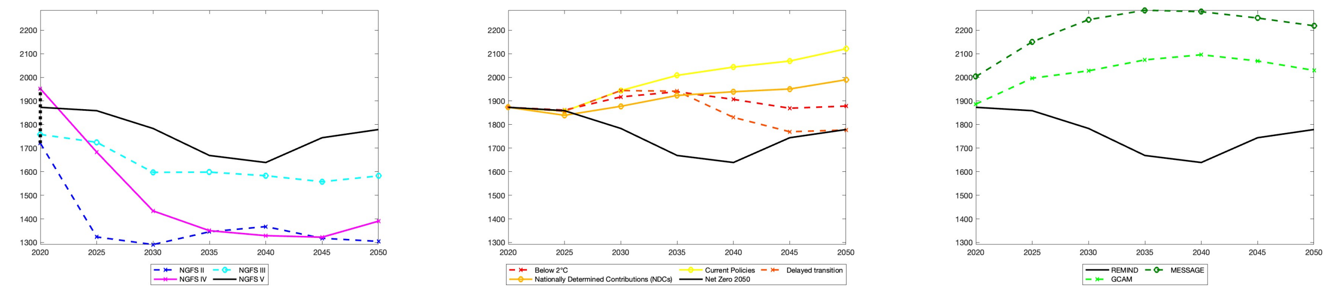
Wilkinson, D., Baer, M., Cormack, C., Köberle, A. C., Lowe, J., & Thompson, E. (2025). *A Climate Scenario Taxonomy for the Financial Sector*. Discussion Paper. UK Centre for Greening Finance and Investment.

Wilson, C., Guivarch, C., Kriegler, E., van Ruijven, B., van Vuuren, D. P., Krey, V., Schwanitz, V. J., & Thompson, E. L. (2021). Evaluating process-based integrated assessment models of climate change mitigation. *Climatic Change*, 166, 3.

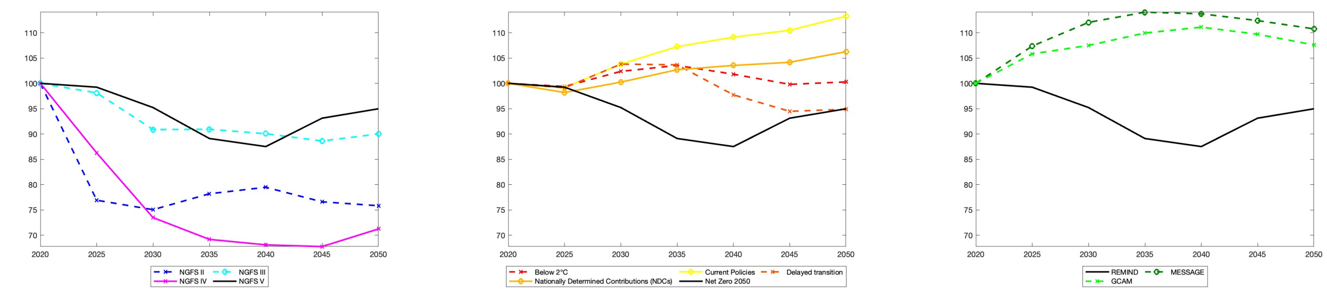
Figure 2: Global steel production pathways across vintages, scenarios and models (REMIND-Net Zero)

Vintages | Scenarios | Models

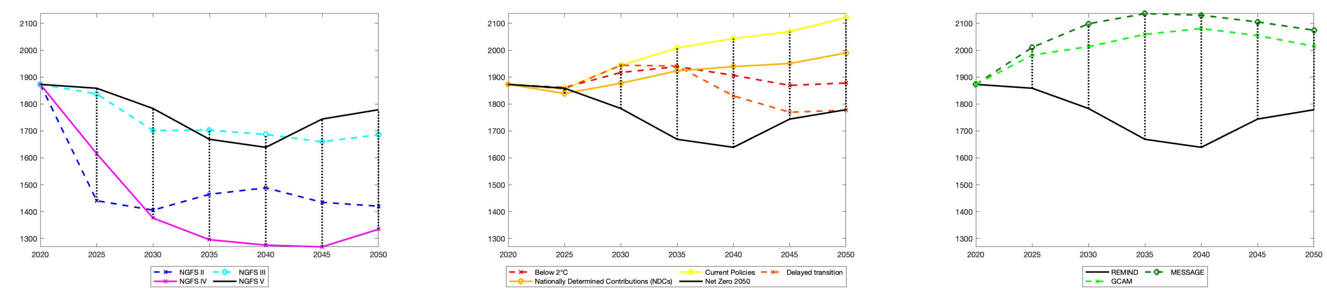
Panel A. Raw variable paths



Panel B. Variable paths indexed to 100

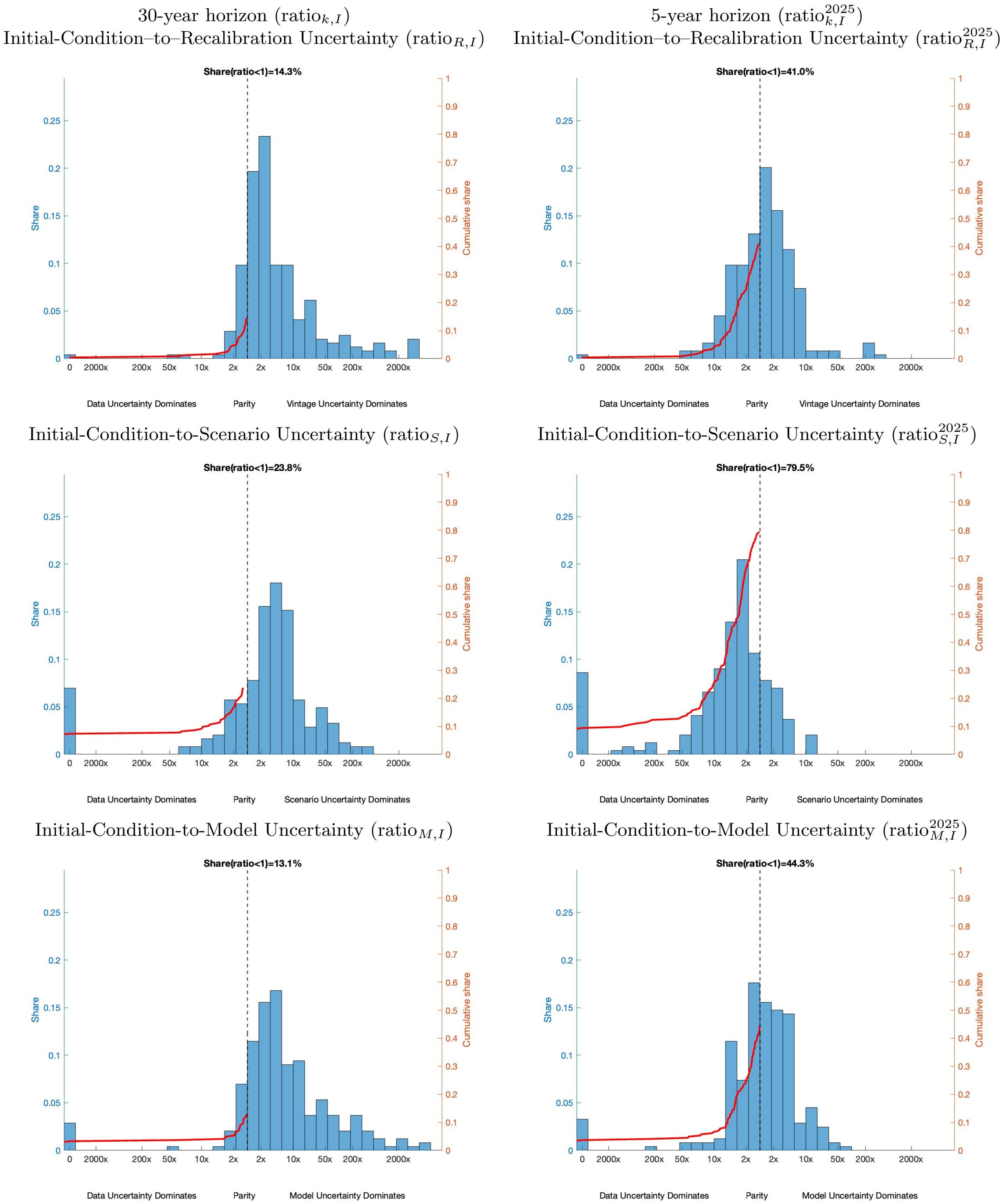


Panel C. Variable paths using a common starting point



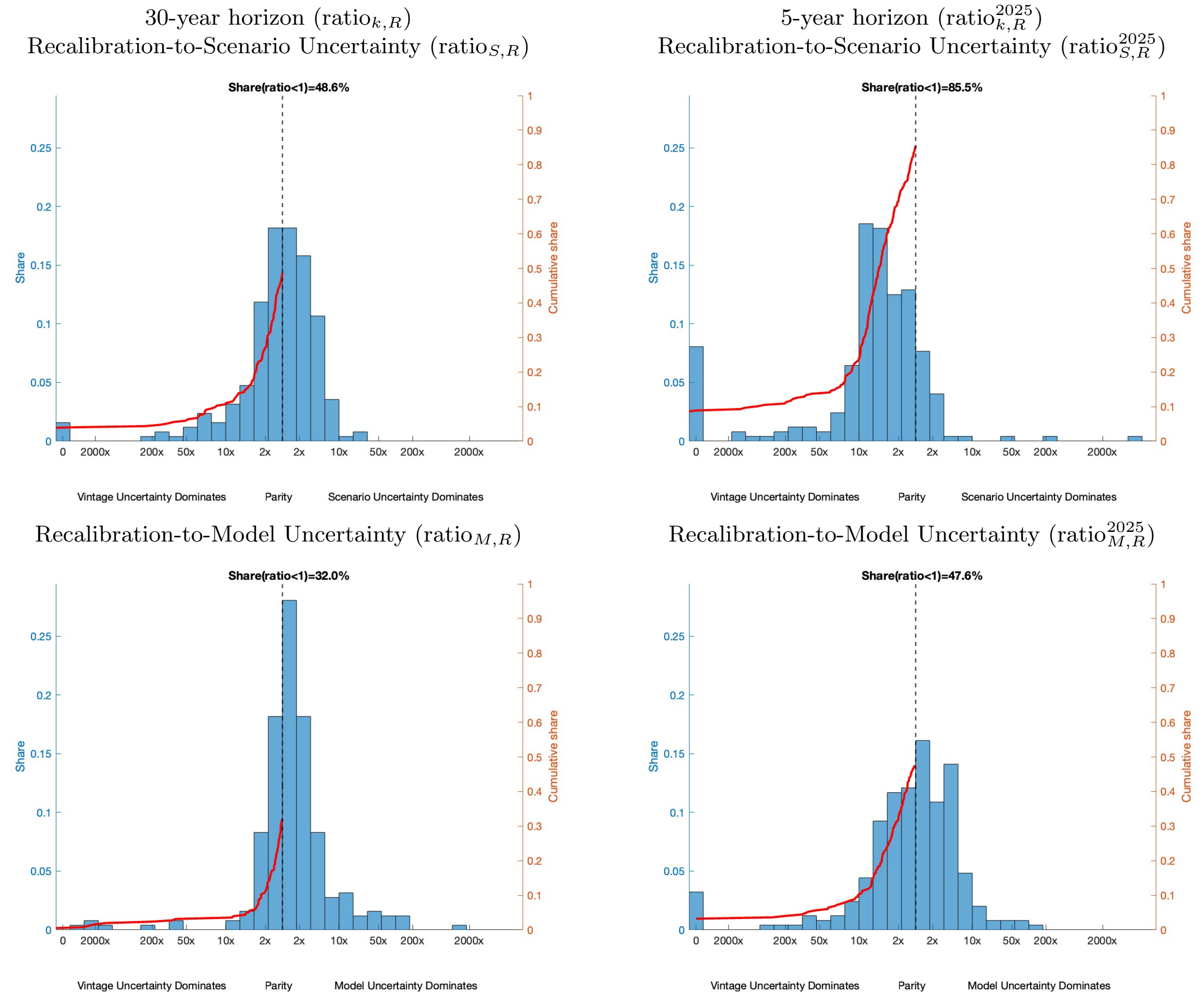
Notes: The figure illustrates the construction of initial condition and path-based uncertainty metrics using global steel production. Panel A shows raw variable paths of steel production (Million ton/year) across vintages, scenarios, and models. For a fixed model and scenario, dispersion in reported values at the base year 2020 reflects initial condition uncertainty. The broken vertical line in the vintages column at 2020 represents the distance between the maximum and minimum reported starting values across NGFS vintages and corresponds directly to the initial condition uncertainty measure. Panel B shows the same paths indexed to 100 at their respective starting values, removing level differences and isolating path-based variation. Panel C aligns all indexed paths to a common reference starting point. At each time step, broken vertical line segments indicate the distance between the contemporaneous minimum and maximum values across pathways, holding the remaining dimensions fixed. The sequence of these extrema defines the maximum and minimum composite paths used to compute path-based uncertainty metrics. NGFS variable name: Production|Steel

