2026 S.-T. Yau High School Science Award

Research Paper

Student name: Qihao Fu , Yigeng Tang ,Tongxi Shen

School: Chongqing Bashu Secondary School

Province: Chong Qing

Country/Region: China

Advisor name: Xu Wang, Rong Huang

Advisor affiliation: Chongqing Bashu Secondary School

Title: From Sampling the Forest to Sampling the Error: Microeconomic Cost-Effectiveness by Optimal Spatial Sampling for Forest Carbon Verification——A Case Study at Saihanba

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

From Sampling the Forest to Sampling the Error: Microeconomic Cost-Effectiveness by Optimal Spatial Sampling for Forest Carbon Verification

A Case Study at Saihanba

Author: Qihao Fu , Yigeng Tang ,Tongxi Shen

Abstract

This paper develops a decision-analytic framework for forest-carbon verification that combines optimal spatial sampling with a microeconomic cost-effectiveness model. A project developer with a limited field budget chooses how many verification plots to buy, where to place them and how to combine them with satellite imagery, so as to minimize the expected sum of field cost, credit deductions under the CCER uncertainty rule and the cost of mis-mapped carbon. Field plots are accurate but costly; satellite imagery is cheap but inaccurate without field calibration. We write the carbon-stock field as a remote-sensing prediction plus a spatially correlated residual, $Y = R + E$, and spend plots on calibrating the satellite model rather than re-measuring what it already sees. A Gaussian-process argument shows why satellite pixels alone cannot pin down the regional mean, and an adapted SMP-cLHS design places plots across the range of the satellite indices. On 576 LiDAR-derived 20 m reference cells in one 23 ha plantation block at Saihanba, with repeated draws for every configuration, the optimized design gives an 80% chance that the project-mean error stays within 5% with about 25 plots, against 32 for zone-random placement and more than 45 for current CCER practice. In the small-sample range of 18 to 54 plots it matches the map accuracy of random placement with 32-46% fewer plots; at 90 plots the advantage is gone. The loss model shows economies of scale in verification: below about 16,000 t CO2e of creditable removals field cost sets the campaign at about 18 plots, while above 50,000-140,000 t CO2e, depending on the design, the 10% deduction line sets it at about 45 plots, and risk aversion moves them slightly lower (all at a carbon price of CNY 70.76 and CNY 3,000 per plot; scenario). The official formula scores the spread of plot values. The optimized design keeps that spread stable, so near the optimal campaign it crosses the 10% line least often (5% of draws at 36 plots, against 10-14%) and it is also the most accurate: private and public interest point the same way, though the rule rewards stability rather than accuracy. At the developer's own optimum, expected over-issued credits are worth 2.7 to 5.1 times the field bill, an external cost the developer does not bear; against current practice, optimized placement roughly halves it at no extra field cost. A model-assisted (GREG) uncertainty rule with a verifier-drawn seed would cut the cost-minimizing campaign from 45 to 27 plots, but it lowers the external cost only together with representative placement. Field costs come from a forestry fee guideline and carbon prices from 2024-2025 market data. Farm- and national-scale figures only illustrate the method and are not estimates; the analysis covers single-period carbon-stock verification and complements rather than replaces the official

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公示

accounting rules.

Keywords: value of information; costly state verification; carbon-asset MRV; cost-effectiveness analysis; optimal spatial sampling; incentive design; CCER

JEL classification: D81, D82, D83, D86, Q23, Q54, Q58, C61

2026 S. -T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

Contents

1 Introduction 6
1.1 Background 6
1.2 Related work 7
1.3 Research gap 9
1.4 The economic decision problem and contributions 10
2 Study area and dataset 11
2.1 Study area 11
2.2 LiDAR-derived reference carbon-stock map 13
2.3 Satellite imagery and spectral indices 14
2.4 Relation between spectral indices and carbon stock 15
3 Methods 15
3.1 Overall design and evaluation 15
3.2 Zonal SMP-cLHS sampling 16
3.3 Seven estimators 19
3.4 Control designs 21
3.5 Data boundary 21
3.6 Theory: error propagation, information gain, and plot value 21
3.7 Economic framework: cost-effectiveness and the marginal value of a plot . . 24
4 Results 26
4.1 Full-factorial main results 26
4.2 SMP gains: size, source, and decay 27
4.3 Ablation across the estimator spectrum 29
4.4 Sample size versus error 32
4.5 Comparison under the CCER mean criterion 33
4.6 Zone-level accuracy check 34
4.7 Supplementary experiments 35
5 Economic analysis: from statistical gains to verification costs 36
5.1 Cost-effectiveness frontier for the map target 36
5.2 Diminishing returns and the shadow price of map accuracy 37
5.3 The project-mean target and threshold risk 38
5.4 Expected total loss: deduction risk and the optimal campaign 39
5.5 Field logistics: does dispersion raise the cost per plot? 43
5.6 Cost and carbon-price scenarios 43
5.7 From carbon stock to creditable removals 44
5.8 Who pays, incentives, and alternatives 45
5.9 Implications for project developers and verification rules 47
6 Discussion 48
6.1 Why optimized sampling improves small-sample mapping 48
6.2 Estimator performance under different sampling designs 48
6.3 Implications for repeated carbon monitoring 49
6.4 Implications for CCER-oriented carbon accounting 49
6.5 Limitations 50
7 Conclusion 50

2026 S. T. Yau High School Science Award
仅用于2026年高中科学奖论文公示

References 51

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

1 Introduction

1.1 Background

Accurate, objective quantification of forest carbon is a precondition for pricing forest carbon credits and for the credibility of the carbon market within China's "dual-carbon" institutional architecture. At the individual-tree level, aboveground biomass can be estimated from field measurements such as diameter at breast height (DBH) and tree height using allometric equations and subsequently converted to carbon stock. Extending such measurements to large forest areas, however, requires sampling because a complete tree census is generally impractical. Operational forest-carbon inventories therefore rely on field plots and statistical inference to estimate carbon stocks over larger areas.

Carbon-stock measurement sits at the front of a two-step chain. The first step is measurement: a project needs a defensible number for how much carbon its forest holds. The second is finance: credits are issued for the verified change in carbon stock against a baseline, so a verified stock estimate in each monitoring period is an input to issuance, discounting and trading. Step two inherits the error of step one, so measurement error converts directly into money, in field-plot costs and in the credits lost when uncertainty crosses a deduction threshold.

Current CCER practice makes the first step expensive. Its uncertainty accounting treats field plots as independent draws within strata defined by species and stand age, so in practice the main lever for precision is plot count: on our study area, the methodology's own Appendix F formula keeps theoretical uncertainty above the 10% no-deduction line until roughly 27 plots, and every extra plot is a field campaign in difficult terrain.

Existing accounting approaches are primarily designed to estimate project-level carbon stocks and their associated uncertainty rather than to provide wall-to-wall spatial information. Spatially explicit carbon-stock estimates are nevertheless important for stand-level forest management, targeted verification, and assessments of disturbance risks such as fire, pests, and forest degradation. Obtaining wall-to-wall reference measurements through field surveys or repeated airborne LiDAR campaigns is generally impractical because of their high cost and limited repeatability. A more scalable approach is to combine a limited number of accurately measured reference plots with wall-to-wall satellite observations and a statistical model that relates remote-sensing covariates to forest carbon stock.

This approach raises two closely related methodological questions: where should a limited number of reference plots be located, and which estimator should be used to combine plot observations with wall-to-wall remote-sensing information? Sampling design and estimator choice are often studied separately, although their effects may interact. A sampling configuration that performs well for one estimator may not provide

2026 Siyu High School Science Award
仅用于2026年度科学奖公示

the same benefit for another, while improved plot placement may also change the relative performance of competing estimators. Sampling design and estimator choice should therefore be evaluated jointly. In this study, we compare alternative sampling designs and estimators within a common experimental framework to examine how their interaction affects forest carbon-stock estimation and mapping under limited sample sizes.

For a project developer the economics are direct. Verification is priced per plot, and Table 35 of the CCER methodology [3] prices uncertainty as well: above 10%, every issued credit is discounted. The gain measured in this paper is on the cost side: optimized placement reaches a given accuracy with fewer plots, and the same plots also yield a wall-to-wall map whose error is quantified (Section 5). Its effect on credit deductions runs through the uncertainty the verifier computes, and Section 5.4 measures that effect under the current rule. The framework is designed to supplement CCER practice: it takes over where the plots go and how their observations are combined, while the official criterion stays as it is.

1.2 Related work

Classical forest inventories commonly use simple random, systematic, or stratified sampling to estimate population-level quantities from a limited number of field plots. These designs and their statistical properties are well established in survey sampling theory (Cochran, 1977), and the CCER afforestation methodology similarly permits plot-based sampling schemes including random, systematic, and stratified designs [3]. Such approaches are well suited to estimating project-level means or totals and their associated uncertainty, but they are not intended to characterize the continuous spatial distribution of forest carbon stocks.

Remote sensing provides a complementary source of information by supplying wall-to-wall covariates related to forest structure and biomass. Statistical or machine-learning models can be fitted between plot-level carbon measurements and remotely sensed variables such as the normalized difference vegetation index (NDVI) and near-infrared reflectance of vegetation (NIRv) and subsequently applied across the study area to generate spatially continuous carbon-stock estimates. The accuracy of such models may nevertheless be constrained by spectral saturation at high biomass levels, variation in canopy structure, and spatially varying relationships between spectral response and carbon stock. These limitations mean that remote-sensing predictions retain spatially structured residual errors even when the overall predictive relationship is strong.

