

Beyond performative transparency: digital technology, governance conditions, and the legitimacy crisis of the Voluntary Carbon Market

Yiming Zhang

School of Advanced International Studies, Johns Hopkins University, Washington, D.C., USA

yimingzhang1112@gmail.com

Abstract. Faced with a credibility crisis, the Voluntary Carbon Market (VCM) increasingly relies on satellite monitoring, blockchain registries, and digital MRV to strengthen transparency and rebuild trust. This paper questions that approach. Information economics holds that signals distinguish quality only when costly to falsify, but digital verification lowers the cost of producing verifiable records for genuine and fraudulent projects alike, so more transparency does not automatically mean more credibility. The paper argues that trustworthy market signals depend on three governance conditions holding at once, data governance independent of commercial interests, enforceable liability for misreporting, and genuine buyer demand for quality. Using a Most Similar Systems Design (MSSD), it compares the international VCM under the Integrity Council, the EU ETS, and China's relaunched Certified Emission Reduction (CCER) mechanism, with within-case analysis through process tracing. All three cases use comparable digital tools, and in all three the party being verified pays for verification, so technology and payment structure function as controls and the real variation lies in governance conditions. First, the key discretion over how many credits are issued lies in setting the counterfactual, which cannot be directly measured, so improving measurement accuracy does not resolve the core problem. Second, liability constrains this discretion, and its presence shows through whether an auditable ledger exists, that is, whether the gap between what should have been sanctioned and what actually was can be calculated. When the three conditions fail to hold together, more abundant and continuous data still produces performative transparency rather than structural integrity.

Keywords: Voluntary Carbon Market, digital MRV, governance conditions, signaling theory, comparative case study

1. Introduction

The voluntary carbon market faces a serious credibility crisis, and the evidence has grown from isolated cases into systematic research. Synthesizing 14 studies using experimental or rigorous observational methods across 2,346 mitigation projects and one-fifth of the credit volume issued to date, Probst et al. [1] estimate that less than 16% of credits issued correspond to real emission reductions. In avoided deforestation, West et al. show by synthetic control that replacing project-set ex-ante baselines with deforestation observed ex-post leaves

only 7% of offsets tied to additional reductions, so emissions already offset are nearly three times the projects' actual mitigation contribution [1, 2].

The deeper issue is not how low these figures are but that rigorous independent teams have not reached consistent conclusions. As discussed in Section 2, offset achievement ratios recalculated by different teams for the same projects differ several-fold, and the main source of divergence is the construction of the counterfactual control group, which cannot be directly observed and must be built using statistical methods [1, 3].

This points to a conclusion more significant than the emissions reduction gap itself, the market has lost not only credit quality but the ability to assess that quality independently. When two rigorous, independent assessments differ several-fold because of an unobservable counterfactual, no single assessment gives buyers a reliable basis for pricing credits.

The market reacted quickly, and what participants said they were waiting for is telling. On Ecosystem Marketplace's estimates, trading volume fell 56% in 2023 and transaction value 61% to \$723 million, with volume falling another 25% in 2024 and value to \$535 million. Some respondents had suspended carbon credit sales while waiting for stronger integrity and quality norms, a broader shift away from nature-based projects followed high-profile reports questioning REDD+ estimates, and retirements under the ten largest standards plateaued at 182 million tons in 2024. These estimates come from a market survey whose publisher does not guarantee the accuracy of responses and treats volume and value as lower bounds [4, 5]. Demand has not disappeared, what has declined is confidence, and sellers are waiting not for better data but for stronger market norms.

The industry's response, by contrast, has been largely technological. Satellite monitoring, blockchain registries, and automated MRV systems have all been promoted as ways to rebuild trust through transparency, a trend Christiansen describes as one where digital technologies are envisioned, in response to criticism, to improve MRV, strengthen transparency, and restore market legitimacy, and which Yu et al. trace into compliance-linked offset programs in China and California [6, 7].

This paper questions that premise. The core logic comes from information economics. Signals can effectively distinguish quality only when falsifying them is costly [8], and digital MRV lowers the cost of generating verifiable records for genuine and fraudulent projects alike. The greatest discretion lies not in measuring what already exists but in setting the counterfactual, and since the counterfactual itself did not occur, no sensor can measure it directly. If a technology does not make misreporting more costly relative to honest reporting, it merely produces "performative transparency" rather than "structural integrity", where the latter describes a situation in which trustworthy data is a prerequisite for market participation rather than a reputational option. This paper therefore argues that trustworthy signals require three governance conditions at once, data governance independent of commercial interests, enforceable liability for misreporting, and genuine buyer demand for quality.

To test this claim, the paper compares three markets that use similar digital tools under different governance arrangements, as set out in Section 4.

2. Literature review

Research on integrity issues in the Voluntary Carbon Market (VCM) falls into five broad strands. Although they focus on different aspects, they all leave the same research gap.

2.1. Crisis evidence

This strand traces the problem to a specific stage but largely stays at the diagnostic level. Beyond the headline figures in Section 1, Probst et al. estimate shares of 11% for cookstove projects, 25% for avoided deforestation, and 68% for HFC-23 abatement, while wind power and improved forest management projects showed no statistically significant reductions [1].

The divergence within this strand is itself evidence. Guizar-Coutiño et al. [3] used a pixel-matching method to analyze 40 projects across nine countries and found deforestation in project areas fell 47% in the first five years, though the reductions were small in absolute terms, concluding that targeted carbon finance can mitigate tropical deforestation. When Probst et al. [1] recalculated the offset achievement ratios using the two datasets, they reached very different results. West's data gave a ratio of 8.2%, Guizar-Coutiño's gave 42%, and among the 26 projects covered by both studies the ratios were 10.5% and 18.5% respectively, with a correlation between the two sets of results of only $r = 0.4$. The principal authors of the two teams have published together, and Probst et al. attribute the discrepancy to differences in methodology and control group selection, noting that estimates are highly sensitive to how the control group is constructed even though it is itself unobservable and must be built using statistical methods [1, 3, 9-11]. This strand has therefore pointed to the counterfactual without asking what makes such discretion possible.