Figure 3: Initial-condition uncertainty ratio distribution



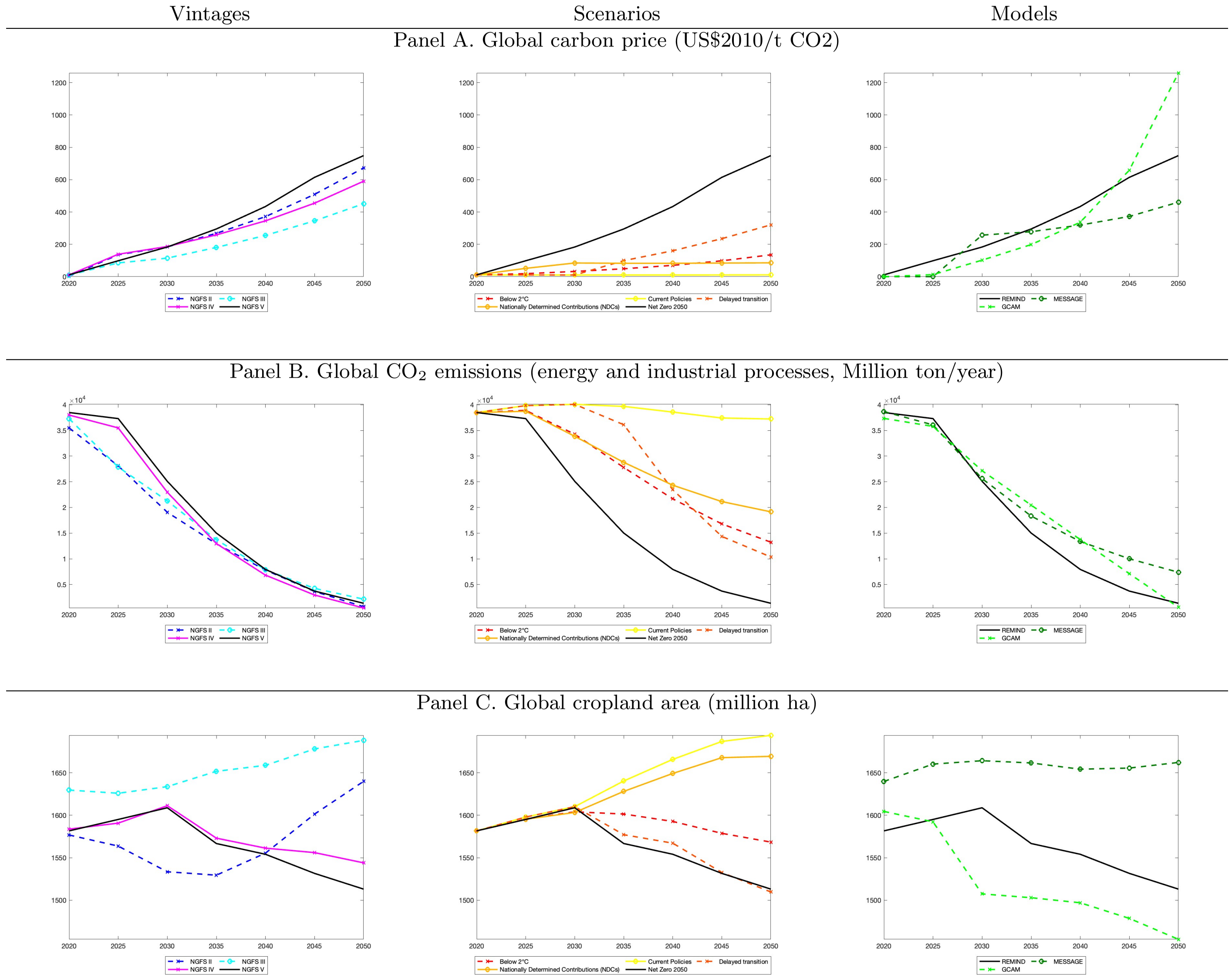
Notes: The figure shows the distribution of initial condition uncertainty ratios, defined as the ratio of path-based uncertainty (scenario, recalibration, or model) to initial condition uncertainty. Ratios are displayed on a base-10 logarithmic scale, i.e. as $\log_{10}$ uncertainty ratios. The leftmost bin at zero corresponds to variables with exactly zero path-based uncertainty and represents a structural mass point rather than dominance. Values below unity (negative $\log_{10}$ values) indicate that initial condition uncertainty dominates, while values above unity (positive $\log_{10}$ values) indicate dominance of path-based uncertainty. Tick labels report selected dominance magnitudes (e.g. $2\times$, $10\times$, $200\times$, $\ldots$), shown symmetrically around parity; dominance direction is indicated by the lower-axis labels. The dashed vertical line marks parity (ratio $=1$, $\log_{10}=0$). The red line overlays the empirical cumulative distribution function (CDF), constructed from the same $\log_{10}$ ratios and plotted only up to parity; its endpoint therefore reports the cumulative share of observations for which initial condition uncertainty dominates. Reported values correspond to the cumulative share of observations with ratios below unity, as read from the empirical cumulative distribution function evaluated at parity.

Figure 4: Recalibration uncertainty ratio distribution



Notes: The figure shows the distribution of recalibration uncertainty ratios, defined as the ratio of scenario or model uncertainty to recalibration uncertainty. Ratios are displayed on a base-10 logarithmic scale, i.e. as $\log_{10}$ uncertainty ratios. The leftmost bin at zero corresponds to variables with exactly zero scenario or model uncertainty and represents a structural mass point rather than dominance. Values below unity (negative $\log_{10}$ values) indicate that recalibration uncertainty dominates, while values above unity (positive $\log_{10}$ values) indicate dominance of scenario or model uncertainty. Tick labels report selected dominance magnitudes (e.g. $2\times$, $10\times$, $200\times$, ...), shown symmetrically around parity; dominance direction is indicated by the lower-axis labels. The dashed vertical line marks parity (ratio = 1, $\log_{10} = 0$). The red line overlays the empirical cumulative distribution function (CDF), constructed from the same $\log_{10}$ ratios and plotted only up to parity; its endpoint therefore reports the cumulative share of observations for which recalibration uncertainty dominates. Reported values correspond to the cumulative share of observations with ratios below unity, as read from the empirical cumulative distribution function evaluated at parity.

Figure 5: Raw variable paths: Global carbon price; Global CO_2 emissions (energy and industrial processes); Global cropland area

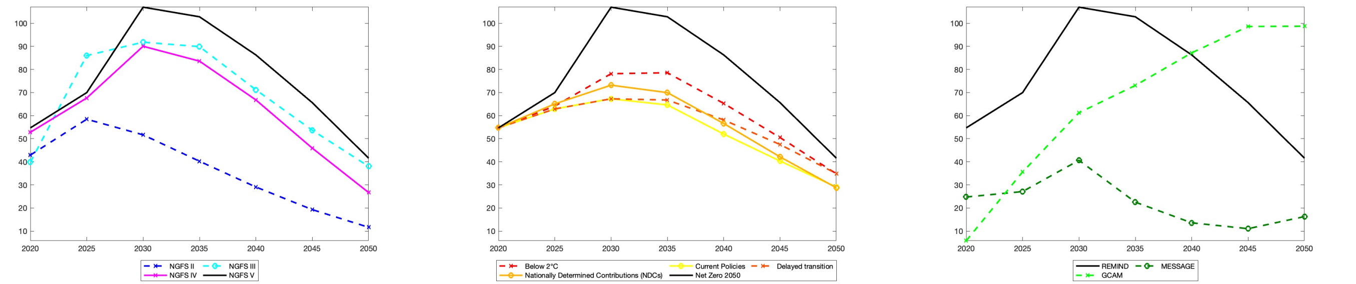


Notes: The figure reports raw variable paths across vintages, scenarios, and models. Values are shown in levels and are not indexed. Dispersion therefore reflects both differences in reported historical starting values and divergence in projected dynamics. This presentation corresponds to Panel A of Figure 2 and is intended to illustrate uncertainty prior to any transformation or alignment of paths. NGFS variable names: Emissions|CO2|Energy and Industrial Processes, Price|Carbon, Land Cover|Cropland

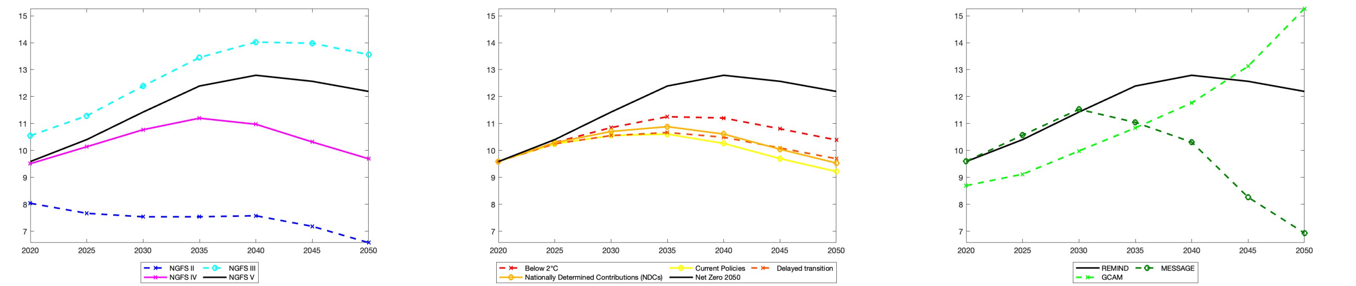
Figure 6: Raw variable paths: Global nuclear electricity investment; Global nuclear electricity generation; Global solar electricity investment

Vintages | Scenarios | Models

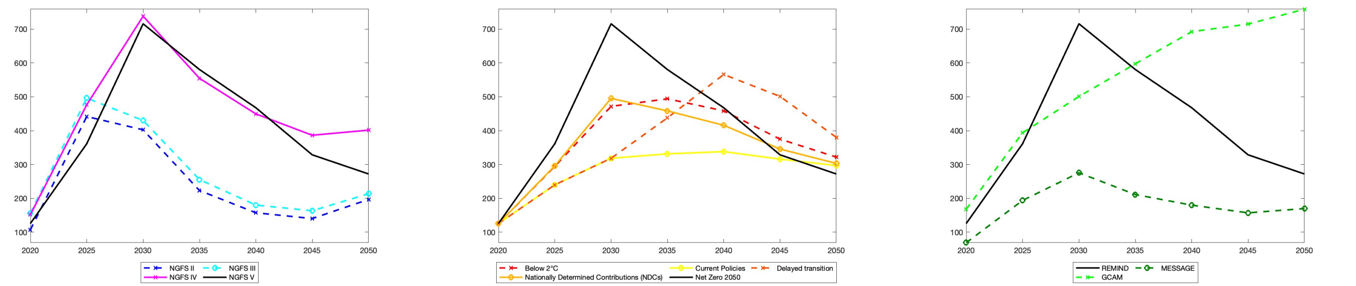
Panel A. Global nuclear electricity investment (billion US$2010/yr)



Panel B. Global nuclear electricity generation (Exa joule/year)



Panel C. Global solar electricity investment (billion US$2010/yr)

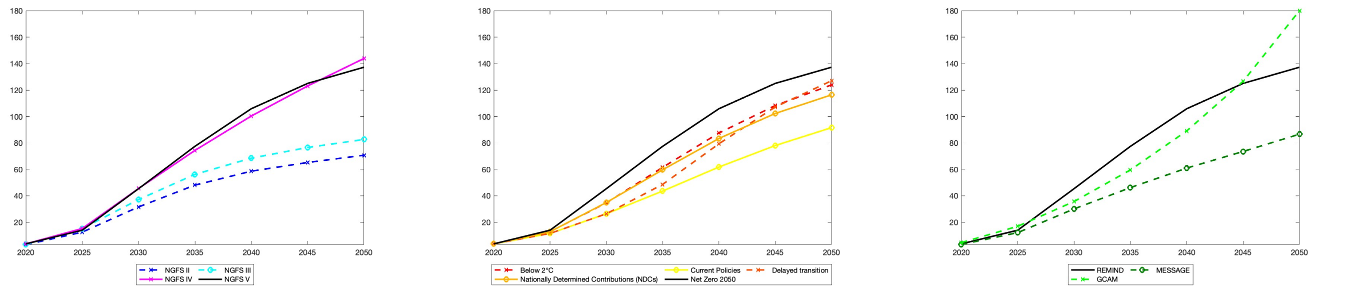


Notes: The figure shows raw variable paths across vintages, scenarios, and models. Paths are presented in levels, without indexation or alignment to a common starting point. As in Panel A of Figure 2, dispersion reflects both Data uncertainty and differences in path-based dynamics. These raw paths provide context for the indexed and path-based uncertainty measures introduced in Section 3. NGFS variable names: Investment|Energy Supply|Electricity|Nuclear, Secondary Energy|Electricity|Nuclear and Investment|Energy Supply|Electricity|Solar