Spatially explicit forest-carbon mapping has long been recognized as important for carbon monitoring and climate-mitigation programs. Gibbs et al. (2007) highlighted the role of forest carbon monitoring in REDD-related applications [15], while Saatchi et al. (2011) demonstrated large-area carbon-stock mapping and associated uncertainty

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

assessment across tropical forests [16]. A central issue in such wall-to-wall products is that prediction errors are often spatially correlated rather than independent. Geostatistical methods provide a formal framework for characterizing this dependence through covariance functions or variograms and for incorporating it into spatial prediction and uncertainty analysis (Goovaerts, 1997; Webster and Oliver, 2007) [13,14]. This spatial dependence forms an important basis for the sampling and inference framework developed in this study.

Geostatistical estimators differ in how they combine spatial dependence with auxiliary information. Ordinary kriging [6, 7] uses the spatial covariance or variogram structure of the target variable to predict values at unsampled locations, whereas regression kriging [8, 9] combines a deterministic regression component derived from auxiliary covariates with spatial interpolation of the remaining residuals. The latter is particularly relevant to remote-sensing-assisted carbon mapping because satellite observations can explain broad spatial variation while geostatistical modelling accounts for residual spatial dependence. The relative performance of these approaches depends on both the predictive strength of the auxiliary variables and the spatial structure remaining in the residual field.

Sampling optimization provides a complementary way to improve the efficiency of field data collection. Minasny and McBratney (2006) introduced conditioned Latin hypercube sampling (cLHS), which selects samples so that their distribution in environmental covariate space approximates that of the target population. Zeng et al. (2015) subsequently developed the Sampling based on Multi-temporal Prior knowledge (SMP) strategy, which combines representativeness in auxiliary-variable space with spatial dispersion and uses simulated annealing to identify efficient sampling configurations. The approach was later incorporated into China's national standard for leaf area index (LAI) product validation [1,2]. However, its application to forest carbon-stock mapping has received much less attention. In particular, it remains unclear how sampling designs based on remote-sensing covariates interact with different spatial estimators and whether gains from improved sample placement are consistent across estimation methods.

Model-assisted survey sampling provides a closely related statistical foundation. Särndal et al. (1992) developed a general framework for improving population inference using auxiliary information while retaining design-based properties [17], and Gregoire and Valentine (2008) reviewed its application to sampling and estimation in natural-resource inventories [18]. Building on this general principle, we treat the wall-to-wall remote-sensing prediction as auxiliary information and use field observations primarily to characterize the remaining spatial residual field. Under this formulation, candidate plot

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

locations can be evaluated according to the reduction in uncertainty they provide for regional carbon-stock estimation and spatial prediction. This links sampling design with estimator performance and provides a common basis for comparing alternative plot configurations under a fixed sampling effort.

Economics of verification. Verification is a purchase of information, and the economics of information treats it as a costly input like any other [27]. Value-of-information theory prices a signal by the decision loss it avoids [19, 20], and Blackwell's comparison of experiments ranks signals by informativeness [32]; a plot design is such an experiment. When one party must pay to observe what another party knows, the problem is one of costly state verification [28], and when the agent's action cannot be observed, an efficient contract conditions on every signal that is informative about that action [29]. Who pays the auditor matters too. In a field experiment with Indian industrial plants, third-party auditors paid by the plants they audited reported emissions more favourably than auditors paid from a central pool [34]; CCER verification bodies are likewise engaged and paid by the project owner. Carbon offsets add a quality-uncertainty problem [21]: buyers cannot check creditable removals, over-crediting has been documented [22], and optimal offset contracts must deal with information the regulator lacks [33]. This paper applies these ideas to the field campaign itself: where plots go, how many are bought, and how the uncertainty rule turns them into payoffs.

1.3 Research gap

Using 576 fixed 20 m × 20 m LiDAR-derived reference plots in Saihanba, we address three research questions concerning sampling efficiency, estimator performance, and their combined effect on forest carbon-stock mapping.

(1) Under small sample sizes, to what extent does SMP-cLHS optimized sampling improve carbon-stock mapping relative to within-zone random, systematic, and contiguous sampling designs? We further examine which components of the optimized design contribute to these improvements and how the benefits change with increasing sample size.

(2) How does estimator performance depend on sampling design and sample size? We compare seven estimators representing different ways of using remote-sensing covariates, spatial dependence, and combinations of the two, and evaluate whether improved placement changes their relative performance.

(3) How much additional mapping accuracy can be achieved when optimized sampling is combined with an appropriate estimator, relative to a conventional project-level mean estimator? For spatial comparison, the latter is represented as a spatially constant map based on the estimated mean carbon stock

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公示

All comparisons use the same reference data, evaluation locations, and random realizations, allowing sampling designs and estimators to be assessed under directly comparable conditions. The experimental design uses only auxiliary information available at the time of sampling and avoids using evaluation data during sample selection or model fitting. In addition to the empirical comparisons, we develop a theoretical formulation based on the spatial covariance structure of the residual field to quantify the information contributed by alternative plot configurations.

The remainder of the paper is organized as follows. Section 2 describes the study area and datasets. Section 3 introduces the sampling designs, estimators, evaluation framework, theory and economic model. Section 4 presents the experimental results, and Section 5 turns them into an economic analysis of verification cost, deduction risk and incentives. Section 6 discusses sampling efficiency, estimator performance and practical implications, and states the limitations. Section 7 concludes. A reader interested mainly in the economics can read Sections 1.4 and 3.7, the summary at the start of Section 4, and then Section 5.

1.4 The economic decision problem and contributions

Type of model. This is a microeconomic, decision-analytic model of a single price-taking firm that buys verification information. It is not a macroeconomic or market-equilibrium model. The carbon price and the accounting rules are taken as given; the firm chooses how much information to buy and where to buy it. The economic question is: in carbon-asset MRV, how should a limited verification budget be allocated across plots to minimize the sum of measurement cost, expected credit-deduction loss and the cost of mis-mapped carbon? Section 5.8 relaxes the rule-taking assumption and asks how the accounting rule itself shapes the firm's choice.

Who decides, and over what. A forest-carbon project developer, or a verifier acting on its behalf, must decide how many field plots to measure in a monitoring campaign, where to put them, and how to combine their measurements with satellite information. Field plots give accurate but costly information. Satellite imagery covers every hectare at low marginal cost, but its carbon predictions carry error. The developer works under two constraints: a field budget, and an accounting standard that ties the uncertainty of the carbon estimate to the credits a project may claim.

The statistical question in Sections 3 and 4 is where information should be collected. The economic question is whether the uncertainty reduction bought by an extra plot is worth its field cost. We use two forms of the problem. The cost-effectiveness form finds the cheapest plot configuration and estimator that meet a stated accuracy target; it needs no money value for map error and matches how monitoring budgets are usually set. The expected-loss form prices a tonne of mis-mapped carbon as a multiple of the carbon price

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

and lets the data choose the campaign size.

Satellite information and field plots are complements in this setting. The satellite prior explains the broad pattern of carbon stock at low cost, and field plots are best spent on the part of the pattern the satellite misses. A plot placed where the satellite is already right buys little, and a plot placed next to another plot buys information the first one already provided. Optimized placement is a way of raising the return on a fixed field budget.

The paper makes four contributions. First, it treats forest-carbon sampling as a constrained purchase of information, in which plots are judged by their marginal value rather than by map quality alone. Second, it prices the information a plot buys in money terms, in the tradition of value-of-information theory [19, 20], and derives a convex cost frontier and a shadow price of accuracy. Third, it evaluates sampling design and estimator choice jointly, under repeated out-of-sample experiments. Fourth, it shows why the current uncertainty rule favours the optimized design near the optimal campaign: the rule rewards a stable spread of plot values, which representative placement delivers (Proposition 4). It also measures the gap between the developer's optimum and the external cost of over-issued credits, and shows that a model-assisted rule and representative placement are complements. Throughout, experiment-based cost savings are kept apart from scenario-based carbon-market figures.

2 Study area and dataset

2.1 Study area

This study was conducted at the Saihanba Mechanical Forest Farm (SMFF) in Weichang County, Hebei Province, northern China (Fig. 1). SMFF is located at the southern edge of the Inner Mongolia Plateau, approximately between 42°02′ to 42°36′ N and 116°51′ to 117°39′ E, with elevations ranging from about 1,010 to 1,940 m above sea level. The region lies in a forest-steppe ecotone and has a cold temperate continental monsoon climate, characterized by long, cold winters and short, cool summers. Long-term records indicate a mean annual temperature of approximately −1.5 °C and mean annual precipitation of about 452.6 mm, with most precipitation occurring during the summer growing season. The forest farm is dominated by plantations established since the 1960s, with Larix principis-rupprechtii and Pinus sylvestris var. mongolica among the principal tree species planted. The farm manages a total area of 93,437.87 ha; official statistics report a forest area of 74,750.30 ha, the area used for scaling in the economic discussion in Section 5.

The experimental area comprised a contiguous 480 m × 480 m forest block subdivided into a regular 24 × 24 grid, yielding 576 fixed 20 m × 20 m reference units

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

and a total area of 23.04 ha. The block has high canopy cover and a relatively homogeneous plantation composition, while retaining substantial variation in tree density, canopy structure, and aboveground carbon stock. This combination of compositional consistency and within-stand structural heterogeneity provides a controlled setting for evaluating the effects of sampling configuration on carbon-stock estimation.

The 20 m × 20 m spatial unit was used consistently throughout the study to align field observations, LiDAR-derived structural information, satellite covariates, and subsequent sampling experiments. The complete 576-unit grid was treated as the reference population. In each sampling experiment, only the selected units were made available to the estimator, while the remaining units were withheld for evaluation. This design allowed alternative sampling strategies and estimators to be compared against the same underlying spatial population and avoided differences caused by changes in reference coverage.