2.2. Governance repair

This strand accepts the diagnosis above but proposes remedies that still leave important issues unresolved. Delacote et al. [10] argue that restoring market integrity requires substantial transparency improvements across the value chain. Drawing on Akerlof's "Market for Lemons" framework, they contend voluntary certification alone is insufficient and view regulation as a key tool. They also cite Bakhtary et al., who find that while buyers pay higher prices for carbon credits with co-benefits, they attend more to the number of attributes a credit has than to its measured climate impact. In later work, Delacote et al. identify ex ante baseline scenarios as the main cause of over-crediting and propose systematic ex post evaluation by an independent evaluation organization, though this body would still be commissioned by project developers or credit buyers. The authors acknowledge its specific form and cost remain unresolved, while flagging buyer demand for quality as a key issue needing further research [10, 11]. Since the commissioning relationship stays the same, the first link in the governance chain remains untouched.

2.3. Technological optimism

This strand does not assume technology solves every problem, but tends to locate the remaining limits in technological maturity and adoption rather than governance preconditions. Siphthorpe et al. [12], evaluating 39 organizations developing blockchain solutions against a technology readiness scale, found more than half at the lowest three levels with a median of level 3, concluding that blockchain is no panacea, with barriers mainly technical, infrastructural, and regulatory. Franke et al. argue information asymmetry can be eliminated once all parties have equal data access, while acknowledging that definitive quality assurance still needs qualified experts [12, 13].

2.4. Critical perspectives

This strand argues that technology alone cannot transform the system, explaining this mainly through the objectives of those deploying it, though usually within a single context. Drawing on her observations of industry events and public communications from digital MRV companies, Christiansen argues that digital MRV mainly addresses two sociopolitical issues facing buyers, reputational risk and anticipated supply

constraints. Building on Goldstein's "non-disruptive disruption", Dempsey's "enterprising nature", and the "socioecological fixes" literature, she concludes that digital technologies alone cannot transform the system [6].

2.5. Transnational private governance

This strand supplies the governance theory, though most of it predates digital MRV's widespread adoption. Auld and Gulbrandsen [14] distinguish procedural from outcome transparency and note that nonstate certification's growth was driven mainly by retailer rather than consumer demand. Schleifer et al. distinguish shallow from deep transparency, find deep disclosure rare in their coding of voluntary sustainability programs, and argue deep transparency is necessary for information users to become "accounting actors" [14, 15]. Bartley contends certification is a compromised outcome, because as a voluntary mechanism its enforcement depends on firms, consumers, and professional auditors, limiting its impact when problems call for a fundamental redistribution of power [16]. These three actor groups map directly onto the three governance conditions developed later here, and the account predates digital verification.

2.6. The gap

These five strands are not opposed to one another, they approach the same unresolved problem from different angles, and critical and reform-oriented research, despite different starting points, ultimately confront the same question, under what conditions is quality actually demanded and paid for. None treats technology as a controlled variable or tests it comparatively across governance structures. The central question here is therefore whether the same digital technology produces different outcomes across different governance structures.

3. Theoretical framework

3.1. The cost condition of signals

The theoretical starting point draws on the classic analysis of quality uncertainty in information economics, one of whose most important implications is often overlooked. Akerlof shows that when buyers can judge quality only through market statistics, sellers have an incentive to supply low-quality products because the returns to high quality accrue to the whole group rather than any individual seller, and that mechanisms such as guarantees and brand names work less by adding information than by changing the underlying cost structure. Discussing education as a form of ability certification, he adds that the certifying institution itself must be credible [17]. Certification does not automatically generate credibility, the credibility of the certifier itself needs independent explanation, not assumption.

Spence offers a sharper criterion. Signals can distinguish between applicants only when signaling costs are negatively correlated with productive capability, since otherwise all applicants facing the same wage schedule invest in the signal the same way and it loses its power to distinguish [8]. This is the paper's core argument. Digital MRV lowers the cost of generating verifiable records for genuine and fraudulent projects alike, so if technology does not make misreporting costlier than honest reporting, it fails Spence's condition. A signal's effectiveness depends not on precision but on how costly it is to fake.

The certifier is no neutral tool either but an actor with its own payoff function. Bolton, Freixas, and Shapiro identify three conflicts of interest, the rating agency's incentive to understate risk to secure business, the issuer's ability to shop for the most favorable rating, and some investors' tendency to simply trust ratings. In

their model, what constrains lying is reputational cost, which needs both a revealing event such as a default and investors able to tell from it whether the agency lied. Their remedies therefore go beyond reputation, to automatic disclosure of all ratings solicited, enforceable oversight of analytical standards, and a lower threshold for liability claims [18]. Two points matter here, the restraint on lying depends on an event that can expose it, and the remedies lie in disclosure rules and legal liability, not better information.

3.2. The three dimensions of transparency

Discussions of transparency in carbon markets rarely address cost, focusing instead on how much is disclosed. The literature offers two distinctions, and this paper adds a third.

The first concerns who is made visible. Auld and Gulbrandsen define procedural transparency as openness about the governance process itself and outcome transparency as openness about regulated behaviors. The two concern different actors, decision-makers and regulated entities, and need not coexist, a governance process can disclose very little about its own operations while still demanding extensive outcome disclosure from others [14].

The second concerns what is made visible. Schleifer, Fiorini, and Auld argue that shallow transparency discloses formal procedures while deep transparency discloses actual processes. In verification, shallow disclosure covers basic certificate-holder information, while deep disclosure covers the certification decisions themselves, such as audit reports. Among 113 voluntary sustainability programs, 54 provided no deep disclosure in any area and only 2 achieved deep transparency across all four [15].

Together, the two distinctions show what digital MRV actually changes. Remote sensing, continuous measurement, and registry systems enhance outcome transparency, but procedural transparency, how baselines are set, who approves the methodology, what disputes arise during validation, has not grown in tandem, and it is precisely there that deep disclosure is scarcest. Technology has made regulated entities more visible, not regulators.