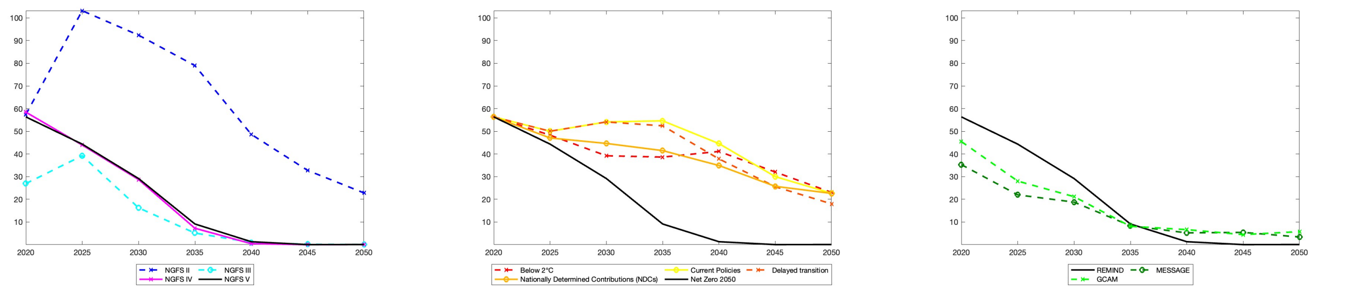
Figure 7: Raw variable paths: Global solar PV electricity generation; Global gas electricity investment; Global gas electricity generation

Vintages	Scenarios	Models

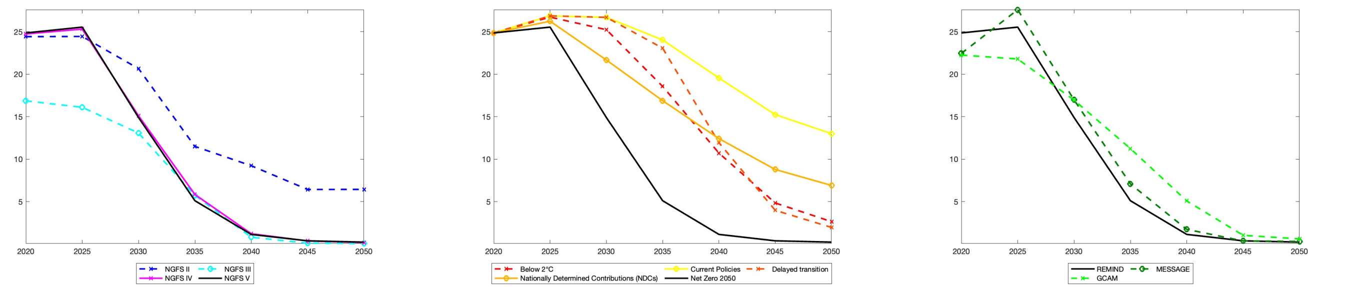
Panel A. Global solar PV electricity generation (Exa joule/year)



Panel B. Global gas electricity investment (billion US$2010/yr)



Panel C. Global gas electricity generation (Exa joule/year)

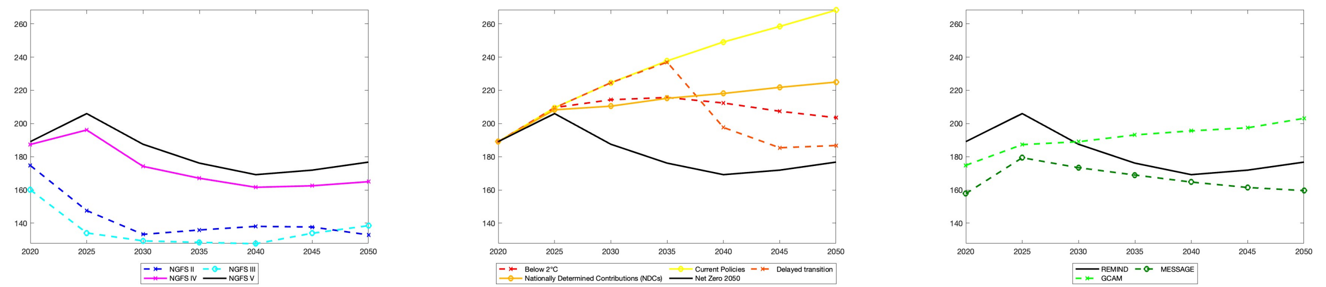


Notes: The figure reports raw variable paths across vintages, scenarios, and models. No indexation or normalization is applied. As in Panel A of Figure 2, variation at the base year reflects initial condition uncertainty, while divergence beyond the base year reflects differences in modeled dynamics. These figures illustrate the magnitude and structure of dispersion prior to constructing path-based uncertainty metrics. NGFS variable names: Secondary Energy|Electricity|Solar|PV, Investment|Energy Supply|Electricity|Gas and Secondary Energy|Electricity|Gas

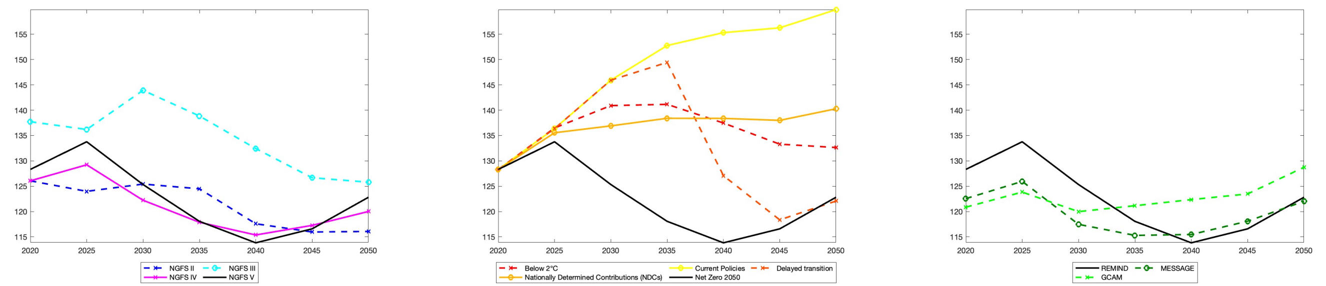
Figure 8: Raw variable paths: Global final energy demand in industry; Global final energy demand in residential and commercial; Global cement production

Vintages | Scenarios | Models

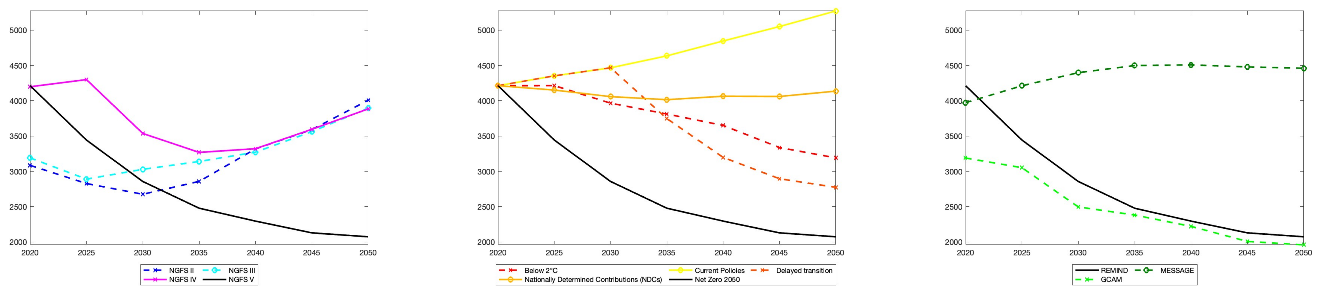
Panel A. Global final energy demand in industry (Exa joule/year)



Panel B. Global final energy demand in residential and commercial (Exa joule/year)



Panel C. Global cement production (Million ton/year)



Notes: The figure displays raw variable paths across vintages, scenarios, and models in levels. The presentation mirrors Panel A of Figure 2 and therefore includes both initial condition differences and path-based variation. These raw paths serve as a descriptive complement to the indexed and aligned representations used to quantify initial condition and path-based uncertainty in Section 3. NGFS variable names: Final Energy|Industry, Final Energy|Residential and Commercial and Production|Cement

Table 1: Initial-condition uncertainty relative to path-based uncertainty across NGFS variables

Variable	**Recalibration** ($\text{ratio}_{R,I}$)	**Scenario** ($\text{ratio}_{S,I}$)	**Model** ($\text{ratio}_{M,I}$)
Global carbon price	Recalibration dominates (+49.9×, +244×)	Scenario dominates (+10.7×, +90.2×)	Model dominates (+12×, +370×)
Global CO_2 emissions	Recalibration dominates (+2.11×, +2.11×)	Mixed (−2.47×, +9.44×)	Mixed (−5.44×, +2.15×)
Global steel production	Recalibration dominates (+1.81×, +2.06×)	Mixed (−10.5×, +1.75×)	Mixed (−1.52×, +2.13×)
Global cement production	Mixed (−1.29×, +3.01×)	Mixed (−1.24×, +2.84×)	Mixed (−1.10×, +2.36×)
Global nuclear electricity investment	Recalibration dominates (+3.24×, +4.77×)	Mixed (−2.10×, +2.67×)	Model dominates (+18.5×, +60.8×)
Global nuclear electricity generation (secondary energy)	Mixed (−1.99×, +1.79×)	Mixed (−15.9×, +1.19×)	Mixed (−5×, +3.95×)
Global solar electricity investment	Recalibration dominates (+3.22×, +7.39×)	Scenario dominates (+2.49×, +8.14×)	Model dominates (+1.36×, +6.97×)
Global solar PV electricity generation (secondary energy)	Recalibration dominates (+5.41×, +62.6×)	Scenario dominates (+4×, +77.3×)	Mixed (−1.02×, +70.4×)
Global gas electricity investment	Recalibration dominates (+1.87×, +2.24×)	Mixed (−5.53×, +1.44×)	Initial-condition dominates (−3.29×, −3.29×)
Global gas electricity generation (secondary energy)	Mixed (−4.42×, +1.03×)	Mixed (−6.04×, +2.37×)	Initial-condition dominates (−1.29×, −1.07×)
Global final energy demand (industry)	Recalibration dominates (+1.64×, +1.64×)	Mixed (−8×, +3.15×)	Mixed (−2.34×, +1.48×)
Global final energy demand (residential and commercial)	Initial-condition dominates (−1.54×, −1.08×)	Mixed (−4.27×, +3.56×)	Mixed (−5.08×, +1.38×)
Global cropland area	Mixed (−1.99×, +2.47×)	Mixed (−18.8×, +3.49×)	Mixed (−1.66×, +3.22×)

Notes: Each cell reports a qualitative dominance classification (top line) and the corresponding signed dominance factors (bottom line) in the form (ratio^{2025}, ratio^{2050}). Positive values indicate that the path-based uncertainty in the numerator (Recalibration, scenario, or model) exceeds initial-condition uncertainty by the reported factor, while negative values indicate dominance of initial-condition uncertainty. Figure 4 reports the same ratios in $\log_{10}$ space; the multiplicative factors shown here are obtained by exponentiating the corresponding $\log_{10}$ ratios and retaining the dominance sign.

Table 2: Recalibration uncertainty relative to scenario and model uncertainty across NGFS variables

Variable	**Scenario** ($\text{ratio}_{S,R}$)	**Model** ($\text{ratio}_{M,R}$)
Global carbon price	Recalibration dominates (−4.66×, −2.66×)	Mixed (−4.15×, +1.18×)
Global CO_2 emissions	Mixed (−5.22×, +5.76×)	Mixed (−11.5×, +1.38×)
Global steel production	Recalibration dominates (−18.9×, −1.44×)	Recalibration dominates (−2.75×, −1.17×)
Global cement production	Scenario dominates (+1.04×, +1.45×)	Model dominates (+1.17×, +1.01×)
Global nuclear electricity investment	Recalibration dominates (−6.80×, −1.79×)	Model dominates (+5.72×, +12.6×)
Global nuclear electricity generation (secondary energy)	Recalibration dominates (−8×, −1.51×)	Mixed (−2.51×, +1.44×)
Global solar electricity investment	Mixed (−1.30×, +1.10×)	Recalibration dominates (−2.37×, −1.17×)
Global solar PV electricity generation (secondary energy)	Mixed (−1.35×, +1.24×)	Recalibration dominates (−5.49×, −1.11×)
Global gas electricity investment	Recalibration dominates (−10.4×, −1.52×)	Recalibration dominates (−6.13×, −7.39×)
Global gas electricity generation (secondary energy)	Mixed (−1.37×, +2.35×)	Mixed (+3.43×, −1.33×)
Global final energy demand (industry)	Mixed (−13.1×, +1.94×)	Mixed (−3.83×, +1×)
Global final energy demand (residential and commercial)	Mixed (−2.78×, +3.82×)	Mixed (−3.30×, +1.40×)
Global cropland area	Mixed (−9.43×, +1.41×)	Model dominates (+1.20×, +1.52×)

Notes: Each cell reports a qualitative dominance classification (top line) and the corresponding signed dominance factors (bottom line) in the form (ratio^{2025}, ratio^{2050}), evaluated horizon-by-horizon relative to unity. Vintage dominates corresponds to $(-,-)$, Scenario/Model dominates to $(+,+)$, and Mixed to $(+,-)$ or $(-,+)$.
Figure 3 reports the same ratios in $\log_{10}$ space; the multiplicative factors shown here are obtained by exponentiating the corresponding $\log_{10}$ ratios and retaining the dominance sign.