Airborne LiDAR data were acquired over the study area in 2024. The point cloud was preprocessed through ground classification and height normalization and was subsequently used to derive terrain- and canopy-structure products, including a canopy height model (CHM). Field measurements collected in association with the LiDAR campaign included diameter at breast height (DBH) and tree height where field observations were available. These measurements and LiDAR-derived structural variables formed the basis for subsequent plot-level biomass and carbon-stock characterization. Detailed field measurements and LiDAR processing procedures are described in the following sections.

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

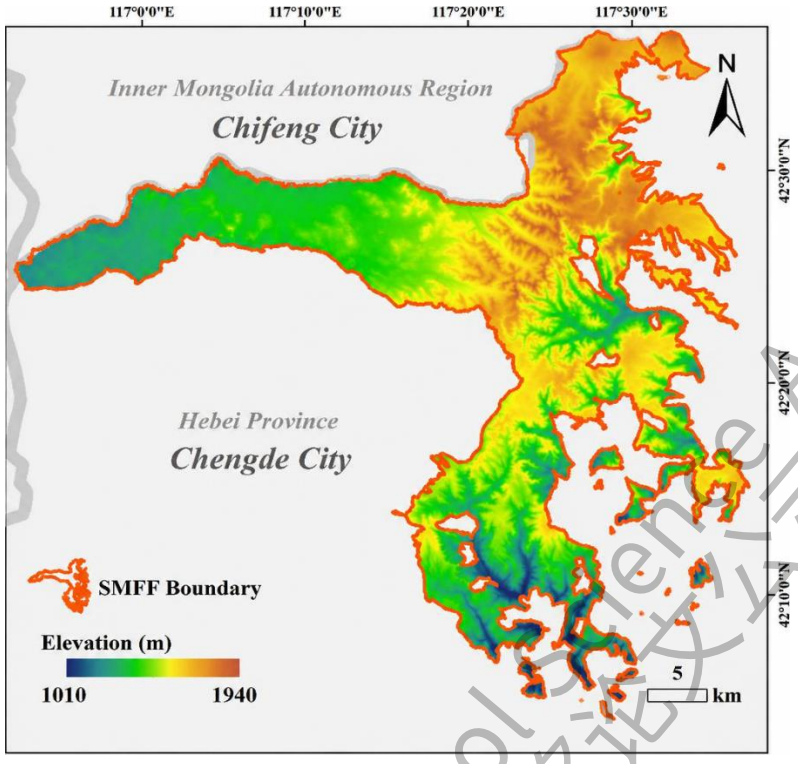


Figure 1 Overview of the Saihanba study area.

2.2 LiDAR-derived reference carbon-stock map

The processing chain is as follows. The airborne LiDAR point cloud yields a canopy height model; individual tree crowns are segmented; single-tree carbon is estimated from crown geometry; each tree is assigned to the pixel containing its crown apex and summed. Pixel carbon density (t C/ha) equals tree carbon (kg) divided by 40 (0.04 ha per pixel). Across the 576 plots in 2024, mean carbon density is 58.85 t C/ha, the standard deviation 17.74, and the coefficient of variation 30.1%. Compared with the highly variable carbon-sink increment, the stock varies less and holds a more stable spatial structure.

We spot-checked the reliability of the LiDAR estimates against field measurements: 101 measured trees (girth converted to DBH, GPS-recorded) were matched to LiDAR trees by crown-polygon IoU, keeping matches above 0.5. The relative RMSE of LiDAR DBH against field values is 6.5% (2024, 88 trees), below the 10% threshold the CCER methodology sets for remote-sensing-assisted estimation. The check has two limits. It stops at DBH, an intermediate variable, so the allometric step from tree size to carbon is not validated; and only 39.6% of the measured trees fall inside the 576-plot area. Every RMSE in this paper is therefore an error against the LiDAR reference map, not against true carbon. Section 4.7 tests how the main comparison changes when the reference map itself is wrong, and Section 6.5 returns to the point.

2026 S. T. Yau High School Science Award 仅用于2026丘成桐中学科学奖论文公示

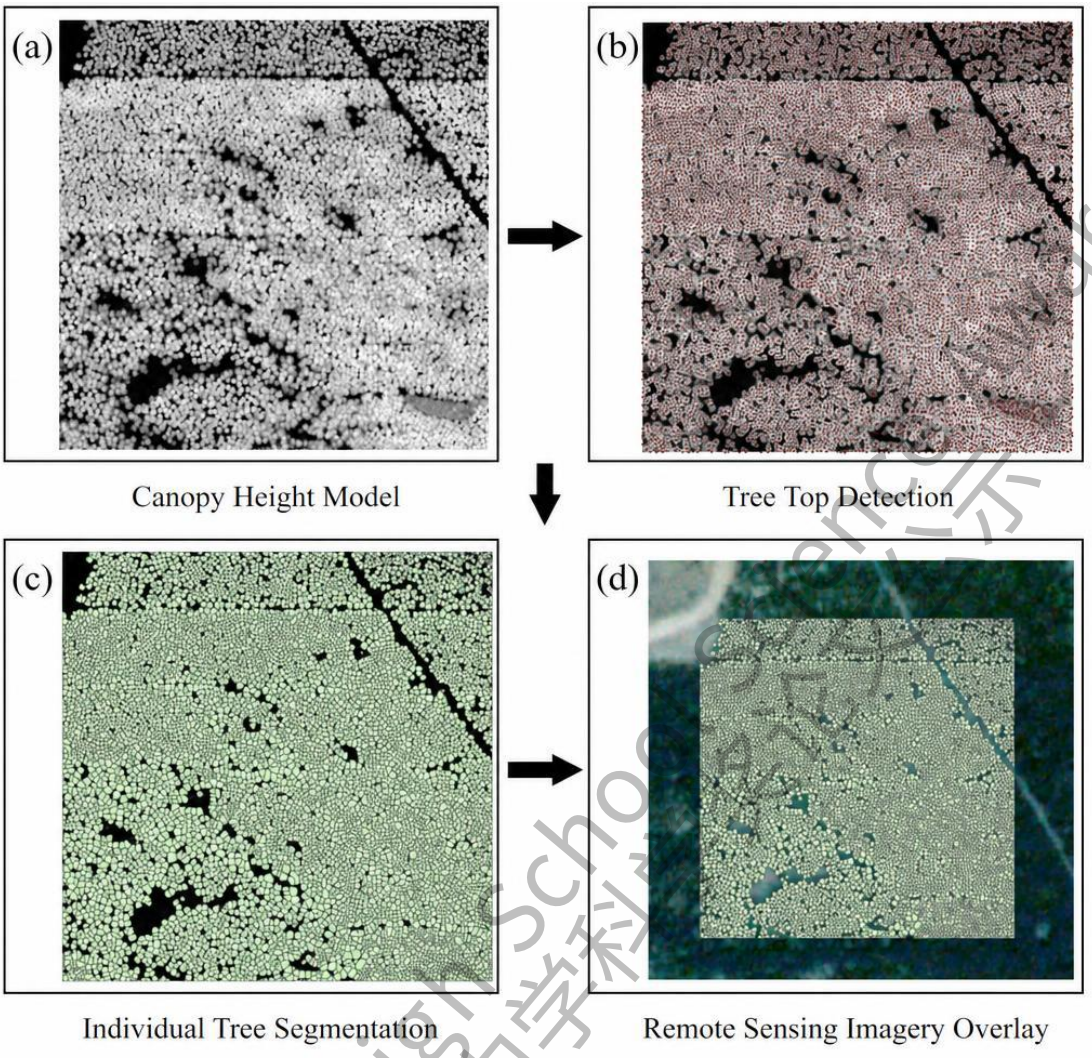


Figure 2 Workflow of individual tree crown segmentation from the CHM

2.3 Satellite imagery and spectral indices

PlanetScope imagery acquired during the 2024 growing season was used to provide wall-to-wall spectral information over the study area. The imagery consisted of eight spectral bands at a spatial resolution of approximately 3 m. Two vegetation indices were retained for subsequent sampling and regression analyses: the normalized difference vegetation index (NDVI) and near-infrared reflectance of vegetation (NIRv). NDVI was calculated from the red and near-infrared bands as

$$NDVI = \frac{NIR-Red}{NIR+Red} \quad (1)$$

and NIRv was calculated as

$$NIRv = NDVI * NIR \quad (2)$$

Spectral values were aggregated to the same 20 m × 20 m reference grid used for the carbon-stock data, ensuring a common spatial support for sampling, model fitting, and evaluation. Across the 576 reference cells, quadratic relationships with carbon stock gave

2026 S. -T. Yau High School Science Award
仅用于2026丘成桐中学科学奖

$R^2 = 0.612$ for NDVI and $R^2 = 0.352$ for NIRv. NDVI and NIRv were themselves strongly correlated ($r = 0.82$), but their imperfect correspondence allowed the two indices to represent partially different aspects of canopy spectral variation.

2.4 Relation between spectral indices and carbon stock

Figure 3 shows the relationships between the two PlanetScope-derived vegetation indices and LiDAR-derived carbon stock across the 576 reference cells. NDVI showed a stronger relationship with carbon stock than NIRv, with quadratic-model R^2 values of 0.612 and 0.352, respectively. The fitted NDVI relationship continued to increase over most of the observed range, indicating that severe high-biomass saturation was not evident within this particular dataset.