A third distinction is therefore needed, concerning what happens once something becomes visible. Schleifer et al. argue deep transparency is necessary for information users to become accounting actors [15], but necessary is not sufficient, whether anyone bears a cost once falsification becomes visible is separate from how much has been disclosed. This paper therefore distinguishes performative transparency, where disclosure may be complete, verifiable, and continuous yet does not alter the discloser's payoff structure, from structural integrity, where credible data becomes necessary for market participation rather than a reputational option. In essence, this third distinction applies the cost condition from Section 3.1 to disclosure itself.

3.3. Three governance conditions

Combining the requirements of the previous two sections makes the problem institutional, under what conditions can data from digital verification serve as a credible signal. This paper proposes three conditions, all of which must be met at once, corresponding one-to-one with the three conflicts of interest Bolton et al. identify [18]. Table 1 sets out each with its theoretical basis and observable indicators.

Table 1. The three governance conditions

Condition	Theoretical basis	Observable indicators
C1 Data governance independent of commercial interests	Akerlof, the certifying institution itself must be credible [17]. Bolton et al., the agency's incentive to understate risk to secure business [18]	(a) legally binding conflict-of-interest restrictions and (b) an entity capable of continuously overseeing verification bodies and aligning their conduct
C2 Enforceable liability for misreporting	Bolton et al., remedies lie in disclosure rules and legal liability [18]. Yu et al., supplier versus buyer liability [7]	(a) a clearly designated liable party with graded and enforceable consequences and (b) an auditable ledger, that is, whether the gap between what should have been sanctioned and what was actually sanctioned can be calculated
C3 Genuine buyer demand for quality	Bolton et al., rating quality declines as the proportion of trusting investors increases [18]. Auld and Gulbrandsen, labels provide little information, and demand may need to be created [14]	(a) whether quality differences are reflected in price and demand and (b) whether buyers bear risks from holding invalid credits

Two points deserve emphasis. The C1 indicator is not who pays the verification fees, since in all three cases verification costs are borne by the party being verified. The second indicator under C2, proposed here, turns something usually treated as a matter of degree into something whose presence or absence can be assessed.

These three conditions must be held together, not separately. If verification lacks independence, even highly precise data remains an extension of the verified party's own account. If no one bears liability, problems verification identifies do not change any actor's payoff. If buyers do not demand quality, the stringency the first two conditions maintain gets competed away because no one will pay for it. This interdependence rules out a causal identification strategy based on any single condition and calls for the comparative case design in Section 4.

4. Research design and methodology

This paper uses the Most Similar Systems Design (MSSD). Seawright and Gerring define it as a case selection method where the selected cases are similar across all measured independent variables except the variable of interest, so a difference in outcome Y can be attributed to that variable X1 rather than the control-variable vector X2. The method requires at least two cases and can be adapted for more complex comparisons [19]. Figure 1 sets out the design used here.

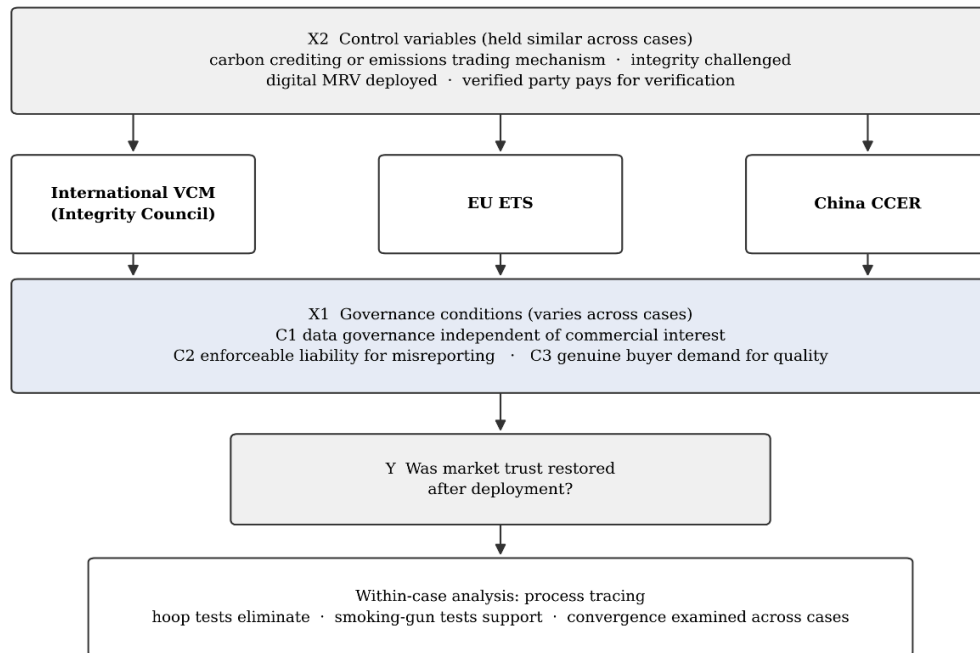


Figure 1. Research design

In this design, X1 refers to the governance conditions under which digital technology is deployed, with its value set by whether each of the three conditions in Section 3.3 is met. Y refers to whether market trust is restored after digital technology is deployed.

Of the four background conditions held similar across the cases in Figure 1, the fourth was established through this paper's own verification of primary materials, that in all three cases verification costs are borne by the party being verified. *Articles 11 and 18 of China's Measures for the Administration of Voluntary Greenhouse Gas Emission Reduction Trading* stipulate that both project validation and emission reduction verification are commissioned by the project owner [20]. The corresponding EU ETS and international VCM arrangements are detailed in Sections 5.2 and 5.1.

Since the payment structure is identical across the three cases, it cannot serve as an explanatory variable and instead functions as a control, like technology.

This paper does not adopt a quantitative identification strategy, for three reasons. First, the large-N matching approach Seawright and Gerring recommend requires a cross-case dataset to draw matches from [19], but too few carbon pricing mechanisms meet the background conditions above to form such a population. Second, as explained in Section 3.3, the three governance conditions are interdependent, and the failure of any one disables the other two, so no treatment variable can be identified in isolation. Third, the evidence here consists of causal-process observations rather than the data-set observations typical of quantitative work [21], valuable not for sample size but for diagnostic relevance to specific causal steps.