The Smith School of Enterprise and the Environment (SSEE)

SSEE was established with a benefaction by the Smith family in 2008 to tackle major environmental challenges by bringing public and private enterprise together with the University of Oxford's world-leading teaching and research.

Research at the Smith School shapes business practices, government policy and strategies to achieve net zero emissions and sustainable development. We offer innovative evidence-based solutions to the environmental challenges facing humanity over the coming decades. We apply expertise in economics, finance, business, and law to tackle environmental and social challenges in six areas: water, climate, energy, biodiversity, food, and the circular economy.

SSEE has several significant external research partnerships and Business Fellows, bringing experts from industry, consulting firms, and related enterprises who seek to address major environmental challenges to the University of Oxford. We offer a variety of open enrolment and custom Executive Education programmes that cater to participants from all over the world. We also provide independent research and advice on environmental strategy, corporate governance, public policy, and long-term innovation.

For more information on SSEE please visit: www.smithschool.ox.ac.uk

Oxford Sustainable Finance Group

Oxford Sustainable Finance Group is a world-leading, multi-disciplinary centre for research and teaching in sustainable finance. We are uniquely placed by virtue of our scale, scope, networks, and leadership to understand the key challenges and opportunities in different contexts, and to work with partners to ambitiously shape the future of sustainable finance.

Aligning finance with sustainability to tackle global environmental and social challenges.

Both financial institutions and the broader financial system must manage the risks and capture the opportunities of the transition to global environmental sustainability. The University of Oxford has world leading researchers and research capabilities relevant to understanding these challenges and opportunities.

Established in 2012, the Oxford Sustainable Finance Group is the focal point for these activities.

The Group is multi-disciplinary and works globally across asset classes, finance professions, and with different parts of the financial system. We are the largest such centre globally and are working to be the world's best place for research and teaching on sustainable finance and investment. The Oxford Sustainable Finance Group is part of the Smith School of Enterprise and the Environment at the University of Oxford.

For more information please visit: sustainablefinance.ox.ac.uk/group

Appendix

A NGFS Background

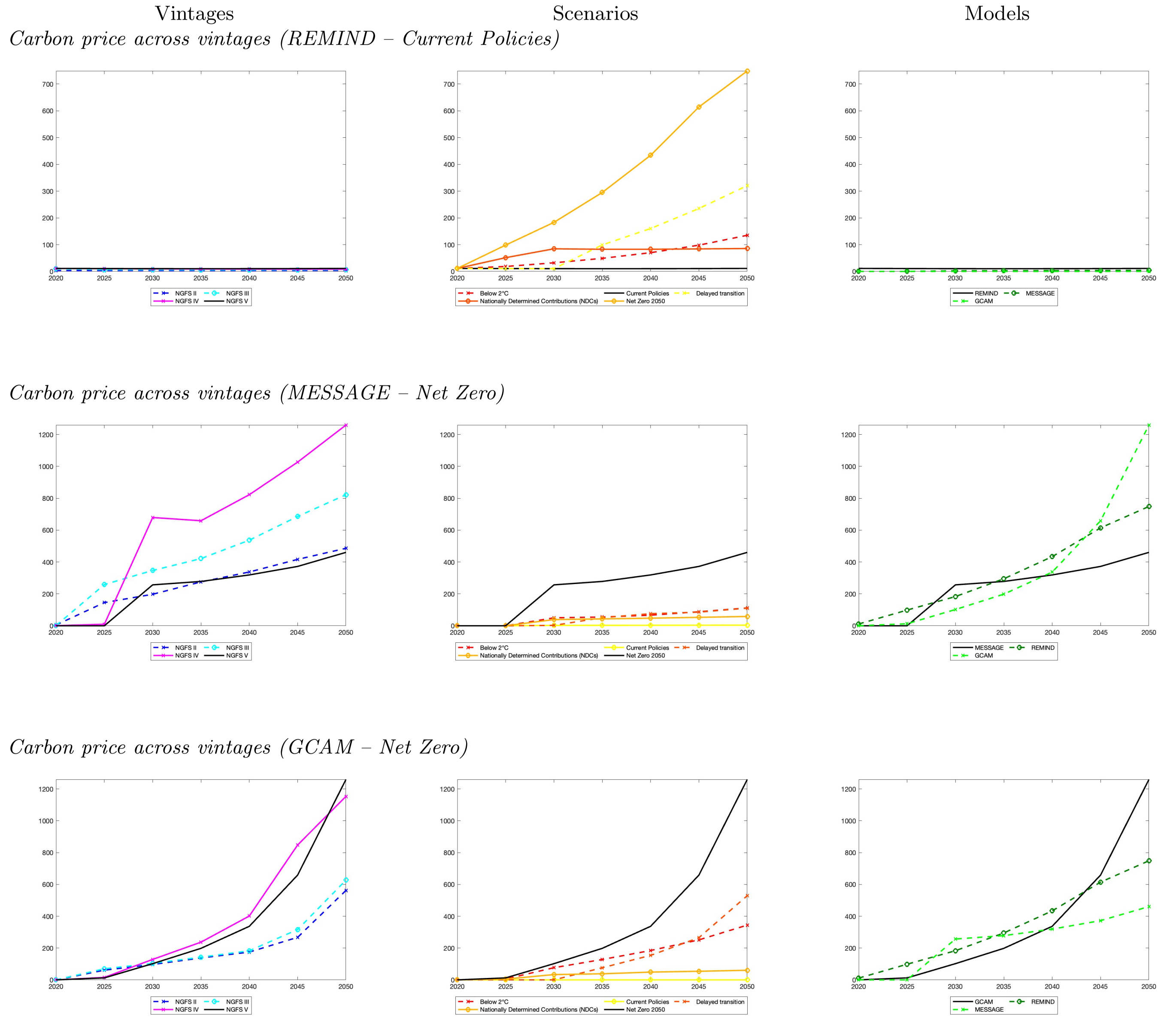
Table A1: Overview of NGFS Integrated Assessment Models (IAMs) used in this study

IAM (acronym spelled out)	Institutional home	High-level structure	Outputs most relevant for stress testing
REMIND–MAgPIE (**R**egional **E**nergy **M**odel of **IN**vestments and **D**evelopment coupled with **M**odel of **Ag**ricultural **P**roduction and its **I**mpact on the **E**nvironment)	Potsdam Institute for Climate Impact Research (PIK) and collaborators	Intertemporal energy–economy optimization model coupled to a global land-use model. Energy system, technology choice, and macroeconomic dynamics are linked iteratively with land-use, agriculture, and bioenergy pathways.	Carbon prices; CO_2 emissions; energy supply and investment by technology; electricity capacity and generation; final energy demand by sector; land-use variables (e.g. cropland).
MESSAGEix–GLOBIOM (**M**odel for **E**nergy **S**upply **S**trategy **A**lternatives and their **G**eneral **E**nvironmental **E**ffect linked with **Gl**obal **L**and-use **O**ptimization **B**iosphere **M**anagement Model)	International Institute for Applied Systems Analysis (IIASA) and collaborators	Global energy system optimization model (MESSAGEix) linked to a detailed land-use, agriculture, and forestry model. Emphasizes least-cost mitigation pathways under policy and resource constraints.	Carbon prices; CO_2 emissions; electricity capacity and investment; fuel production; energy demand by sector; bioenergy and land-use related outputs (where reported).
GCAM (**G**lobal **C**hange **A**nalysis **M**odel)	Pacific Northwest National Laboratory (PNNL) and collaborators	Integrated partial-equilibrium framework linking energy, economy, agriculture, land use, and climate systems. Market equilibrium is solved across regions and sectors with endogenous technology substitution.	CO_2 emissions; energy supply mix; electricity capacity and generation; final energy demand; land-use variables; technology deployment trajectories relevant for transition risk analysis.

Notes: The NGFS publishes scenario "vintages" annually. For each vintage, scenarios are generated by the modelling teams using updated initial condition, calibration choices, and model versions. The table provides a high-level orientation for non-IAM readers and is not intended as a complete technical description of each model.

B Additional Results

Figure A1: Carbon price across vintages (REMIND – Current Policies; MESSAGE – Net Zero; GCAM – Net Zero)

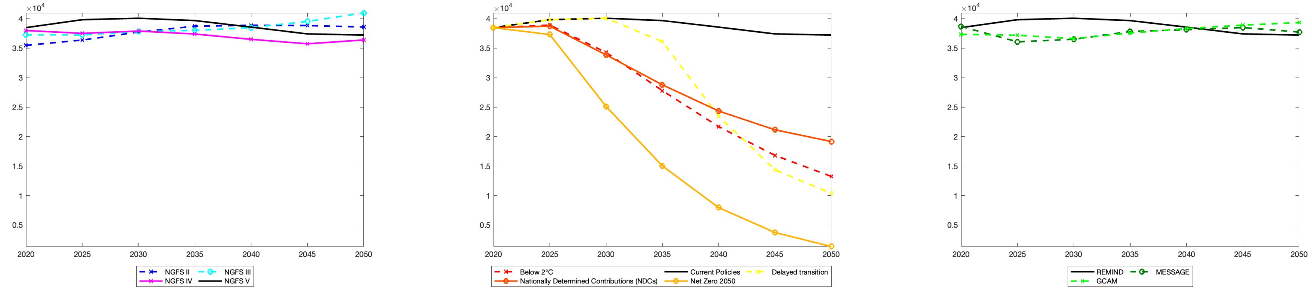


Notes: The figure reports raw variable paths across NGFS scenario vintages, scenarios, and integrated assessment models (IAMs). Paths are shown in levels and are not indexed or aligned to a common starting point. Dispersion at the base year (2020) therefore reflects *initial condition uncertainty*, arising from differences in reported historical starting values across NGFS vintages. Divergence beyond the base year reflects a combination of recalibration uncertainty, scenario uncertainty, and model uncertainty as defined in Section 3. These figures correspond to Panel A of the framework illustrated in Figures 3–7 and are intended as descriptive complements to the indexed and path-based uncertainty metrics reported in Tables 1–2.

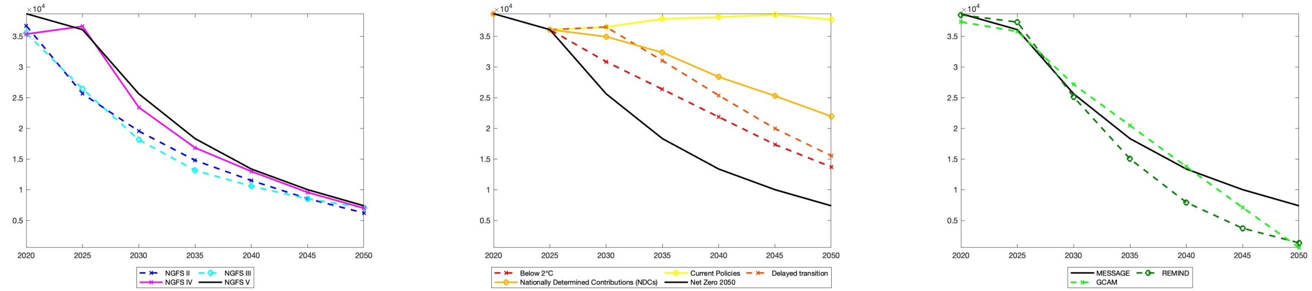
Figure A2: CO_2 emissions (energy and industrial processes) across vintages (REMIND – Current Policies; MESSAGE – Net Zero; GCAM – Net Zero)