Substantial variation in carbon stock nevertheless remained at similar NDVI and NIRv values. Thus, although the spectral indices captured an important component of spatial variation in carbon stock, spectral information alone did not fully explain the reference carbon-stock distribution. This residual variation provides the motivation for explicitly modelling spatially structured prediction errors in the subsequent analysis.

NDVI and NIRv were correlated ($r = 0.82$) but were not interchangeable. We therefore retained both variables in the bivariate sampling design and regression analyses and subsequently tested whether using both indices improved sampling performance relative to an NDVI-only design.

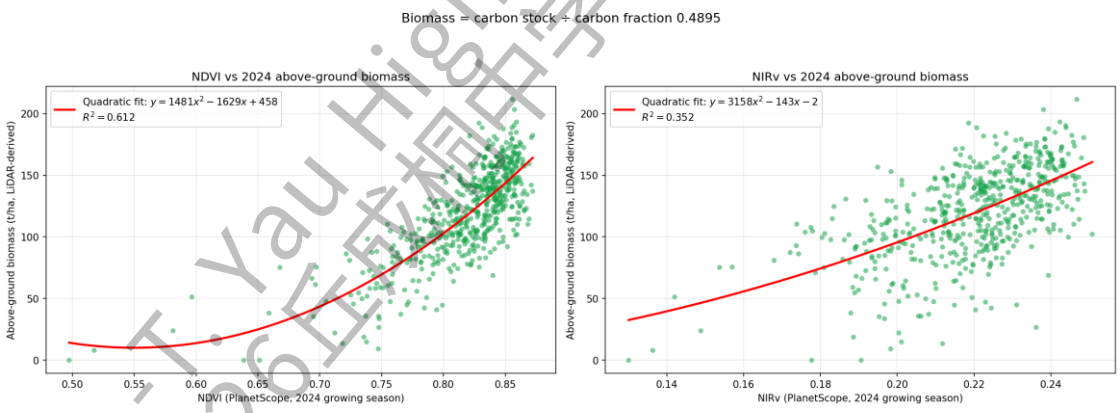


Figure 3 NDVI versus 2024 above-ground biomass for the 576 plots (biomass = carbon stock ÷ carbon fraction 0.4895, so R^2 is the same for carbon; quadratic fit $R^2 = 0.612$; right panel: NIRv, $R^2 = 0.352$)

3 Methods

3.1 Overall design and evaluation

We formulate forest carbon-stock mapping as the combination of a remote-sensing prediction component and a spatial residual component. Let the study domain D consist of N equal-area reference cells, with $Y(s)$ denoting the reference carbon density at location s. For a given sampled set, a regression model based on wall-to-wall remotely sensed covariates provides an initial prediction $R(s)$. The corresponding residual field is

2026 S.T. Yau High School Science Award 仅用于2026丘成桐中学科学奖公示

defined as

$$E(s) = Y(s) - R(s), Y(s) = R(s) + E(s) \quad (3)$$

Under this formulation, remotely sensed covariates explain the systematic component of carbon-stock variation, while sampled reference observations provide information on both regression calibration and the remaining spatially structured residuals. For equal-area cells, the regional mean can be written as

$$\mu_Y = \frac{1}{N}\sum_{i=1}^{N} Y_i = \mu_R + \mu_E \quad (4)$$

Prediction accuracy was evaluated only at unsampled reference cells. Let U denote the set of unsampled cells. Mapping accuracy was quantified using

$$\mathrm{RMSE} = \sqrt{\frac{1}{|U|}\sum_{s\in U}\left(Y(s) - \hat{Y}(s)\right)^2} \quad (5)$$

The main experiment followed a full-factorial design comprising three sampling strategies, seven estimators, and three sample sizes ($n = 27$,54 and 90). Each combination was repeated across 100 random realizations, yielding 6,300 sampling-estimation runs; the within-zone contiguous control was evaluated with E4 only (Section 4.4). Within each run, the sampled cells were used for model fitting and spatial estimation, whereas all remaining cells were withheld from fitting and used for evaluation. This ensured that all sampling designs and estimators were compared against the same reference population without using evaluation values during model construction.

In addition to the main factorial experiment, we evaluated performance across a wider range of sample sizes to characterize the relationship between sampling effort and prediction error. Additional sensitivity analyses examined the effects of alternative model specifications and univariate versus bivariate sampling attributes.

3.2 Zonal SMP-cLHS sampling

The optimized sampling design was adapted from the SMP framework of Zeng et al. (2015), which jointly considers representativeness in auxiliary-variable space and spatial dispersion. SMP has subsequently been used in China's national standard for leaf area index (LAI) product validation [1,2]. We re-implemented the sampling procedure in Python following the original objective-function formulation:

For a candidate sample set S, attribute representativeness was quantified using the discrepancy term

$$\mathrm{dif}(S) = \frac{1}{n}\sum_{i=1}^{n}|c_i - 1| \quad (6)$$

where the population distribution of each sampling attribute was divided into n equal-frequency classes and c_i denotes the number of selected cells falling in class i. A perfectly balanced Latin-hypercube representation therefore gives $dif = 0$. The discrepancy was

2026 S.T. Yau High School Science Award 仅用于2026丘成桐中学科学奖公示

calculated separately for NDVI and NIRv and combined in the sampling objective.

Spatial dispersion was characterized using the nearest-neighbour index (NNI):

$$\mathrm{NNI}(S) = \frac{\overline{d}_{nn}(S)}{E[d]}, E[d] = \frac{1}{2\sqrt{\rho}}, \rho = \frac{k}{|D|} \tag{7}$$

where $\overline{d}_{nn}(S)$ is the mean nearest-neighbour distance among selected cells, k is the number of selected cells, and $|D|$ is the area of the sampling domain. Values of $NNI > 1$ indicate greater spatial dispersion than expected under complete spatial randomness.

Attribute representativeness and spatial dispersion were combined through

$$\mathrm{OF}(S) = \frac{\mathrm{dif}(S)}{\mathrm{NNI}(S)} \tag{8}$$

which was minimized using simulated annealing. Candidate swaps that reduced the objective were preferentially accepted, while the annealing schedule also allowed occasional uphill moves to reduce sensitivity to local minima. The initial temperature was 1, temperature was reduced by 5% every 10 iterations, and optimization was terminated after 10,000 iterations or when the stopping criterion fell below 0.01.

To ensure spatial coverage across the study block, we introduced an additional zonal constraint. The 24 × 24 reference grid was divided into nine 8 × 8 zones. Each zone contributed an equal number of sampled cells, $q \in \{3,6,10\}$, corresponding to total sample sizes of $n = 27,54$, and 90, respectively. Simulated-annealing swaps were restricted to cells within the same zone, so that every final design contained exactly q sampled cells in each of the nine zones. This constraint prevented sample concentration in a small portion of the study area while preserving SMP optimization within each zone.

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

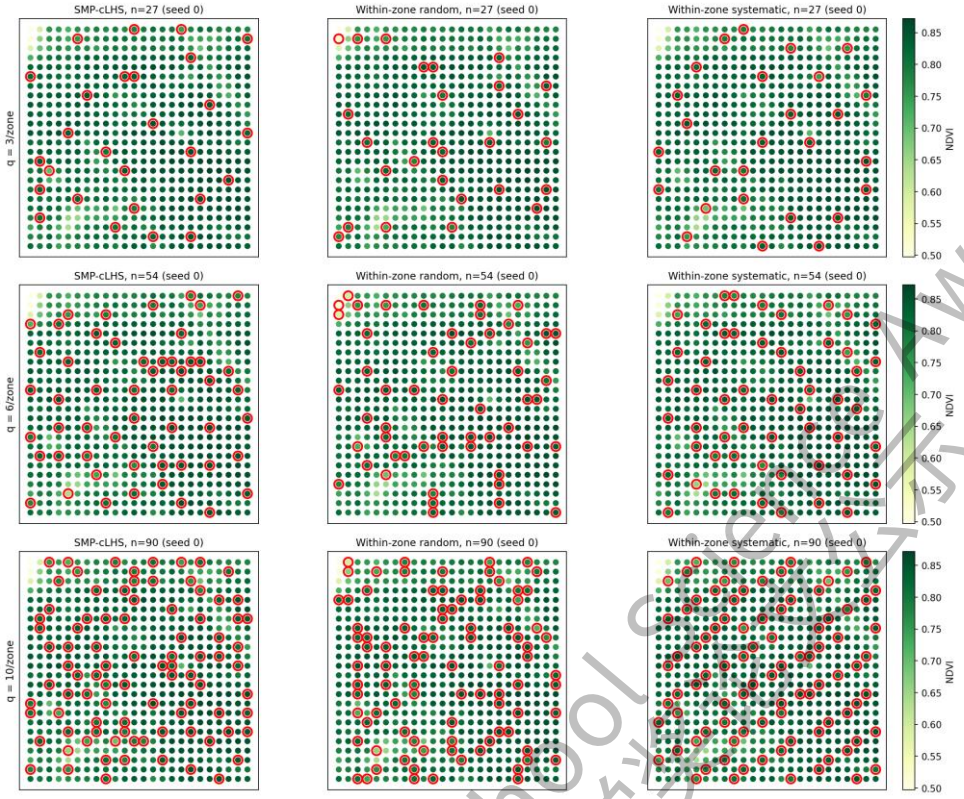


Figure 4 Sample layouts of the three sampling designs (columns: SMP, within-zone random, within-zone systematic; rows: n = 27, 54, 90; background is NDVI, pale patches are low-NDVI sparse areas, red circles mark selected plots)

Figure 4 shows what histogram matching does to plot placement. With random and systematic designs, small samples leave the low-NDVI sparse patches (pale areas) under-covered; SMP raises coverage of exactly those areas. Figure 5 quantifies the quantile-bin matching, and Figure 6 traces one annealing run.