The within-case analysis uses process tracing. Collier classifies tests by whether passing them is necessary or sufficient for accepting an inference, ranging from straw-in-the-wind tests that discriminate weakly to doubly decisive tests that both confirm one hypothesis and rule out others. This paper uses two, hoop tests, which set a necessary criterion and can rule out an explanation without confirming it, and smoking-gun tests, which set a sufficient criterion and strongly support an explanation without rejecting it on failure [21].

A 2019 refinement, used in Section 6, is also relevant. Zaks expanded the relationship between rival explanations beyond mutual exclusivity to three types, coincident, congruent, and inclusive [21], so the motive-based explanation in the critical literature need not be eliminated to make room for the condition-based explanation proposed here.

This design also has clear limits. Seawright and Gerring note that cases broadly representative of the population give the strongest basis for generalization, and classify the most similar method as confirmatory when the hypothesis concerns a specific X-Y relationship [19]. The three cases here were selected for institutional variation, not representativeness, so no statistical extrapolation is claimed. Collier also cautions that fully ruling out a hypothesis is often difficult in social science [21], so the inferences in Section 5 are bounded by the evidence available.

5. Case analysis

5.1. The international Voluntary Carbon Market

The international VCM is both where the crisis originated and where the technological turn has been most visible. This section traces the sequence, first with a hoop test that rules out one rival explanation, then testing the third governance condition through an internal market comparison.

5.1.1. The 2023 shock and market response

Following empirical studies and high-profile media reports questioning the accuracy of emission reduction estimates under REDD+ methodologies, the market contracted sharply, and Ecosystem Marketplace noted that some respondents had suspended carbon credit sales while waiting for the market to establish stronger integrity and quality norms [4]. What market participants identified as missing was not better data but stronger norms.

5.1.2. The measurement explanation fails a hoop test

One straightforward explanation locates the problem in measurement. If project-specific forest cover layers are more accurate than the global layers used in independent assessments, over-crediting may simply reflect a measurement dispute. On Collier's classification this is a hoop test the explanation must pass to stay under consideration [21]. Swinfield et al. tested it directly and found the opposite. Among the 36 projects reporting certified measurements, the median deforestation rate measured with the European Union Tropical Moist Forest Annual Change Collection was 0.26% per year against a certified median of 0.17%, so underestimating deforestation in control areas cannot explain the observed over-crediting. Across all credits assessed, the global over-crediting ratio was 10.7 (95% bootstrapped CI 5.4 to 26.5), and on the subset of 17 projects with complete data, sequential substitution attributes roughly 2% of over-crediting to the choice of remote sensing data for the project area, 57% to reference area selection, minus 37% to remote sensing for the reference area, and a residual of about 78% to ex ante modeling. The authors caution that the subsample is small, most effects are not statistically significant, and considerable uncertainty remains around the ex ante modeling attribution [9].

The measurement explanation therefore fails the hoop test. Digital verification improves the measurement stage, which accounts for only about 2% of over-crediting, while the greater discretion sits in setting the counterfactual, which never occurred and so no sensor can measure it, the empirical counterpart to Section 3.1.

5.1.3. The institutional response

The scope the Integrity Council set for itself is telling, it assesses programs and methodologies rather than individual projects, and lists verifier oversight as work still to be done [22]. Its REDD+ assessment concluded

in November 2024 with three methodologies approved [23].

The reform nevertheless made real progress at one source while leaving another untouched. Under VM0048, project developers no longer set baselines from self-selected reference areas, a change Swinfield et al. view positively since it removes project-selected reference areas as a major bias source and shifts the analysis to independent data providers with no financial stake in issuance, though such methodologies still rely on ex ante predictions and VM0048 still lets nested projects use ex ante spatial modeling [9, 23]. Measured against the earlier decomposition, the reform addresses the source behind about 57% of over-crediting while leaving the 78% source largely untouched.

The same decisions reveal the state of the second governance condition. Verra excluded five older REDD+ methodologies from assessment, methodologies accounting for most REDD+ credits on the market and roughly a quarter of all credits retired in 2023, with the only consequence being that credits issued under them cannot receive the CCP label [23]. The strongest available measure was therefore denying future label eligibility. No liable party was designated, no auditable ledger existed, and credits already issued and used stayed unaffected.

5.1.4. A 2024 contrast internal to the market

The third governance condition can be tested here through an intra-market comparison. Ecosystem Marketplace records that CCP-approved REDD+ and afforestation credits had not yet entered the market, while trading volumes rose in the clusters where fugitive emissions and ozone-depleting substances projects won CCP approval. The average price of ACR credits doubled and CAR-registered projects' total transaction value rose 23%, with overall market impact limited only because the share of approved project types was still small in 2024 [5].

The value of this comparison lies in isolating the CCP label's effect from other factors. The market, year, and buyers were the same, and the only difference was whether the label had already attached to a category. Where it had, buyers paid for it. The third condition is therefore not impossible to meet, it depends on whether a quality label can attach to the relevant credits, and for REDD+, the category where the crisis originated, none yet existed.

5.1.5. Reading the sequence

The first governance condition was not met at any point. Methodological flexibility was open to both project proponents and certification bodies, both with financial incentives to increase credit issuance, and Verra's program fee schedule supports this, while the standard setter assesses no individual projects and had not established oversight of verifiers [9, 22]. On Collier's classification this is a smoking-gun test, sufficient to establish that the first condition fails, though its absence would not establish the reverse [21]. Technology and rules were updated in parallel, yet the three conditions were never met at once. The result is what this paper calls performative transparency, disclosure grew more complete and continuous while the step that determines credit issuance, and who bears its consequences, stayed unchanged.

5.2. The European Union Emissions Trading System

The EU ETS is often seen as the publicly regulated counterpart to the VCM, but attributing the difference to state authority versus private self-regulation is inaccurate. Calel et al. describe the EU ETS arrangement, each polluter must submit an annual emissions report approved by an accredited third-party auditor, and because the auditor is hired by the polluter, the authors identify an obvious conflict of interest [24]. The structure matches the voluntary market, confirming the background condition established in Section 4.