Vintages | Scenarios | Models

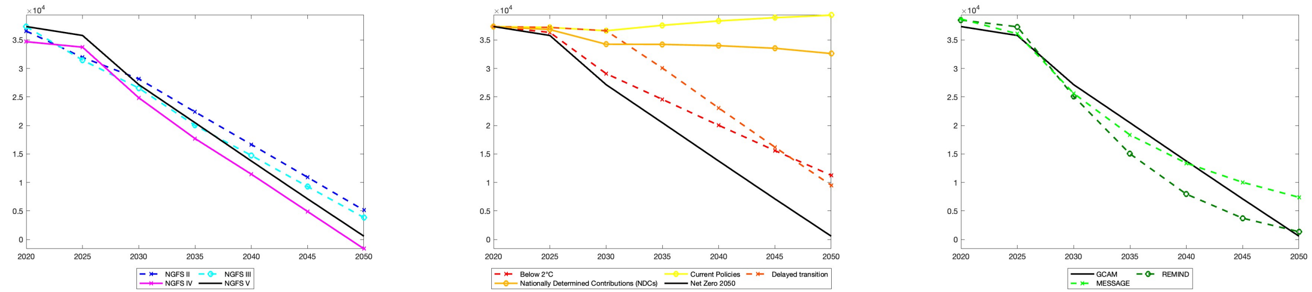
CO_2 emissions (energy and industrial processes) across vintages (REMIND – Current Policies)



CO_2 emissions (energy and industrial processes) across vintages (MESSAGE – Net Zero)



CO_2 emissions (energy and industrial processes) across vintages (GCAM – Net Zero)

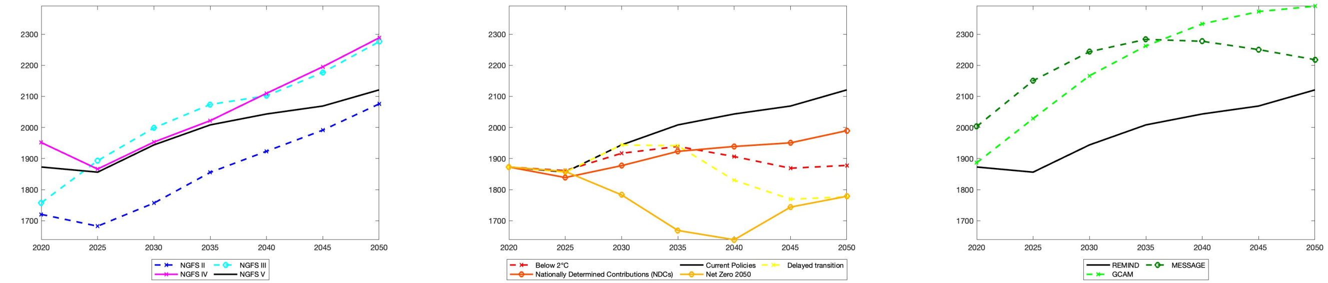


Notes: The figure reports raw variable paths across NGFS scenario vintages, scenarios, and integrated assessment models (IAMs). Paths are shown in levels and are not indexed or aligned to a common starting point. Dispersion at the base year (2020) therefore reflects *initial condition uncertainty*, arising from differences in reported historical starting values across NGFS vintages. Divergence beyond the base year reflects a combination of recalibration uncertainty, scenario uncertainty, and model uncertainty as defined in Section 3. These figures correspond to Panel A of the framework illustrated in Figures 3–7 and are intended as descriptive complements to the indexed and path-based uncertainty metrics reported in Tables 1–2.

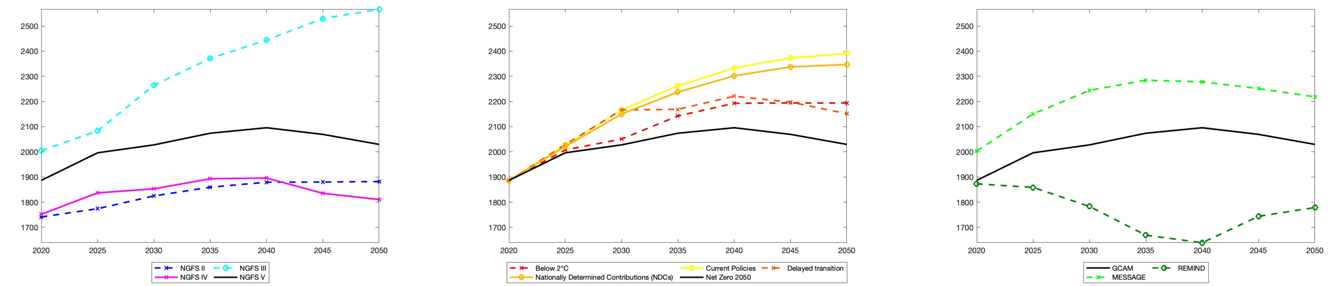
Figure A3: Global steel production across vintages (REMIND – Current Policies; MESSAGE – Net Zero; GCAM – Net Zero)

Vintages | Scenarios | Models

Global steel production across vintages (REMIND – Current Policies)



Global steel production across vintages (GCAM – Net Zero)

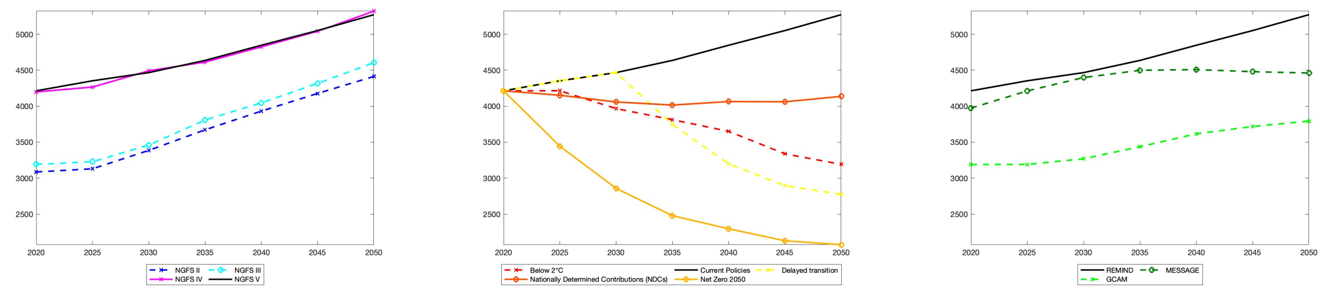


Notes: The figure reports raw variable paths across NGFS scenario vintages, scenarios, and integrated assessment models (IAMs). Paths are shown in levels and are not indexed or aligned to a common starting point. Dispersion at the base year (2020) therefore reflects *initial condition uncertainty*, arising from differences in reported historical starting values across NGFS vintages. Divergence beyond the base year reflects a combination of recalibration uncertainty, scenario uncertainty, and model uncertainty as defined in Section 3. These figures correspond to Panel A of the framework illustrated in Figures 3–7 and are intended as descriptive complements to the indexed and path-based uncertainty metrics reported in Tables 1–2.

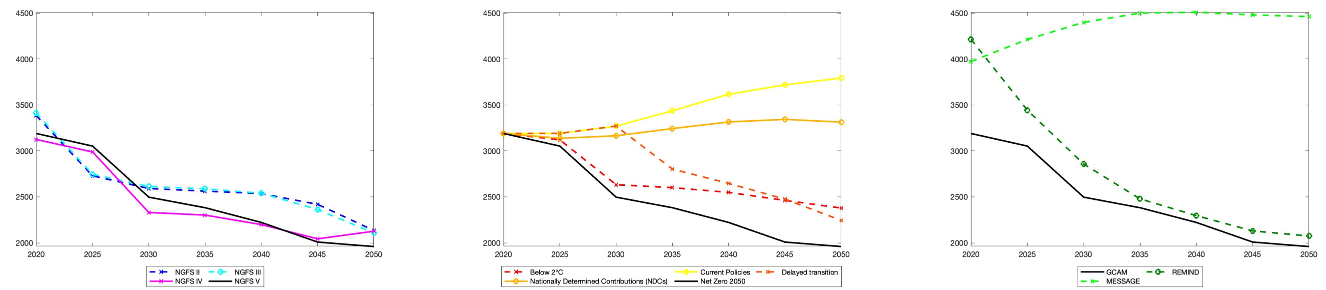
Figure A4: Global cement production across vintages (REMIND – Current Policies; MESSAGE – Net Zero; GCAM – Net Zero)

Vintages | Scenarios | Models

Global cement production across vintages (REMIND – Current Policies)



Global cement production across vintages (GCAM – Net Zero)

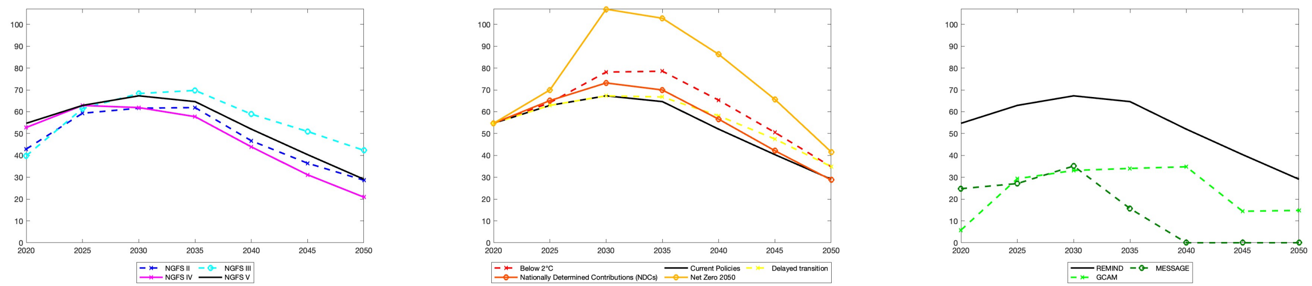


Notes: The figure reports raw variable paths across NGFS scenario vintages, scenarios, and integrated assessment models (IAMs). Paths are shown in levels and are not indexed or aligned to a common starting point. Dispersion at the base year (2020) therefore reflects *initial condition uncertainty*, arising from differences in reported historical starting values across NGFS vintages. Divergence beyond the base year reflects a combination of recalibration uncertainty, scenario uncertainty, and model uncertainty as defined in Section 3. These figures correspond to Panel A of the framework illustrated in Figures 3–7 and are intended as descriptive complements to the indexed and path-based uncertainty metrics reported in Table 1–2.

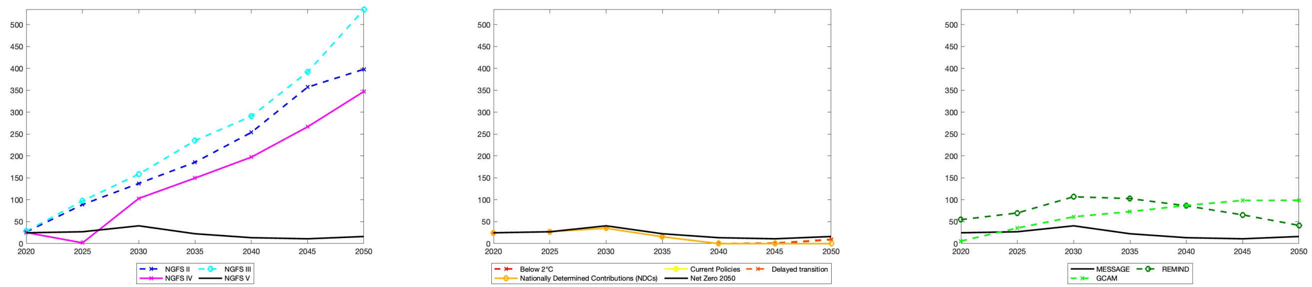
Figure A5: Nuclear electricity investment across vintages (REMIND – Current Policies; MESSAGE – Net Zero; GCAM – Net Zero)

Vintages | Scenarios | Models

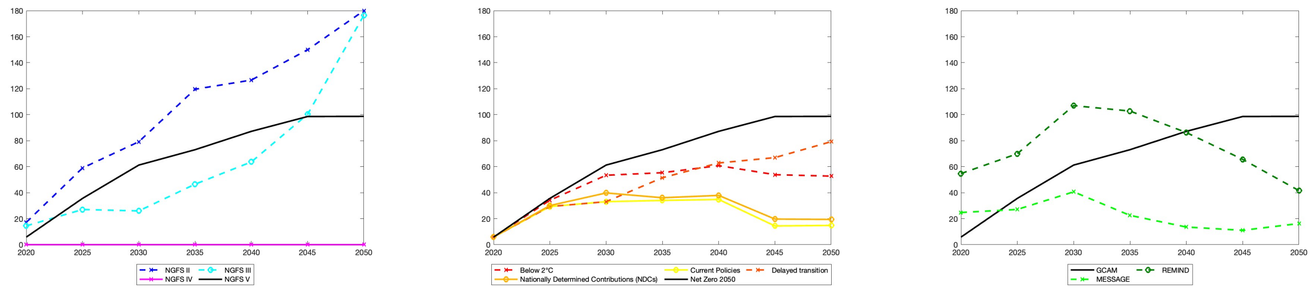
Nuclear electricity investment across vintages (REMIND – Current Policies)



Nuclear electricity investment across vintages (MESSAGE – Net Zero)



Nuclear electricity investment across vintages (GCAM – Net Zero)

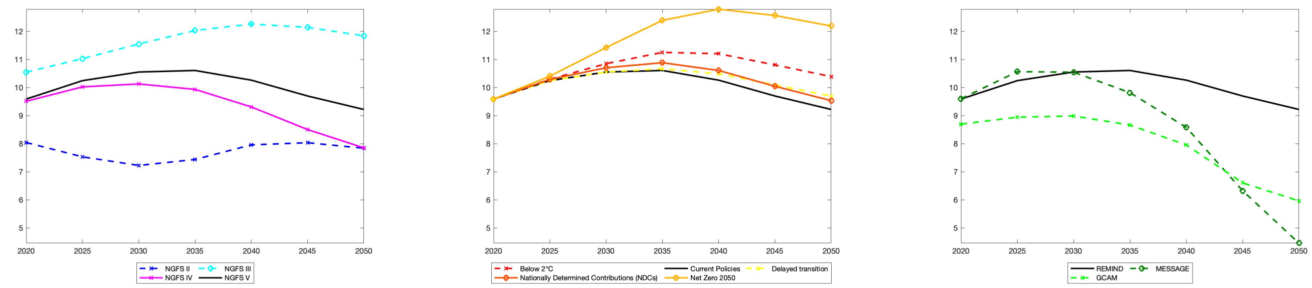


Notes: The figure reports raw variable paths across NGFS scenario vintages, scenarios, and integrated assessment models (IAMs). Paths are shown in levels and are not indexed or aligned to a common starting point. Dispersion at the base year (2020) therefore reflects *initial condition uncertainty*, arising from differences in reported historical starting values across NGFS vintages. Divergence beyond the base year reflects a combination of recalibration uncertainty, scenario uncertainty, and model uncertainty as defined in Section 3. These figures correspond to Panel A of the framework illustrated in Figures 3–7 and are intended as descriptive complements to the indexed and path-based uncertainty metrics reported in Tables 1–2.