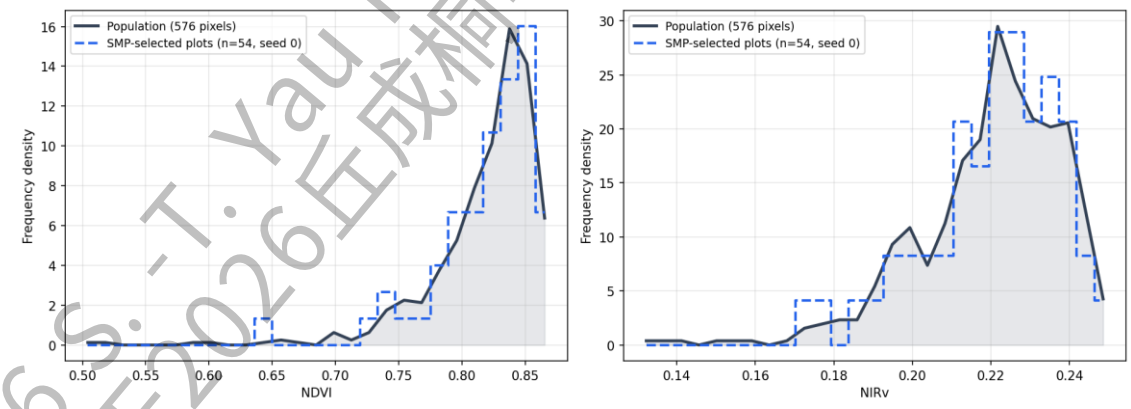


Figure 5 Frequency distributions of the population and the SMP sample (left: NDVI; right: NIRv; solid: all 576 pixels; dashed: the 54 SMP-selected plots; the agreement of the two curves is the attribute representativeness)

2026 S.-T. Yau High School Award
仅用于2026丘成桐中学科学奖

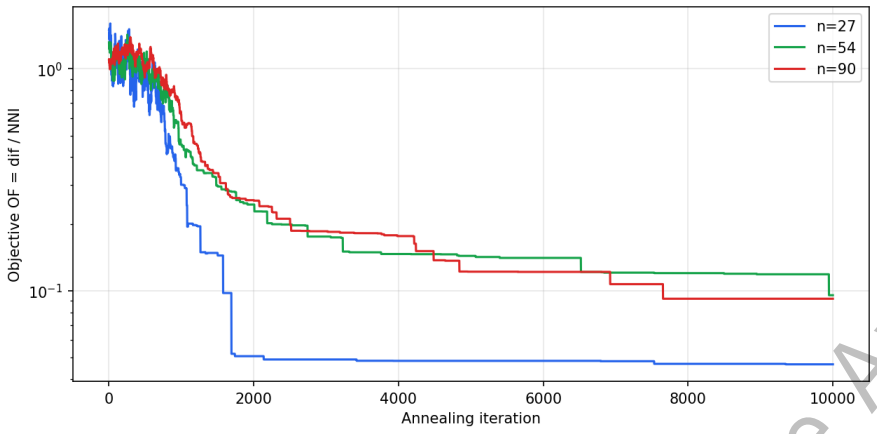


Figure 6 Convergence of the simulated-annealing objective (one typical run each at n = 27, 54, 90; log scale)

3.3 Seven estimators

The seven estimators span the full spectrum from attribute-only to space-only to fusion of both, and E5, E6, and E7 are fusion estimators built in this study. Every model is fitted on the n sampled plots alone, and no unsampled reference value enters training, parameter estimation, or weighting:

Their formal definitions follow. E1 (pure regression) uses the full quadratic in the two satellite-observable indices as the prior:

$$R(s) = \beta_0 + \sum_{j=1}^{2} \left[\beta_{1j} x_j(s) + \beta_{2j} x_j^2(s)\right] \tag{9}$$

E2 (inverse-distance weighting) weights plot observations by the inverse p-th power of distance:

$$\hat{z}_{IDW}(s_0) = \frac{\sum_{i \in S} d(s_0, s_i)^{-p} z(s_i)}{\sum_{i \in S} d(s_0, s_i)^{-p}} \tag{10}$$

Table 1 Definition of the seven estimators

No.	Method	Information used
E1	Quadratic regression on NDVI+NIRv	Attribute only (spectral)
E2	IDW interpolation (p = 2; known points take LiDAR values)	Space only
E3	Ordinary kriging (de-mean, interpolate, add back)	Space only
E4	E1 regression surface + kriged residual correction (additive fusion)	Additive fusion
E5	w·E1 + (1−w)·E2, w a global LOO weight	Weighted fusion
E6	w·E1 + (1−w)·E3, w a global LOO weight	Weighted fusion
E7	Per-pixel fusion of E1 and E3 weighted by error maps, w(s) ∝ 1/residual²(s)	Per-pixel weighted fusion

The covariance structure of the spatial components (E2/E3/E4) follows a nugget-plus-exponential variogram:

$$\gamma(h) = \tau^2 + \sigma^2\left(1 - e^{-h/\ell}\right) \tag{11}$$

E3 (ordinary kriging) performs optimal linear unbiased interpolation on the de-meaned field, with the kriging variance giving the theoretical uncertainty:

$$\hat{z}_K(s_0) = \sum_{i=1}^{n} \lambda_i z(s_i), \mathrm{Var}_K(s_0) = (\tau^2 + \sigma^2) - c^{\top} C^{-1} c \tag{12}$$

2026 S. T. Yau High School Science Award 仅用于2026丘成桐中学科学奖公示

E4 (regression plus kriged residuals) interpolates the residuals by ordinary kriging on top of the regression prior. This is the core step of spatial error sampling: a full-coverage prior R as the base map, plots learning the error field:

$$\hat{Y}(s_0) = R(s_0) + \sum_{i \in S} \lambda_i (s_0)(Y(s_i) - R(s_i)) \quad (13)$$

with weights λ solved from the covariance equations induced by Eq. (11). The E5 and E6 built in this study weight the regression and spatial information by the inverse LOO error variances:

$$w = \frac{1/v_{reg}}{1/v_{reg} + 1/v_{space}} \quad (14)$$

E7 pushes the weighting down to individual pixels, fusing point by point according to the regression error map and the kriging variance map:

$$w(s) = \frac{1/\hat{e}^2_{reg}(s)}{1/\hat{e}^2_{reg}(s) + 1/\sigma^2_{krig}(s)} \quad (15)$$

Three implementation details matter. First, the variogram fit (Eq. (11); [10]). Full-sample diagnostics put the correlation of 2024 regression residuals at only 0.20 at 20 m lag (γ (24 m)/sill $\approx$ 0.83), so residuals are nugget-dominated; a no-nugget exponential does not fit, its range collapses to a few meters and interpolation does nothing. On an n = 54 SMP sample the direct variogram of carbon is $\tau^2 = 144$, $\sigma^2 = 167$, range $\ell = 200$ m (long-range structure, consistent with Moran's I = 0.43 [11]); regression residuals give $\tau^2 = 12$, $\sigma^2 = 88$, $\ell = 1.7$ m. Second, E3 de-means before interpolation. With a mean near 59 t C/ha, zero-mean simple kriging shrinks systematically (bias about −10 t C/ha before correction, about −2.6 after de-meaning and adding back). Third, the weights in Eqs. (14) and (15) come from the data with no hand tuning: the kriging variance map σ^2_krig(s) = $(\tau^2 + \sigma^2) - c^T C^{-1} c$ gives the theoretical interpolation error growing away from plots, and the regression error ê_reg(s) spreads the LOO squared residuals of the n plots by IDW （floored at 25% of the global mean squared residual to avoid singularity）. Substituting into Eq. (15), pixels near plots lean toward spatial interpolation and far pixels lean toward regression; the global weights of E5 and E6 likewise follow Eq. (14) from LOO error variances.

2026 S.T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

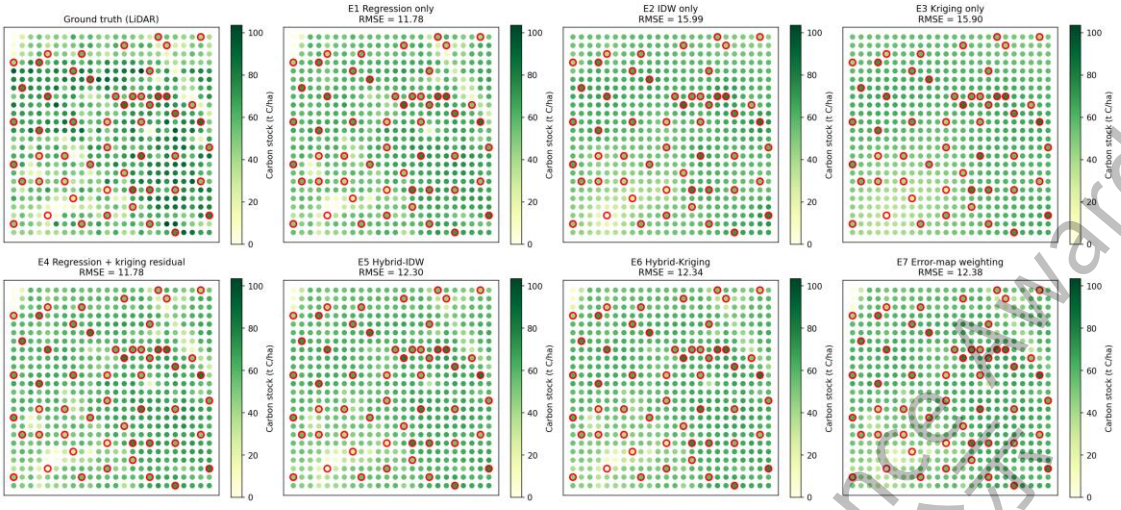


Figure 7 Maps of the seven estimators against ground truth (panels left to right: truth, then E1 through E7; n = 54, SMP, seed 0)

3.4 Control designs

Three controls share the same zonal quotas as SMP, to isolate the contribution of histogram matching plus dispersion optimization: within-zone random (equal random draws per zone), within-zone systematic (evenly spaced in row-major order), and within-zone contiguous (one q-pixel rectangle per zone: 1×3 for 3, 2×3 for 6, 2×5 for 10, placed at random). The contiguous design mimics the field habit of measuring neighboring plots in one trip. A fourth baseline is current CCER practice: whole-area simple random sampling estimating only the mean, whose mapping equivalent is the mean-flat map filled with the sample mean, RMSE about 17.8 to 18.2 and nearly constant in n. The mean criterion does improve with more plots (Table 5), but pixel-wise RMSE splits into the spatial standard deviation of carbon (a constant near 17.7) plus a shrinking mean-estimation term, and the constant dominates, so the curve stays flat.