The main difference lies in the institutional apparatus. *Article 43 of Implementing Regulation 2018/2067* requires the verifier to be independent and impartial and treats consulting services for the monitoring and

reporting process as a conflict of interest, while *Article 25* requires a verification report to be reviewed before issuance by someone uninvolved in that verification. The Integrity Council, by contrast, still lists oversight of validation and verification companies as work yet to be completed [22, 25].

That such arrangements exist does not prove they work. Calem et al. [24] found the same observed non-compliance rate, 0.2%, among regulators with and without authority to appoint auditors, and a lower rate, 0.0% against 0.6%, where regulators could train and supervise them, but they stress these are descriptive statistics. In more rigorous tests neither appointment nor supervisory powers produced significant differences, only site inspection powers did, and beyond fines no compelling evidence supported tools such as enhanced auditor oversight [24]. The first governance condition in the EU ETS is therefore specified institutionally but not demonstrated empirically, since if detailed independence rules do not alter compliance behavior, whether the condition holds cannot be read off the rules alone.

The second governance condition provides the clearest cross-case contrast here. Fines for excess emissions were set at 40 euros per tonne and raised to 100 from 2008, and official data show 98.8% of installations have complied since 2006, yet the observed non-compliance should have produced roughly 13 billion euros in fines against no more than about 2.1 billion collected, with fine probability and severity explaining only about a tenth of the variation in compliance rates, a pattern known as Harrington's paradox [24].

The significance of these figures lies not in weak enforcement but in the fact that the gap itself can be calculated. A clearly identified liable party and a statutory unit penalty make an auditable ledger possible, satisfying the second indicator in Section 3.3. In the voluntary carbon market, how much should have been sanctioned cannot even be asked. The difference lies not in enforcement intensity but in whether accounting is possible at all.

The third governance condition takes a different form in a compliance market. Demand for accuracy arises not from voluntary preference but from the cap and the surrender obligation. Polluters must surrender allowances matching their emissions, so the accuracy of emissions reports directly determines their legal liability. This matches Auld and Gulbrandsen's observation that demand for quality may need to be created [14], here by a mandatory obligation rather than buyer preference.

One alternative explanation must be ruled out, that the EU ETS is ineffective overall. Using a generalized synthetic control method, Bayer and Aklin find the system cut CO₂ emissions by roughly 1.2 billion tonnes between 2008 and 2016, equal to 3.8% of total EU emissions and close to half of member states' Kyoto pledges [26]. Persistently low carbon prices do not therefore mean the mechanism has failed.

5.3. China's CCER scheme

China's CCER is the only case where the governance architecture was rebuilt before the market reopened, running from suspension to legislation to relaunch, which makes it possible to see how institutions are designed absent market pressure.

5.3.1. *The old scheme and its 2017 suspension*

The CCER mechanism launched in 2012. One feature of the old scheme matters here, by March 2017 forest carbon sinks made up only 2.1% of registered projects [27], so China had little exposure to the category where the international crisis was concentrated. In March 2017 the acceptance of applications was suspended. Zhu et al. record this as implementing requirements for streamlining administration and delegating powers, covering five application categories while letting already registered projects keep trading, whereas Yu et al. record it as meant to strengthen the regulatory framework [7, 27]. This paper does not judge the motivation, but on either account, what was suspended was market entry, not the market itself.

5.3.2. *What was built during the suspension*

The results of the reform appear in the *Measures for the Administration of Voluntary Greenhouse Gas Emission Reduction Trading (Trial)*, issued in 2023. On independence, *Article 37* severs funding, asset, and management ties between verification bodies and the project owners commissioning them and bars those bodies from preparing owners' project documents, *Article 18* bars one body from both validating and verifying the same project, and *Article 30* requires a clean five-year record plus a risk fund or insurance matching business risk. *Articles 11 and 18* nonetheless leave the payment structure unchanged [20].

On liability, *Article 43* escalates from an order to rectify and a fine for false materials, through revocation and a three-year bar for fabricated data, to cancellation equal to the false portion, and finally compulsory cancellation, return of any shortfall, a higher fine, and an open-ended bar. *Article 44* extends liability to verification bodies and withdraws responsible staff's qualifications, and *Article 49* required projects filed before 14 March 2017 to reapply, writing the old-new break into law [20]. This contrasts with Section 5.1, where Verra excluded older methodologies from assessment while credits already issued stayed unaffected.

Zhu et al.'s institutional comparison further shows the revised CCER requires emission reduction disclosure and whole-process supervision absent from the old CCER, CDM, and VCS, though the CDM and VCS are not blank on project disclosure, both requiring a 30-day publication before verification [27]. Two qualifications apply, the comparison concerns institutional design rather than effectiveness, and two authors are affiliated with the Ministry of Ecology and Environment's Policy Research Center. These design advantages establish only the rule component of the first governance condition, not its behavioral one.

5.3.3. *The 2024 relaunch and its present state*

The mechanism relaunched in January 2024, and the third governance condition is built in institutionally. *Article 28* provides that certified reductions may offset allowance surrender obligations in the national and local emissions trading markets and must be cancelled in the registry once used [20]. Technology is deployed rather than withheld. Yu et al. record that the methodologies for coal mine methane utilization and energy-efficient highway tunnel lighting mandate continuous IoT monitoring with hourly recording integrated into the national platform, that new afforestation methodologies incorporate remote sensing including airborne LiDAR-assisted sequestration monitoring, and that the platform cross-verifies reported data against independent sources such as electricity grid settlement data [7]. This confirms the treatment of technology as a control variable in Section 4.

5.3.4. *Reading the sequence*

The first two governance conditions receive the most complete treatment at the level of rules among the three cases, but evidence of their behavioral effects is still lacking. The second indicator in Section 3.3 also cannot yet be evaluated, because the mechanism relaunched too recently to have generated an enforcement record, so the auditable ledger has been specified but remains empty. Set against the EU ETS, where the ledger exists and already shows a gap, the indicator distinguishes three states, no ledger, a ledger with an enforcement record, and a ledger without one.