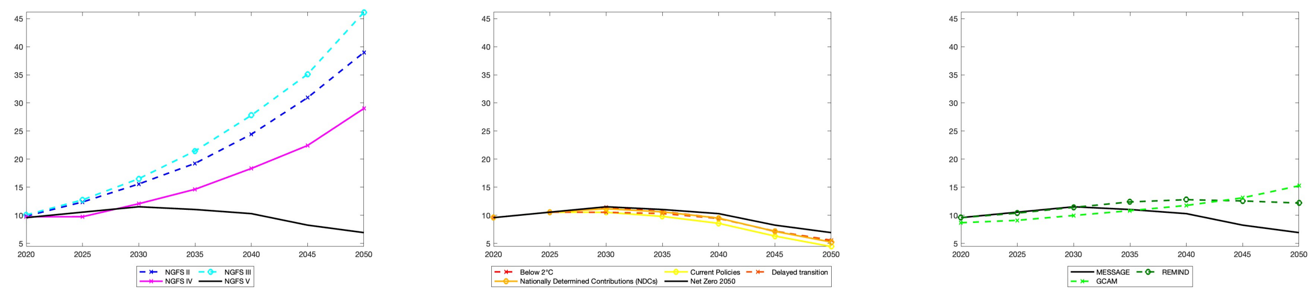
Figure A6: Nuclear electricity generation across vintages (REMIND – Current Policies; MESSAGE – Net Zero; GCAM – Net Zero)

Vintages | Scenarios | Models

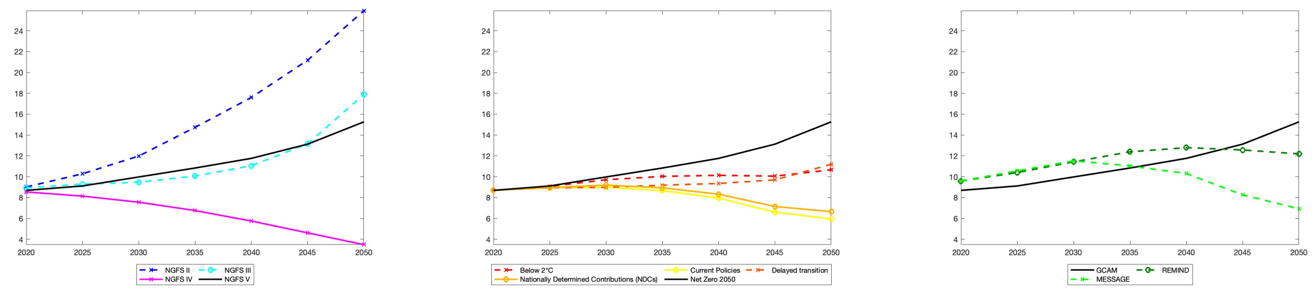
Nuclear electricity generation across vintages (REMIND – Current Policies)



Nuclear electricity generation across vintages (MESSAGE – Net Zero)

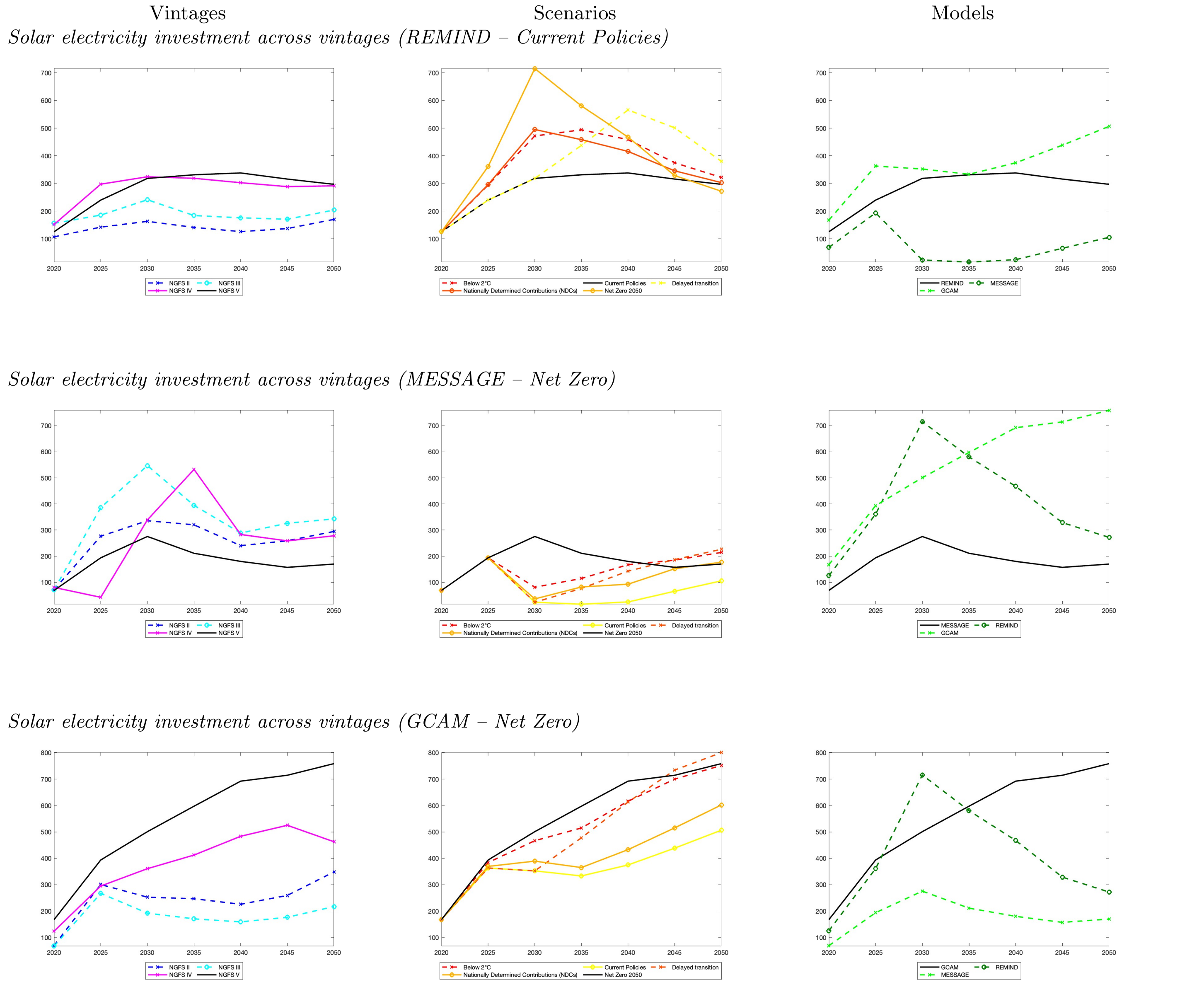


Nuclear electricity generation across vintages (GCAM – Net Zero)



Notes: The figure reports raw variable paths across NGFS scenario vintages, scenarios, and integrated assessment models (IAMs). Paths are shown in levels and are not indexed or aligned to a common starting point. Dispersion at the base year (2020) therefore reflects *initial condition uncertainty*, arising from differences in reported historical starting values across NGFS vintages. Divergence beyond the base year reflects a combination of recalibration uncertainty, scenario uncertainty, and model uncertainty as defined in Section 3. These figures correspond to Panel A of the framework illustrated in Figures 3–7 and are intended as descriptive complements to the indexed and path-based uncertainty metrics reported in Tables 1–2.

Figure A7: Solar electricity investment across vintages (REMIND – Current Policies; MESSAGE – Net Zero; GCAM – Net Zero)

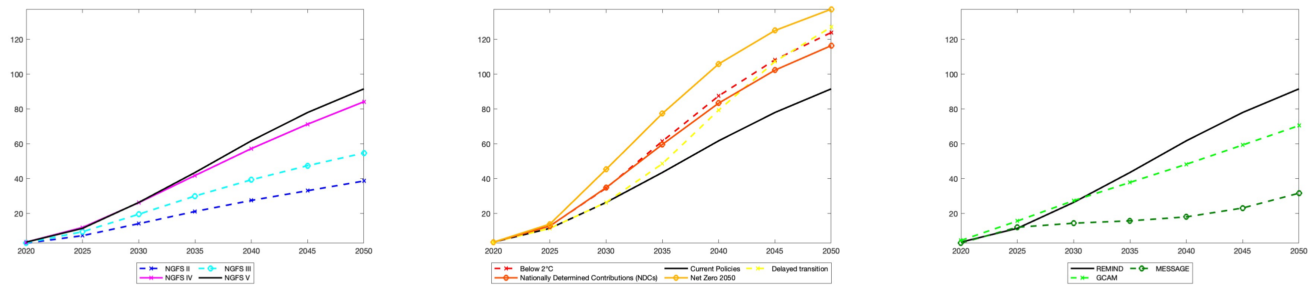


Notes: The figure reports raw variable paths across NGFS scenario vintages, scenarios, and integrated assessment models (IAMs). Paths are shown in levels and are not indexed or aligned to a common starting point. Dispersion at the base year (2020) therefore reflects *initial condition uncertainty*, arising from differences in reported historical starting values across NGFS vintages. Divergence beyond the base year reflects a combination of recalibration uncertainty, scenario uncertainty, and model uncertainty as defined in Section 3. These figures correspond to Panel A of the framework illustrated in Figures 3–7 and are intended as descriptive complements to the indexed and path-based uncertainty metrics reported in Tables 1–2.