3.5 Data boundary

Every experiment uses only information observable in 2024: PlanetScope indices over the full area, plus 2024 LiDAR values at the sampled plots. The 2018 plot table in the dataset was back-derived from the 2024 data, so it never enters any prior, to avoid leakage. The same rule governs the sampling attributes: only satellite-observable NDVI and NIRv may enter the dif score, and no carbon or LiDAR structural information may influence plot selection.

3.6 Theory: error propagation, information gain, and plot value

The combination of estimators and sampling designs above follows from Gaussian-process conditional inference, which also yields three propositions that can be checked on the Saihanba data. Let D be the study domain, with area |D| and N = 576 cells. Model the error field $E = (E(s_1), \ldots, E(s_N))^{\top}$ as a zero-mean Gaussian process with stationary

2026 S. Tuyau High School Science Award
仅用于2026五成川学科学术论文公示

covariance function C(·), so that $\Sigma = [C(s_i - s_j)]$ is its N × N prior covariance matrix. For a plot set S, Σ_{SS} is the covariance among the plots, $\Sigma_{:S}$ the N × |S| covariance between every cell and the plots (the colon means "all cells"; $\Sigma_{S:} = \Sigma_{:S}^{\top}$), and $\Omega_{SS} = \mathrm{diag}(\omega_j^2)$ the field-measurement error variance at the plots. The regional mean is $\mu_E = a^{\top}E$, where $a = (1/N, \ldots, 1/N)^{\top}$ holds the equal-area weights of Eq. (4).

Proposition 1 (Spatial aggregation error). For a second-order stationary field, the variance of the regional mean of the error field is the average of its covariance matrix; on a continuous domain it becomes a double integral of the covariance function:

$$\mathrm{Var}(\overline{E}_D) = \frac{1}{N^2}\mathbf{1}^{\top}\Sigma\mathbf{1} \approx \frac{1}{|D|^2}\int_D \int_D C(s-t)\,dsdt \approx \frac{1}{|D|}\int_{\mathbb{R}^2} C(h)\,dh \tag{16}$$

The last approximation holds when the region is much larger than the correlation length (|D| ≫ ℓ²). **Proof.** Take the exponential kernel

$$C(h) = \sigma^2 \exp(-\frac{\|h\|}{\ell}) \tag{17}$$

with sill σ² and correlation length ℓ. In polar coordinates $\int_0^{\infty} e^{-r/\ell}\, r dr = \ell^2$, so $\int_{\mathbb{R}^2} C(h)\,dh = 2\pi\sigma^2\ell^2$. Substituting into Eq. (16) gives

$$\mathrm{Var}(\overline{E}_D) \approx \frac{2\pi\sigma^2\ell^2}{|D|}, N_{\mathrm{eff}} = \frac{\sigma^2}{\mathrm{Var}(\overline{E}_D)} \approx \frac{|D|}{2\pi\ell^2} \tag{18}$$

where N_{eff} is the number of independent cells that would give the same variance of the mean. On the discrete grid the exact value is $N_{\mathrm{eff}} = N^2/(\mathbf{1}^{\top}R\mathbf{1})$, with R the correlation matrix, so N_{eff} can never exceed N. This completes the proof.

Plugging in the Saihanba parameters shows what this buys. For the carbon field ℓ ≈ 200 m on a 480 m block, so |D| ≫ ℓ² does not hold and Eq. (18) is only a rough guide ($N_{\mathrm{eff}} \approx 0.9$). The exact value on the 576-cell grid is $N_{\mathrm{eff}} \approx 3.0$: because their errors move together, the 576 map pixels carry about three independent observations of the regional mean. The regression-residual field has ℓ ≈ 1.7 m, far below the 20 m cell size, so at the scale of the grid it behaves as white noise and $N_{\mathrm{eff}} \approx 576$, one independent observation per cell (Figure 8). Two consequences follow for Section 4. Adding remote-sensing pixels cannot tighten the regional mean much, because their errors move together. And the residual left after the regression has almost no spatial structure for kriging to exploit, so the plots earn their value mainly by calibrating the satellite regression, which is why placing them across the range of the satellite indices matters.

Corollary 1 (Effective sample size). The effective number of independent observations in wall-to-wall information is $N_{\mathrm{eff}} = \sigma^2/\mathrm{Var}(\overline{E}_D)$. It is close to $|D|/(2\pi\ell^2)$ when |D| ≫ ℓ² and is bounded above by the number of cells. More pixels do not buy proportionally more information (Figure 8b).

Proposition 2 (Conditioning weakly reduces posterior variance under the

2026 S.-T. Yau High School Science Award 仅用于2026丘成桐中学科学奖论文公示

assumed covariance model). Given a plot set S, the conditional covariance of the error field follows the standard Gaussian-process update:

$$\Sigma_{E|S} = \Sigma - \Sigma_{:S}(\Sigma_{SS} + \Omega_{SS})^{-1}\Sigma_{S:} \tag{19}$$

For the regional mean $\mu_E = a^\top E$, with $a = (1/N, \ldots, 1/N)^\top$ the vector of equal-area weights (every entry 1/N, so $a^\top E$ is the plain average of the cells), the posterior variance is

$$V(S) = a^\top \Sigma_{E|S} a = V(\emptyset) - G(S), G(S) = a^\top \Sigma_{:S}(\Sigma_{SS} + \Omega_{SS})^{-1}\Sigma_{S:}a \tag{20}$$

The second term, G(S), is the information gain of S. Since $\Sigma_{S:}a$ is the vector of covariances between each plot and the regional mean, G(S) is a quadratic form in that vector, not a sum over cells. **Proof.** Write $b = \Sigma_{S:}a$ and $M = \Sigma_{SS} + \Omega_{SS}$. M is positive definite, hence so is its inverse, and $G(S) = b^\top M^{-1} b \geq 0$, with equality if and only if b = 0, that is, if the plots are uncorrelated with the regional mean. Hence V(S) ≤ V(∅). This completes the proof.

Under the assumed covariance model, adding ground plots to remote sensing can only lower the variance of the regional-mean estimate. E4 (Eq. (13)) carries out this conditional inference on the prior-plus-residual decomposition, and E5/E6/E7 (Eqs. (14)(15)) approximate it with weights.

Scope of Proposition 2. The result concerns the model-implied posterior variance under a correctly specified covariance structure. It does not guarantee a lower out-of-sample RMSE when covariance parameters are estimated from a few plots or when the model is misspecified, so empirical performance is judged separately on held-out cells (Section 4).

Proposition 3 (Plot value). Adding a candidate plot j to an existing set S reduces the posterior variance of the regional mean by

$$\Delta(j \mid S) = V(S) - V(S \cup \{j\}) = \frac{(a^\top \Sigma_{:j|S})^2}{\Sigma_{jj|S} + \omega_j^2} \tag{21}$$

Proof. Here $\Sigma_{:j|S}$ is the j-th column of the posterior covariance in Eq. (19), $\Sigma_{jj|S}$ the posterior variance at the candidate cell before it is measured, and ω_j^2 its measurement-error variance. Both are posterior quantities and differ from the prior covariance Σ of Eq. (16); ω² is the field-measurement error and is unrelated to the variogram nugget τ² of Eq. (11). Adding j is a rank-one update of Eq. (19), and the Sherman-Morrison formula gives the drop in Eq. (21) directly. This completes the proof.

Corollary 2 (Plot selection rule). The value of a plot is its squared posterior covariance with the regional mean divided by the total noise of the new observation. A good plot sits where uncertainty is still high, covaries with errors over the whole area, does not duplicate information already in hand, and can be measured precisely. The

2026 S. T. Yau High School Science Award 仅用于2026丘成桐中学科学奖论文公示

optimal layout is $S^* = \mathrm{argmax}_{|S|=k} G(S)$.

The SMP objective (Eq. (8)) is an operational approximation of this principle in attribute space: histogram matching sends plots into attribute intervals where the prior is uncertain, and NNI dispersion avoids duplicate information; the zonal quota (Section 3.2) plays the role of a coverage constraint. The three propositions and two corollaries place sampling design, estimator choice, and budget allocation on one information-theoretic footing (Figure 8).