The Chinese case also carries a vulnerability at precisely the stage shown critical in Section 5.1. The new methodologies divide additionality demonstration into exemption, simplified demonstration, and general demonstration, with widely recognized technology types of good data quality exempted altogether, a standardization that reduces administrative burden and improves predictability but risks issuing credits to financially viable projects lacking genuine additionality [7, 27]. Since the largest over-crediting source lies in setting the counterfactual, the exemption route effectively removes this test for certain categories. Two systems with almost opposite governance models thus run into problems at the same stage, in the international VCM discretion arises from ex ante baselines projects set themselves, while in China it arises from rules

exempting certain categories from demonstration. The mechanisms differ, but the location of the problem does not.

No judgment can yet be made on the outcome variable. Yu et al. state explicitly that the Ministry of Ecology and Environment has made offset integrity a priority, but that the effectiveness of the new safeguards still awaits assessment through ongoing implementation [7].

6. Discussion

The patterns from the cross-case comparison fall into two parts, and only together do they form a complete explanation.

First, the discretion determining how many credits are issued is concentrated in setting the counterfactual, which never occurred and so cannot be directly measured. Four independent lines of evidence point to the same stage. Swinfield et al. attribute about 78% of over-crediting to ex ante modeling against only about 2% to the choice of remote sensing data for the project area [9]. Delacote et al. independently identify ex ante baseline scenarios as the main cause [11]. Probst et al. find estimates highly sensitive to the construction of the unobservable control group [1]. West et al. attribute the discrepancy directly to the gap between ex ante baselines and ex post counterfactuals [2]. Coming from different teams, methods, and samples, this convergence is persuasive. Discretion is not spread across the process but concentrated in an object that cannot, in principle, be directly observed, so however much measurement accuracy improves, it cannot reach the step that determines the outcome. The Chinese case confirms this from another direction, exempting certain technology types from demonstrating additionality removes the test of the counterfactual altogether [27]. That two systems with almost opposite governance models run into problems at the same stage shows the problem's location is set by the nature of that stage, not the design of any particular institution.

Table 2. The three cases against the three governance conditions

	C1 Independence	C2 Liability and ledger	C3 Buyer demand
International VCM	Not met. Methodological flexibility opens to certifiers with a financial stake, and the standard setter assesses no individual projects and has not established verifier oversight	Not met. No designated liable party, and the strongest instrument is denial of future CCP label eligibility	Untestable in REDD+. Where the label had attached, buyers paid for it
EU ETS	Specified but not demonstrated. <i>Articles 43 and 25</i> impose independence, but no compelling evidence of effect	Met. Statutory unit penalty and identified liable party, roughly 13 billion euros owed against no more than 2.1 billion collected, the gap calculable	Created by the cap and the surrender obligation
China CCER	Most complete at the level of rules (<i>Articles 37, 18, 30</i>), behavioral evidence absent	Liable parties designated on both supplier and verifier sides with graded consequences (<i>Articles 43, 44</i>), ledger specified but empty	Created by offset eligibility (<i>Article 28</i>)

Second, what constrains discretion is not measurement but liability, and the existence of liability can be observed through whether an auditable ledger exists. Table 2 summarizes the three cases against the three

conditions, and they differ most sharply on this second dimension [20, 24].

The auditable ledger is an indicator this paper proposes. Its value lies in turning a phenomenon usually treated as a matter of degree into a question of whether it exists at all. Enforcement strictness is a continuous variable observers may reasonably dispute, but whether the amount owed can even be calculated is binary. The European shortfall can be calculated not because enforcement is strict but, on the contrary, because a clearly defined liable party and a statutory penalty make accounting possible even when enforcement is lax. Technology cannot reach the step that determines the outcome, only institutions can, and the observable indicator of institutional difference is the ledger, not data volume, the empirical manifestation of the distinction between performative transparency and structural integrity.

The relationship between this paper and the critical literature is not one of opposition but a division of labor. As discussed in Section 2.4, Christiansen traces digital MRV to the reputational risk and anticipated supply constraints facing buyers and argues that digital technologies themselves will not transform the system, though this does not mean transformation is out of reach [6]. On the Zaks framework adopted in Section 4, both explanations can hold at once rather than being mutually exclusive. If buyers care about reputational risk rather than quality itself, the third governance condition does not hold, which is precisely why technology can produce only performative transparency. Christiansen's finding explains why the third condition fails, while this paper sets out what follows from that failure. The same logic applies to the governance repair literature. The independent evaluation organization Delacote et al. propose would still be commissioned by project developers or credit buyers, so the commissioning relationship stays the same and the proposal never reaches the first condition, and the authors themselves flag buyer demand as a crucial requirement needing further study [11]. The critical and reformist strands reach the same unresolved question from opposite directions, and the three governance conditions offer one comparable, falsifiable answer.

Several trade-offs deserve acknowledgment. Delacote et al. note that ex post evaluation shifts the risk of compromised integrity from the market as a whole to project developers and local communities and may limit supply [11]. Auld and Gulbrandsen, citing Bernstein, note a legitimacy trade-off whereby higher participation and transparency increase buy-in but slow decisions [14]. Swinfield et al. found two projects in Madagascar and one in Peru with mean over-crediting ratios below one, meaning they achieved more avoided deforestation than claimed [9]. Over-crediting is therefore not universal, technology is ineffective when the three conditions fail, not every project suffers credit inflation.

Two of the three outcomes cannot yet be adequately observed. No credits had been issued under the international VCM's approved methodologies when the decision was taken, and China's CCER relaunched too recently for its safeguards to be assessed [7, 23]. Only the EU ETS has a sufficiently long record, so this paper offers specification and preliminary testing of the governance conditions, not confirmation of outcomes. Some of the literature relied on here also occupies a position worth noting. Three of the four authors of Siphthorpe et al. are employed by Shell, and two authors of Zhu et al. are affiliated with the Policy Research Center of the Ministry of Ecology and Environment [12, 27]. One testable prediction follows, if China's CCER publishes an enforcement record letting that gap be calculated, the credibility of its market signal should exceed that of the international VCM over the same period.