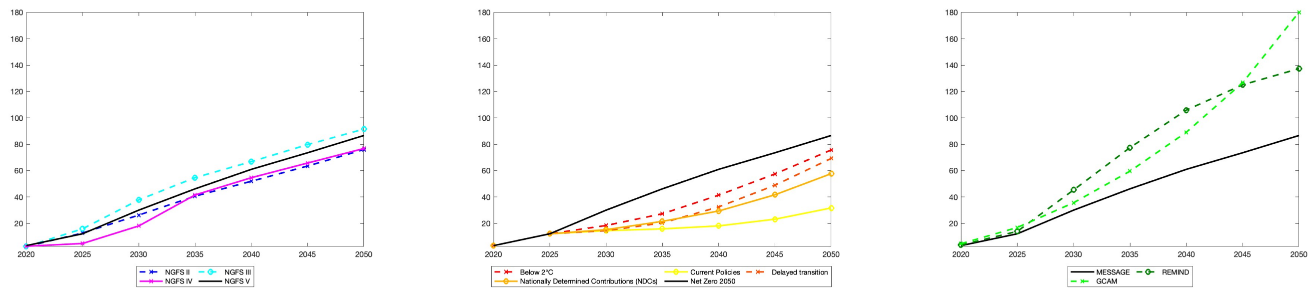
Figure A8: Solar PV electricity generation across vintages (REMIND – Current Policies; MESSAGE – Net Zero; GCAM – Net Zero)

Vintages | Scenarios | Models

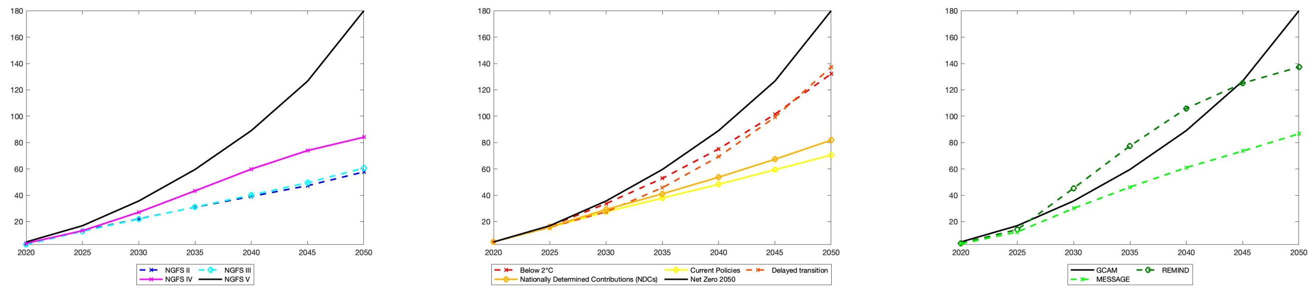
Solar PV electricity generation across vintages (REMIND – Current Policies)



Solar PV electricity generation across vintages (MESSAGE – Net Zero)



Solar PV electricity generation across vintages (GCAM – Net Zero)

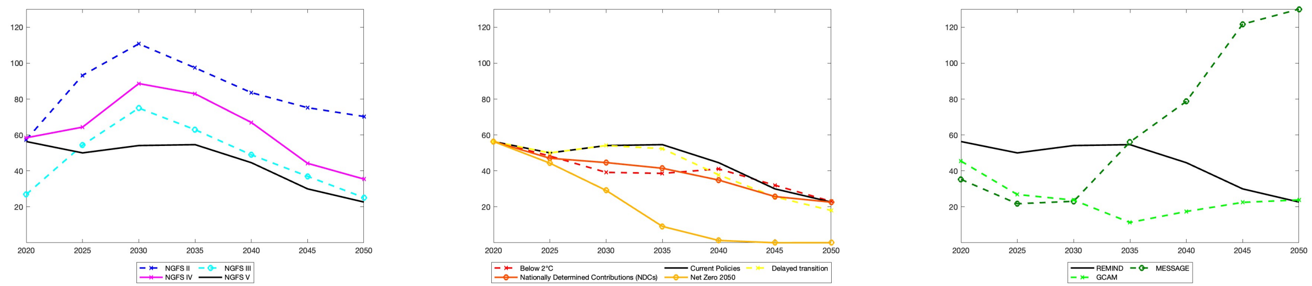


Notes: The figure reports raw variable paths across NGFS scenario vintages, scenarios, and integrated assessment models (IAMs). Paths are shown in levels and are not indexed or aligned to a common starting point. Dispersion at the base year (2020) therefore reflects *initial condition uncertainty*, arising from differences in reported historical starting values across NGFS vintages. Divergence beyond the base year reflects a combination of recalibration uncertainty, scenario uncertainty, and model uncertainty as defined in Section 3. These figures correspond to Panel A of the framework illustrated in Figures 3–7 and are intended as descriptive complements to the indexed and path-based uncertainty metrics reported in Tables 1–2.

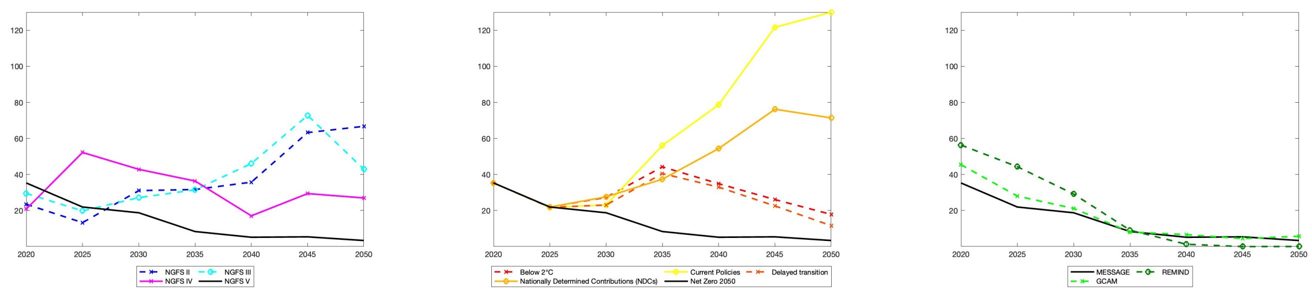
Figure A9: Gas electricity investment across vintages (REMIND – Current Policies; MESSAGE – Net Zero; GCAM – Net Zero)

Vintages | Scenarios | Models

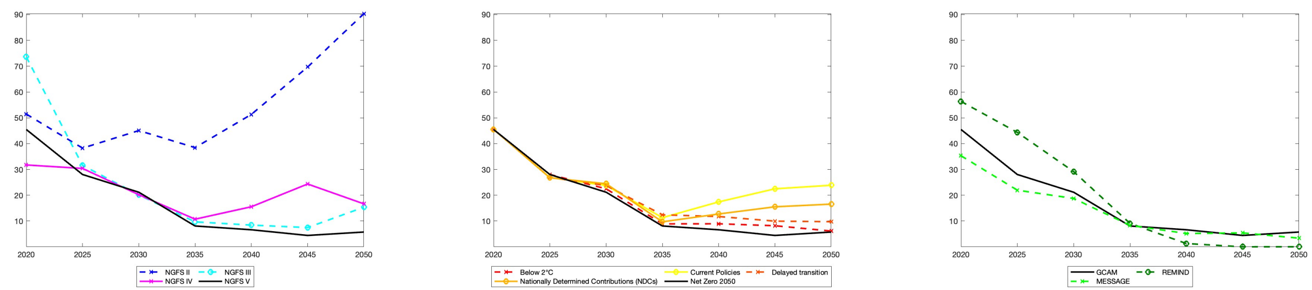
Gas electricity investment across vintages (REMIND – Current Policies)



Gas electricity investment across vintages (MESSAGE – Net Zero)



Gas electricity investment across vintages (GCAM – Net Zero)

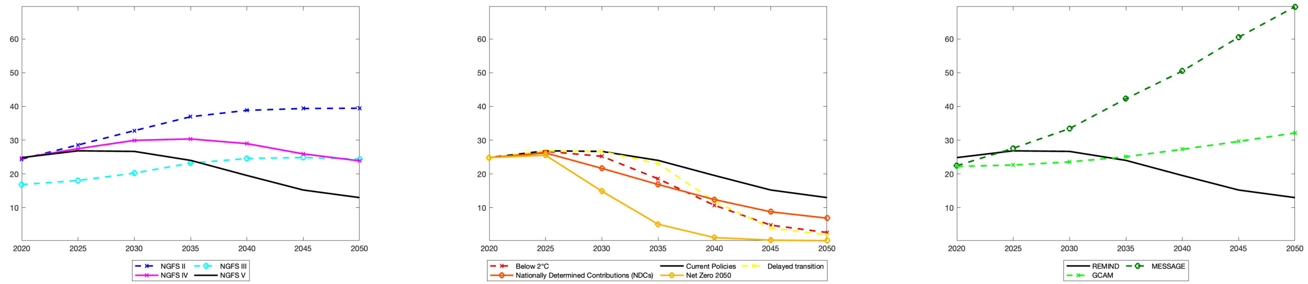


Notes: The figure reports raw variable paths across NGFS scenario vintages, scenarios, and integrated assessment models (IAMs). Paths are shown in levels and are not indexed or aligned to a common starting point. Dispersion at the base year (2020) therefore reflects *initial condition uncertainty*, arising from differences in reported historical starting values across NGFS vintages. Divergence beyond the base year reflects a combination of recalibration uncertainty, scenario uncertainty, and model uncertainty as defined in Section 3. These figures correspond to Panel A of the framework illustrated in Figures 3–7 and are intended as descriptive complements to the indexed and path-based uncertainty metrics reported in Tables 1–2.

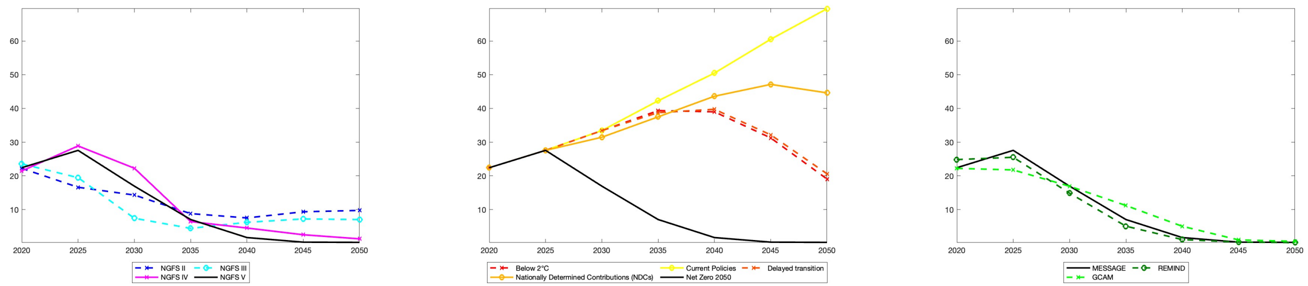
Figure A10: Gas electricity generation across vintages (REMIND – Current Policies; MESSAGE – Net Zero; GCAM – Net Zero)

Vintages | Scenarios | Models

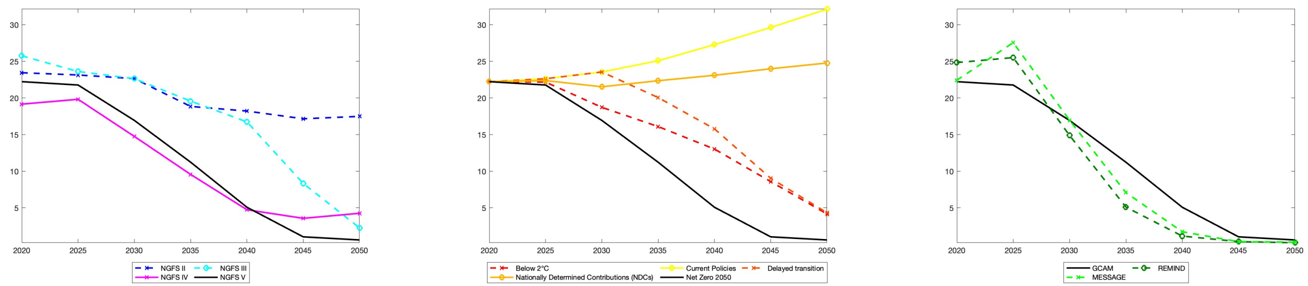
Gas electricity generation across vintages (REMIND – Current Policies)



Gas electricity generation across vintages (MESSAGE – Net Zero)



Gas electricity generation across vintages (GCAM – Net Zero)

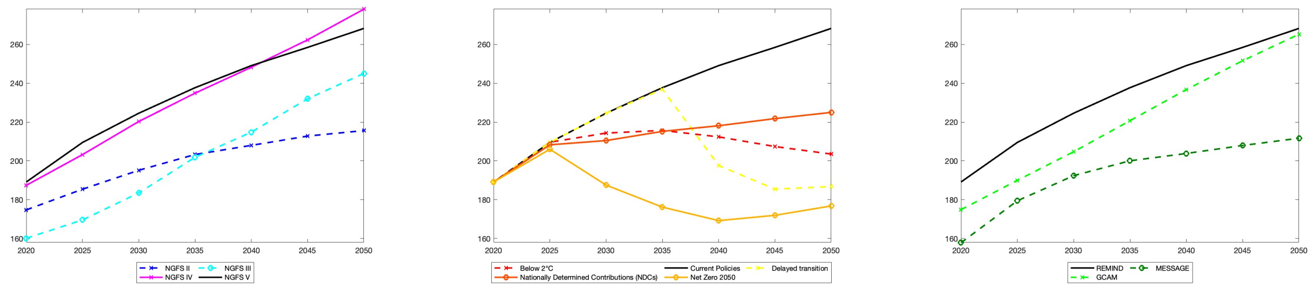


Notes: The figure reports raw variable paths across NGFS scenario vintages, scenarios, and integrated assessment models (IAMs). Paths are shown in levels and are not indexed or aligned to a common starting point. Dispersion at the base year (2020) therefore reflects *initial condition uncertainty*, arising from differences in reported historical starting values across NGFS vintages. Divergence beyond the base year reflects a combination of recalibration uncertainty, scenario uncertainty, and model uncertainty as defined in Section 3. These figures correspond to Panel A of the framework illustrated in Figures 3–7 and are intended as descriptive complements to the indexed and path-based uncertainty metrics reported in Tables 1–2.