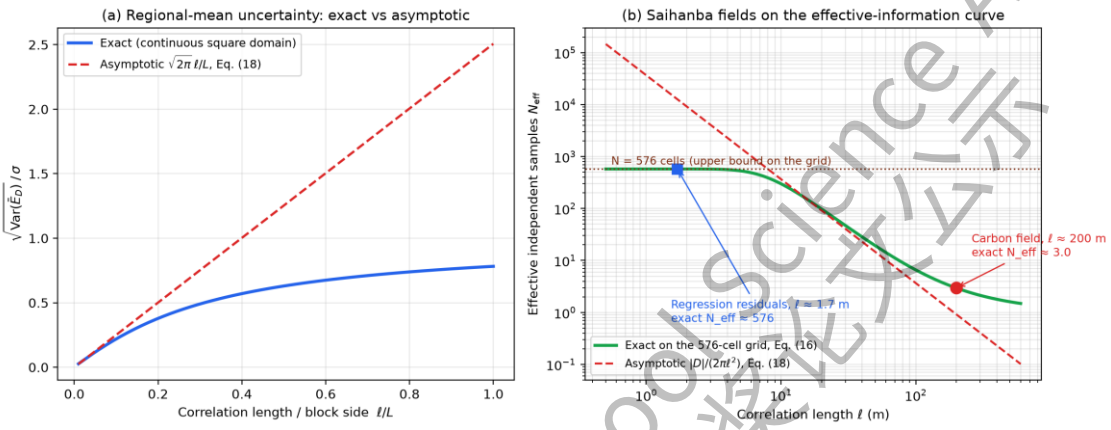


Figure 8 Effect of spatial correlation on regional-mean uncertainty and effective sample size (left: exact value on a continuous square domain versus the asymptotic formula of Eq. (18); right: exact N_{eff} on the 576-cell grid, Eq. (16), and the asymptotic curve, with the Saihanba carbon and residual fields)

3.7 Economic framework: cost-effectiveness and the marginal value of a plot

Decision variables and budget. Let S be the set of plots, n = |S| its size, e the estimator, and c_j the full field cost of plot j (travel, crew time, measurement and quality control). The campaign must respect the budget B:

$$\sum_{j \in S} c_j \leq B \tag{22}$$

Expected total verification loss. The developer is a single price-taking firm, so the model is microeconomic. It chooses the plot set and the estimator to minimize the expected sum of three losses, all in CNY: field measurement cost, the credit deduction triggered by the uncertainty rule, and the cost of carbon that the map puts in the wrong place:

$$\mathbb{E}[L(S,e)] = \sum_{j \in S} c_j + pQ\mathbb{E}[D(u(S))] + \lambda_C A_m \mathrm{RMSE}(S,e) \tag{23}$$

Here p is the carbon price in CNY per t CO2e, Q the creditable quantity in t CO2e, D(·) the deduction share in Table 35 of the CCER methodology [3], and u(S) the uncertainty computed by its Appendix F formula from the plot values. In the third term A_m is the mapped area in hectares, so A_m·RMSE turns an error in t C/ha into tonnes of carbon, and λ_C is the money value of one tonne of mis-mapped carbon. RMSE is used rather than mean absolute error because it weights large local errors more heavily, and

2026 S. T. Yau High School Science Award 仅用于2026丘成桐中学科学奖公示

large errors are the ones that misdirect management or an audit; since RMSE ≥ MAE, A_m·RMSE is an upper bound on the absolute carbon misallocated. We anchor λ_C to the market: $\lambda_C = \kappa \cdot p_C$, where $p_C = (44/12)\cdot p$ is the carbon price per tonne of carbon (1 t C = 3.67 t CO2e). With κ = 1 a mis-mapped tonne costs its market value; κ > 1 adds the cost of decisions taken on the wrong stands. The first two terms are observable: plot cost from a fee guideline, and E[D(u)] from repeated draws. κ is not, so Section 5.4 solves Eq. (23) for κ = 0, 1, 2 and 5.

Cost-effectiveness problem. Eq. (23) has a useful special case. When the deduction term is held fixed and the map-error value is unknown, the developer chooses the cheapest configuration that meets an accuracy target A. Two targets are used, because two users need different things. A map user (forest management, targeted verification, risk assessment) needs pixel-level accuracy, measured by the mapping RMSE of Section 3.1. The accounting rules need the project mean. For the map:

$$C^*(A) = \min_{S,e} \sum_{j \in S} c_j \quad \text{subject to} \quad \mathrm{RMSE}(S, e) \le A \tag{24}$$

For the project mean, a single campaign is one random draw, so the target is stated as a probability rather than an average:

$$\Pr\left(\frac{|\hat{\mu}(S,e) - \mu|}{\mu} \le u^*\right) \ge 1 - \alpha \tag{25}$$

with u^* the developer's target relative error of the project mean, set to 5% in the results, and 1 − α the required assurance, set at 80% and 95%. The 5% level is a margin a developer may choose in order to stay below the 10% line at which Table 35 of the CCER methodology [3] begins to discount credits; it is not itself a regulatory threshold.

Shadow price of accuracy. With equal plot costs, c_j = c, problem (24) has the Lagrangian L = c·n + η(RMSE(S, e) − A), and the multiplier $\eta = -dC^*/dA$ is the shadow price of accuracy: the extra field cost of tightening the map target by one unit. We write η, not λ, to keep it apart from the kriging weights of Eq. (13) and from the map-error value λ_C. $C^*(A)$ cannot be differentiated on a grid of sample sizes, so we use its discrete approximation between two sample sizes $n_1 < n_2$, and in Section 5.2 a fitted continuous frontier:

$$\eta \approx \frac{c(n_2 - n_1)}{\mathrm{RMSE}(n_1) - \mathrm{RMSE}(n_2)} \tag{26}$$

in CNY per t C/ha of RMSE. Dividing by A_m gives CNY per tonne of mis-mapped carbon, which can be compared directly with λ_C: adding plots pays off when $\lambda_C > \eta/A_m$. A rising η is the economic form of diminishing returns.

Marginal value of a plot. Proposition 3 gives Δ_j, the reduction in the posterior variance of the regional mean when plot j is added. It is the information a plot buys for the mean target (25), in units of (t C/ha)². The map target (24) is measured in RMSE, a

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公开

different quantity, so we keep the two apart: Proposition 3 motivates where plots go, and η prices map accuracy.

What the SMP objective optimizes. The SMP objective (Eq. (8)) rewards feature-space representativeness and spatial dispersion. It contains no money term, so it is not a direct solution of problem (24). It is an operational approximation: when plot costs are equal, a design that buys more information per plot reaches A with fewer plots. Section 5 tests whether it does, against random placement and against current CCER practice, and checks the equal-cost assumption by measuring the field route each design requires.

Credit deductions. Table 35 of the CCER methodology [3] discounts issued credits by 6% when the uncertainty computed by its Appendix F formula lies between 10% and 20%. That formula depends on the number of plots and their spread, not on the estimator, so a better estimator alone does not change the reported value. Plot placement does change it, because it changes the spread of the plot values. Section 5.4 measures this effect for each design.

4 Results

Section 4 in brief. For readers mainly interested in the economics, the statistical results come down to four facts. First, with few plots, where the plots go matters more than which estimator is used: at 27 plots optimized placement cuts mapping RMSE by about a fifth against random placement (13.48 against 17.24 t C/ha, 100 draws). Second, the gain shrinks as plots become plentiful and is gone by 90 plots. Third, the plots mainly calibrate the satellite regression; interpolating the remaining residuals adds little. Fourth, on the official criterion, the project mean, the optimized design also has a smaller error than current practice. Section 5 puts money values on these facts.

4.1 Full-factorial main results

Table 2 lists the mapping RMSE of all 21 sample-size-by-estimator treatments under three designs (mean over 100 draws, unsampled-pixel criterion). Under SMP the best value at every n comes from E4; under random and systematic designs the E5/E6/E7 hybrids close much of the gap to SMP; the space-only estimators (E2/E3) trail the regression family (E1) in every combination.

Table 2 Full-factorial results: mapping RMSE (t C/ha, mean of 100 draws, unsampled pixels)

n	Estimator	SMP-cLHS	Zone random	Zone systematic
27	E1 Regression	13.50	17.25	16.92
27	E2 IDW	16.60	16.82	16.58
27	E3 Kriging	16.76	17.05	16.74
27	E4 Reg.+kriged resid.	13.48	17.24	16.90
27	E5 Hybrid-IDW	13.83	14.78	14.59
27	E6 Hybrid-kriging	13.95	14.87	14.66
27	E7 Error-map weight	13.61	15.07	14.52

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

54	E1 Regression	11.55	13.72	12.41
54	E2 IDW	15.49	15.98	15.61
54	E3 Kriging	15.77	16.23	15.82
54	E4 Reg.+kriged resid.	11.44	13.60	12.30
54	E5 Hybrid-IDW	12.10	12.72	12.16
54	E6 Hybrid-kriging	12.26	12.75	12.25
54	E7 Error-map weight	12.07	12.68	12.29
90	E1 Regression	11.21	11.46	11.37
90	E2 IDW	15.06	15.41	15.13
90	E3 Kriging	15.24	15.60	15.31
90	E4 Reg.+kriged resid.	11.09	11.26	11.21
90	E5 Hybrid-IDW	11.57	11.58	11.51
90	E6 Hybrid-kriging	11.64	11.62	11.58
90	E7 Error-map weight	11.66	11.61	11.59