Finally, the three governance conditions were already foreshadowed in the transnational private governance literature. As Bartley argues, cited in Section 2.5, certification is a compromised outcome precisely because enforcement relies on firms, consumers, and professional auditors, and these three parties correspond to the supplier, buyer, and verifier here [16]. That judgment predates digital MRV, and the structure it describes has not been altered by the technology's arrival, which itself supports the paper's central claim, the conditions that make a signal credible are governance conditions, and technology has not changed what those conditions are.

7. Policy implication

The three recommendations correspond to the three governance conditions. This paper does not advocate any single jurisdiction's model, Yu et al. note that no single governance model is inherently superior and effective design depends on context [7], so each recommendation is organized around a mechanism rather than a country.

7.1. Sever the link between payment and outcome

In all three cases the party paying the verification fee is the party being verified, and this structure is unlikely to change. Reversing it is neither realistic nor necessary. The workable approach is to ensure the paying party cannot influence the result, and the problem has three levels for which institutional tools already exist. First, on conflicts of interest, *Article 43 of the EU Implementing Regulation* and *Article 37 of the Measures* address consulting and asset ties between the verifier and the verified party. Second, on the reliability of results, *Article 25* of the same EU regulation mandates independent review [20, 25]. The third level concerns ratings shopping. Bolton et al. note that up-front payment eliminates ratings inflation but not shopping, so all solicited ratings must be disclosed, and they also document a more radical proposal, a body randomly assigning rating agencies with fees paid in advance to it, eliminating both conflicts of interest and shopping at once, though the proposal was dropped during the U.S. legislative process [18].

These tools must be used together, since up-front payment weakens the incentive to invest in accuracy, and enforceable oversight of analytical standards must also exist [18]. The verifier oversight the Integrity Council had not yet established, noted in Section 5.1, is precisely that gap.

7.2. Designate a liable party and build an auditable ledger

The key issue is not the severity of the penalty but whether liability rests on an identifiable party, whether consequences are graded and enforceable, and whether the gap between what should have been sanctioned and what was actually sanctioned can be calculated. Existing models supply the first two elements. The EU ETS establishes measurable unit liability through a statutory per-unit penalty, *Articles 43 and 44 of the Measures* apply graded consequences to both suppliers and verifiers, and *Article 30* requires verification bodies to hold a risk fund or insurance commensurate with their business risk [20, 24].

The third element, the ledger itself, is the most operational recommendation offered here. The EU ETS shows that even when enforcement is imperfect, the gap between what was owed and what was collected remains calculable, while the international VCM cannot perform the same accounting, not for lack of data but because no party has been designated as liable. Standard setters should therefore periodically publish two figures, credits that should have been sanctioned under their own rules and credits actually sanctioned. This needs no new technology and no new institution, only a designated liable party and a public record.

7.3. Turn quality from preference into eligibility

The third condition is hardest to address directly because it lies outside institutional designers' control. The problem is not buyers' morality but the absence of a mechanism that makes quality differences visible and priced. Delacote et al. note that developers lack incentives to conduct ex post evaluations precisely because no mechanism signals the resulting quality, and propose a two-tier classification where only credits that underwent open, scientifically grounded ex post evaluation are eligible for firms' climate claims, with the rest classified as climate contributions excluded from such claims. Buyers intending to use credits as offsets would then, by design, have to care about quality [11].

This shifts the third governance condition from an expectation about buyer preferences to an eligibility rule, the structural integrity defined in Section 3.2, and the within-case comparison in Section 5.1 offers preliminary evidence of feasibility. Complementary tools include mandatory disclosure, corporate ESG ratings, and tax incentives, while risk-sharing is needed to avoid constraining supply [11].

Implementation also matters. Calel et al. suggest governments building a carbon market avoid creating new institutions from scratch wherever possible, noting the EU ETS was built on an existing pollution control directive, and that relying on tested reporting systems and legal mechanisms avoids solved problems and reduces blind spots at the boundaries between old and new regulatory bodies [24]. The three recommendations are therefore best implemented by adding carbon-specific requirements to existing legal and reporting infrastructure, not by creating new institutions.

8. Conclusion

A comparison of the three cases answers this paper's question in the affirmative, but the source of the differences is not technology. All three markets deploy digital MRV, and in all three the party paying for verification is the party being verified, so neither technology nor payment structure explains the differences in outcomes.

This pattern falls into two parts. First, the discretion over how many credits are issued sits mainly in setting the counterfactual, which never actually occurred and so no sensor can measure it directly, and technology improves only the observable side of the process. Second, what constrains this discretion is not measurement but liability, observable through whether an auditable ledger exists, specifically whether the gap between what should have been sanctioned and what was actually sanctioned can be calculated. On this dimension the three cases differ far more than they do in technical sophistication.

The three governance conditions proposed in Section 3.3 must therefore be held together. They are interdependent, and the failure of any one is enough to make the other two ineffective. When the three conditions do not hold together, more abundant, precise, and continuous data still yields performative transparency, disclosure that may be complete and verifiable while leaving the discloser's payoff structure unchanged. Structural integrity becomes possible only when credible data is a requirement of market participation rather than a reputational option.

These conclusions are limited by the current stage of the evidence. Two of the three outcomes cannot yet be adequately observed, and only the EU ETS has a record of sufficient length, so this paper offers specification and preliminary testing of the governance conditions rather than confirmation of outcomes. As approved methodologies begin issuing credits and China's CCER accumulates an enforcement record, the auditable ledger proposed here can serve as an object of observation for that work.

This points to a broader conclusion. The credibility of a signal carried by technology does not depend on the technology itself but on whether the surrounding institutions have already made falsifying that signal sufficiently costly.

References

- [1] Probst, B. S., Toetzke, M., Kontoleon, A., Díaz Anadón, L., Minx, J. C., Haya, B. K., Schneider, L., Trotter, P. A., West, T. A. P., Gill-Wiehl, A., & Hoffmann, V. H. (2024). Systematic assessment of the achieved emission reductions of carbon crediting projects. *Nature Communications*, 15, 9562.
- [2] West, T. A. P., Wunder, S., Sills, E. O., Börner, J., Rifai, S. W., Neidermeier, A. N., Frey, G. P., & Kontoleon, A. (2023). Action needed to make carbon offsets from forest conservation work for climate change mitigation.