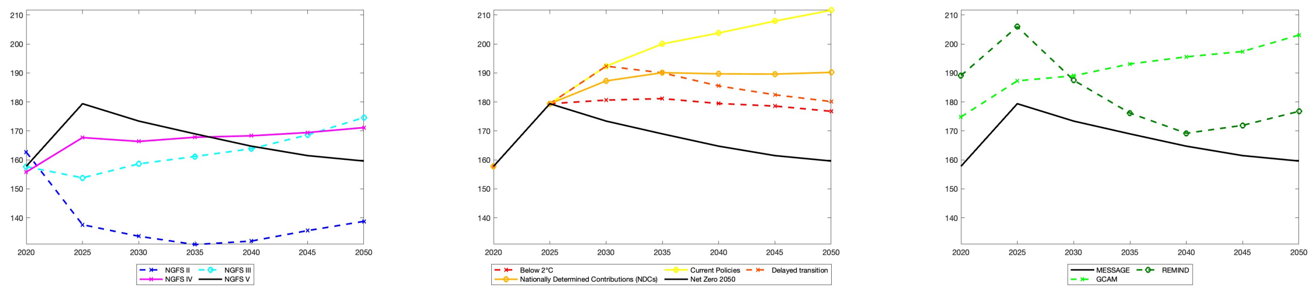
Figure A11: Final energy demand in industry across vintages (REMIND – Current Policies; MESSAGE – Net Zero; GCAM – Net Zero)

Vintages | Scenarios | Models

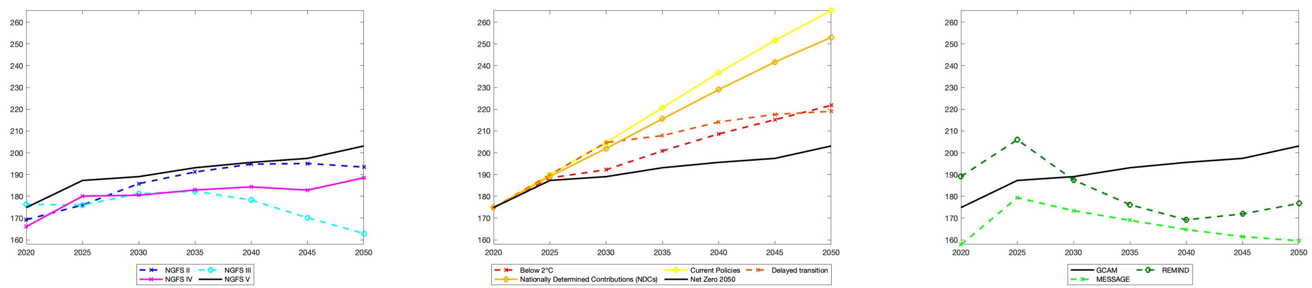
Final energy demand in industry across vintages (REMIND – Current Policies)



Final energy demand in industry across vintages (MESSAGE – Net Zero)



Final energy demand in industry across vintages (GCAM – Net Zero)

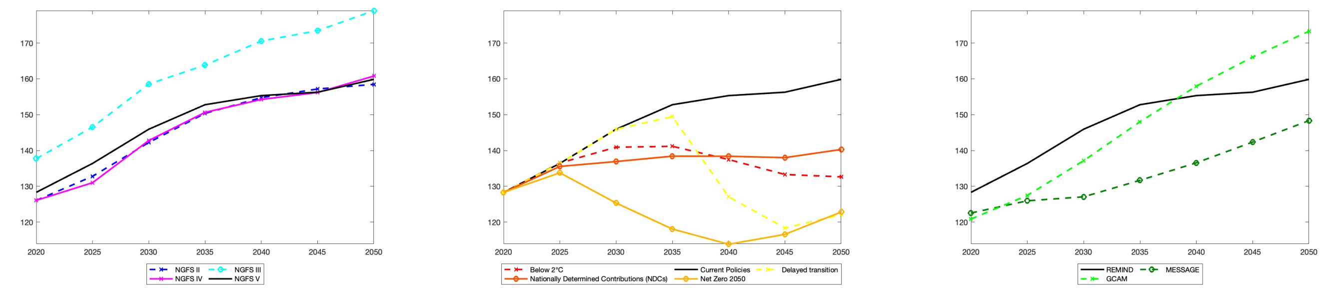


Notes: The figure reports raw variable paths across NGFS scenario vintages, scenarios, and integrated assessment models (IAMs). Paths are shown in levels and are not indexed or aligned to a common starting point. Dispersion at the base year (2020) therefore reflects *initial condition uncertainty*, arising from differences in reported historical starting values across NGFS vintages. Divergence beyond the base year reflects a combination of recalibration uncertainty, scenario uncertainty, and model uncertainty as defined in Section 3. These figures correspond to Panel A of the framework illustrated in Figures 3–7 and are intended as descriptive complements to the indexed and path-based uncertainty metrics reported in Tables 1–2.

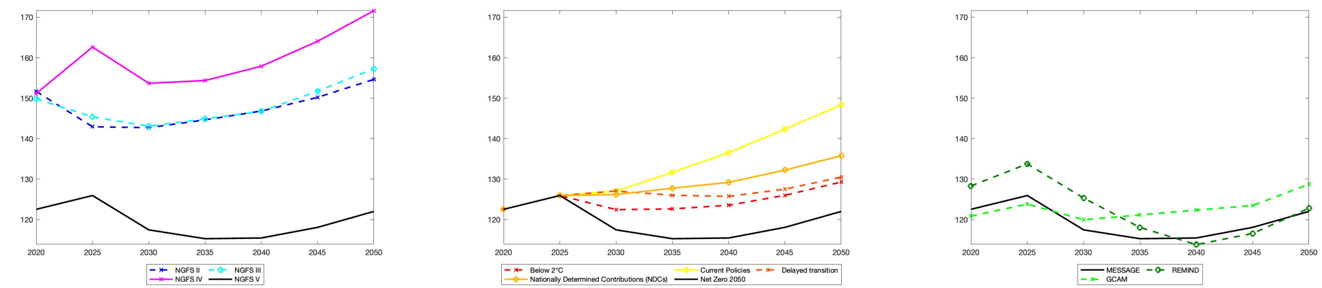
Figure A12: Final energy demand in residential and commercial across vintages (REMIND – Current Policies; MESSAGE – Net Zero; GCAM – Net Zero)

Vintages | Scenarios | Models

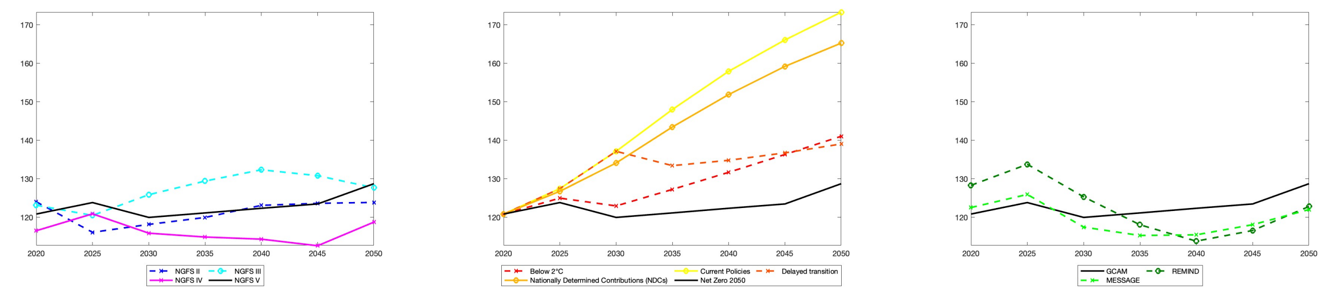
Final energy demand in residential and commercial across vintages (REMIND – Current Policies)



Final energy demand in residential and commercial across vintages (MESSAGE – Net Zero)



Final energy demand in residential and commercial across vintages (GCAM – Net Zero)

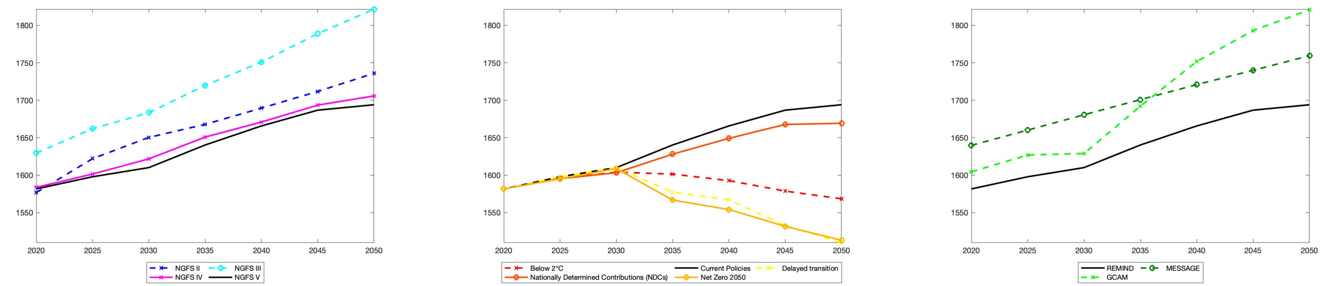


Notes: The figure reports raw variable paths across NGFS scenario vintages, scenarios, and integrated assessment models (IAMs). Paths are shown in levels and are not indexed or aligned to a common starting point. Dispersion at the base year (2020) therefore reflects *initial condition uncertainty*, arising from differences in reported historical starting values across NGFS vintages. Divergence beyond the base year reflects a combination of recalibration uncertainty, scenario uncertainty, and model uncertainty as defined in Section 3. These figures correspond to Panel A of the framework illustrated in Figures 3–7 and are intended as descriptive complements to the indexed and path-based uncertainty metrics reported in Tables 1–2.

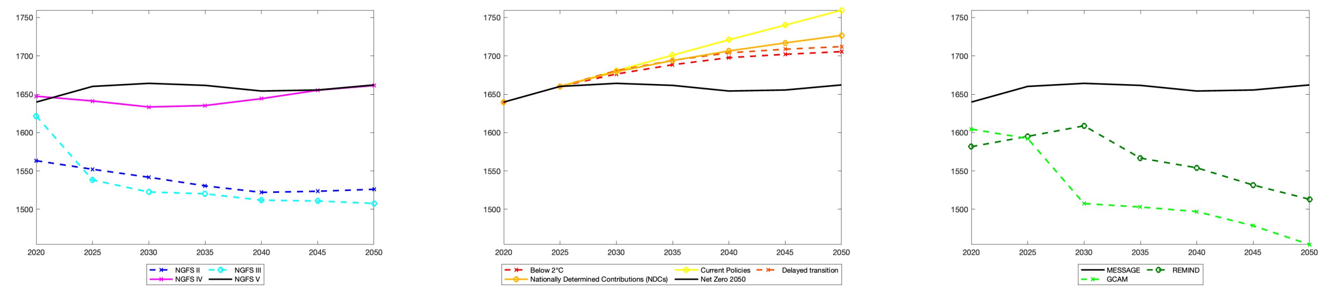
Figure A13: Cropland area across vintages (REMIND – Current Policies; MESSAGE – Net Zero; GCAM – Net Zero)

Vintages Scenarios Models

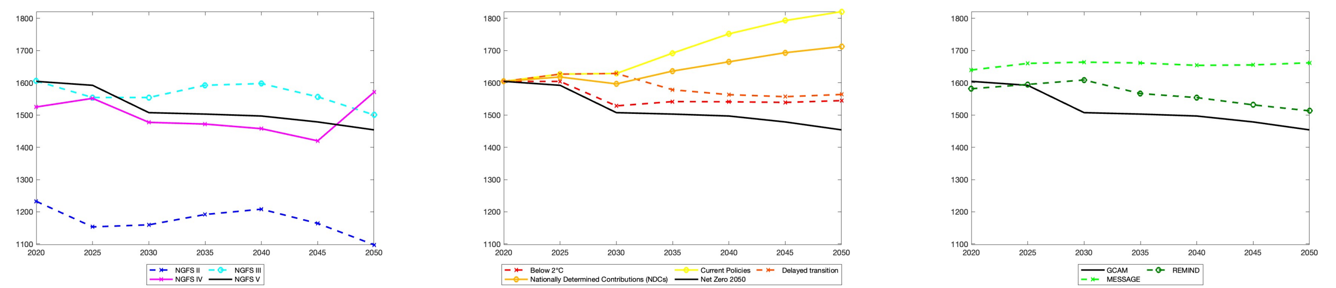
Cropland area across vintages (REMIND – Current Policies)



Cropland area across vintages (MESSAGE – Net Zero)



Cropland area across vintages (GCAM – Net Zero)



Notes: The figure reports raw variable paths across NGFS scenario vintages, scenarios, and integrated assessment models (IAMs). Paths are shown in levels and are not indexed or aligned to a common starting point. Dispersion at the base year (2020) therefore reflects *initial condition uncertainty*, arising from differences in reported historical starting values across NGFS vintages. Divergence beyond the base year reflects a combination of recalibration uncertainty, scenario uncertainty, and model uncertainty as defined in Section 3. These figures correspond to Panel A of the framework illustrated in Figures 3–7 and are intended as descriptive complements to the indexed and path-based uncertainty metrics reported in Tables 1–2.