- Science*, 381(6660), 873-877.
- [3] Guizar-Coutiño, A., Jones, J. P. G., Balmford, A., Carmenta, R., & Coomes, D. A. (2022). A global evaluation of the effectiveness of voluntary REDD+ projects at reducing deforestation and degradation in the moist tropics. *Conservation Biology*, 36(6), e13970.
 - [4] Ecosystem Marketplace. (2024). *On the path to maturity: State of the voluntary carbon market 2024*. Forest Trends Association.
 - [5] Ecosystem Marketplace. (2025). *Meeting the moment: Renewing trust in carbon finance*. State of the voluntary carbon market 2025. Forest Trends Association.
 - [6] Christiansen, K. L. (2025). Relegitimising the voluntary carbon market: Visions of digital monitoring, reporting and verification. *Environment and Planning A: Economy and Space*, 57(8), 1190-1205.
 - [7] Yu, R.-X., Gordon, J. A., Dai, F., Lu, S.-Z., Zhang, D., & He, K.-B. (2026). A comparative review of compliance-linked carbon offset programs in China and California: The new CCER and the Compliance Offset Program. *Advances in Climate Change Research*, 17, 877-889.
 - [8] Spence, M. (1973). Job market signaling. *Quarterly Journal of Economics*, 87(3), 355-374.
 - [9] Swinfield, T., Williams, A., Coomes, D., Dales, M., Ferris, P., Guizar-Coutiño, A., Hartup, J., Holland, J., Jaffer, S., Jones, J. P. G., Lam, M. O. K., Keshav, S., Madhavapeddy, A., Toye-Scott, E., West, T. A. P., & Balmford, A. (2026). Learning lessons from over-crediting to ensure additionality in forest carbon credits. *Nature Communications*, 17, 3944.
 - [10] Delacote, P., L'Horty, T., Kontoleon, A., West, T. A. P., Creti, A., Filewod, B., LeVelly, G., Guizar-Coutiño, A., Groom, B., & Elias, M. (2024). Strong transparency required for carbon credit mechanisms. *Nature Sustainability*, 7(6), 706-713.
 - [11] Delacote, P., Chabé-Ferret, S., Creti, A., Duffy, K., Elias, M., Guizar-Coutiño, A., Filewod, B., Groom, B., Kontoleon, A., LeVelly, G., L'Horty, T., Missirian, A., & West, T. A. P. (2025). Restoring credibility in carbon offsets through systematic ex-post evaluation. *Nature Sustainability*, 8(7), 733-740. DOI:10.1038/s41893-025-01589-7
 - [12] Siphthorpe, A., Brink, S., Van Leeuwen, T., & Staffell, I. (2022). Blockchain solutions for carbon markets are nearing maturity. *One Earth*, 5(7), 779-791.
 - [13] Franke, L., Schletz, M., & Salomo, S. (2020). Designing a blockchain model for the Paris Agreement's carbon market mechanism. *Sustainability*, 12(3), 1068.
 - [14] Auld, G., & Gulbrandsen, L. H. (2010). Transparency in nonstate certification: Consequences for accountability and legitimacy. *Global Environmental Politics*, 10(3), 97-119.
 - [15] Schleifer, P., Fiorini, M., & Auld, G. (2019). Transparency in transnational governance: The determinants of information disclosure of voluntary sustainability programs. *Regulation & Governance*, 13(4), 488-506.
 - [16] Bartley, T. (2007). Institutional emergence in an era of globalization: The rise of transnational private regulation of labor and environmental conditions. *American Journal of Sociology*, 113(2), 297-351.
 - [17] Akerlof, G. A. (1970). The market for "lemons": Quality uncertainty and the market mechanism. *Quarterly Journal of Economics*, 84(3), 488-500.
 - [18] Bolton, P., Freixas, X., & Shapiro, J. (2012). The credit ratings game. *Journal of Finance*, 67(1), 85-111.
 - [19] Seawright, J., & Gerring, J. (2008). Case selection techniques in case study research: A menu of qualitative and quantitative options. *Political Research Quarterly*, 61(2), 294-308.
 - [20] Ministry of Ecology and Environment & State Administration for Market Regulation. (2023). *Measures for the administration of voluntary greenhouse gas emission reduction trading (Trial)* [in Chinese]. Order No. 31.
 - [21] Collier, D. (2011). Understanding process tracing. *PS: Political Science and Politics*, 44(4), 823-830.
 - [22] Integrity Council for the Voluntary Carbon Market. (2024). *Fact sheet: Integrity Council for the Voluntary Carbon Market*. <https://icvcm.org/wp-content/uploads/2024/09/CCP-Fact-Sheet-FINAL-01-8-24.pdf>
 - [23] Integrity Council for the Voluntary Carbon Market. (2024, November 15). *Integrity Council approves three REDD+ methodologies*. <https://icvcm.org/integrity-council-approves-three-redd-methodologies/>

- [24] Calel, R., Dechezleprêtre, A., & Venmans, F. (2025). Policing carbon markets. *Climate Policy*, 25(9), 1489-1507.
- [25] Commission Implementing Regulation (EU) 2018/2067 of 19 December 2018 on the verification of data and on the accreditation of verifiers pursuant to Directive 2003/87/EC of the European Parliament and of the Council. *Official Journal of the European Union*, L 334, 31.12.2018, p. 94.
- [26] Bayer, P., & Aklin, M. (2020). The European Union Emissions Trading System reduced CO2 emissions despite low prices. *Proceedings of the National Academy of Sciences*, 117(16), 8804-8812.
DOI:10.1073/pnas.1918128117
- [27] Zhu, L., Wang, M., Huang, D., & Lu, Y. (2024). The current development and outlook of China certified emission reduction trading scheme. *Chinese Journal of Urban and Environmental Studies*, 12(3), 2450015.