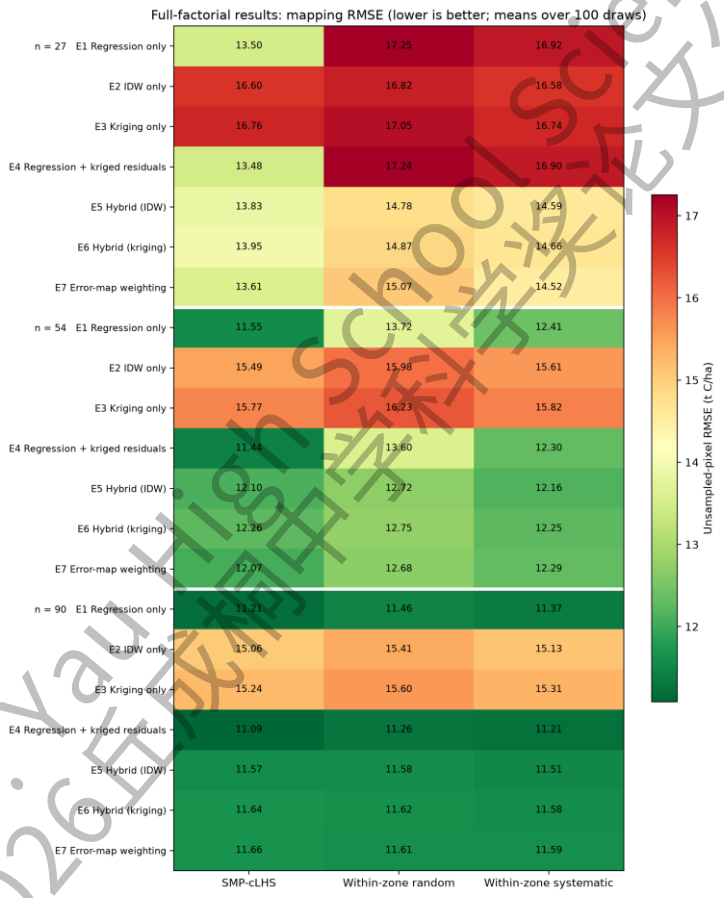


Figure 9 Summary heatmap of full-factorial results (100 draws; green means lower RMSE)

4.2 SMP gains: size, source, and decay

Converting Table 2 into the improvement of SMP over within-zone random (Table 3), the gain is defined as:

$$\text{Gain}(n) = \frac{\text{RMSE}_{rand}(n) - \text{RMSE}_{SMP}(n)}{\text{RMSE}_{rand}(n)} \times 100\% \quad (27)$$

the subscripts naming the sampling design and the expectation taken over 100 random draws. The regression family (E1/E4) gains most: about 22% at n = 27, 16% at n = 54,

2026 S.-T. Yau High School Science Award 仅用于2026丘成桐中学科学奖论文公示

and only 1.5% to 2.2% at n = 90. The space-only estimators (E2/E3) gain 1% to 3%, with the hybrids in between at small n.

Table 3 RMSE improvement of SMP over the within-zone random baseline (positive favors SMP)

n	Regression E1/E4	Spatial E2/E3	Hybrid E5/E6/E7
27	+21.7% / +21.8%	+1.3% ~ +1.7%	+6.2% ~ +9.7%
54	+15.8% / +15.9%	+2.8% ~ +3.1%	+3.9% ~ +4.8%
90	+2.2% / +1.5%	+2.3% ~ +2.3%	−0.4% ~ +0.1%

The gains concentrate at small n and decay as n grows. The main risk of regression on small samples is a training distribution that drifts from the population: samples crowding the mid NDVI range force extrapolation at both ends. Histogram matching removes exactly this risk by forcing the sample's attribute coverage onto the population's. By n = 90 a random sample already covers the range, and the marginal return of optimization shrinks. Space-only estimators gain little from SMP because NNI dispersion and interpolation want overlapping things, and IDW and kriging are insensitive to exact locations as long as plots do not clump.

Figure 10 shows the three designs differently. Each plot's territory, the set of pixels closest to it, is colored by the plot's carbon value: SMP territories are small and even; zone random throws off large pale territories where one low-carbon plot commands too much area; contiguous territories are large and skewed, with corners reaching deep into unsampled regions.

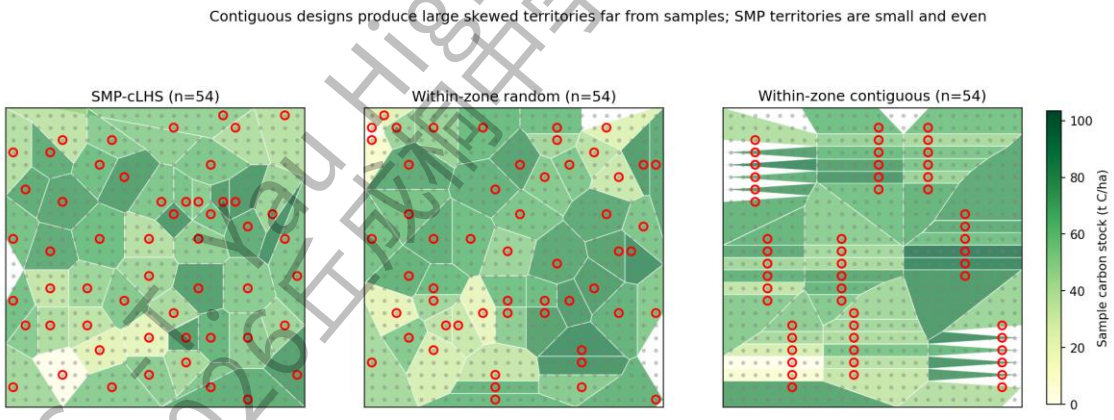


Figure 10 Voronoi territories of the plots, colored by plot carbon (SMP, within-zone random, within-zone contiguous)

Figure 11 shows where the regression errors are. NDVI and carbon are each split into tertiles and crossed into nine color classes; most pixels sit on the diagonal, high NDVI with high carbon, which is where regression works. Off-diagonal blocks are the exceptions: high NDVI but low carbon, or the reverse. These anomalous pixels are the main source of regression error, and the SMP plots (red circles) deliberately cover them. So histogram matching also brings these relation anomalies into the sample.

2026 Siyuan High School Science Award 仅用于2026年思源中学科学论文奖

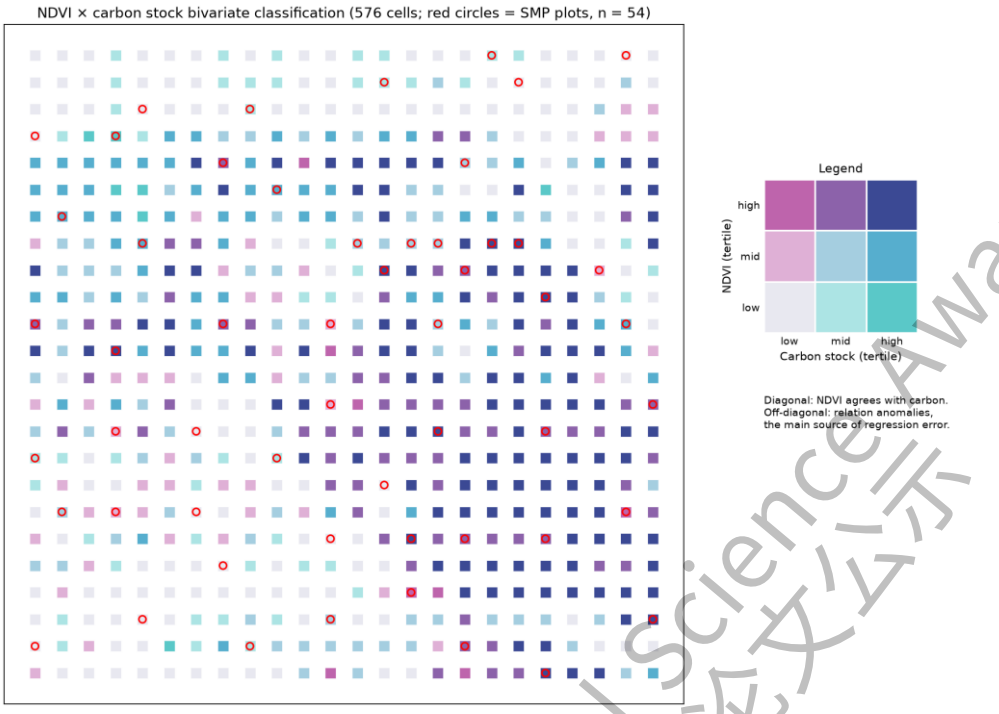


Figure 11 Bivariate NDVI × carbon classification map (each split into tertiles, nine classes; red circles are SMP plots)

The zone-level check (full data in Section 4.6) adds a local view. At n = 27 random sampling lets the error grow sharply in the most heterogeneous zone Z1, where E4 reaches RMSE 31.5, and SMP cuts it by about 40% (to 19.0). Besides lowering average RMSE, SMP removes the local failures that random sampling produces at small n.

4.3 Ablation across the estimator spectrum

Data decides the balance between attribute and spatial information. Here the NDVI prior has $R^2 \approx 0.63$ while 54-point pure spatial interpolation reaches only 0.22 to 0.23, so E1 dominates E2/E3. A parallel analysis of the carbon-stock increment on the same plots, not reported here, found the reverse (prior $R^2 = 0.03$, spatial correction dominant): change the target and the two information types swap places.

E4 is best in 5 of the 9 sampling-by-size combinations: all three sample sizes under SMP and n = 90 under both zone designs. Its kriging correction is small, within 0.02 to 0.16 of E1 in Table 2, but consistently positive. The variogram explains it (Figure 12): the direct carbon variogram carries long-range structure ($\tau^2 = 144$, $\sigma^2 = 167$, range 200 m) while regression residuals are nugget-dominated ($\tau^2 = 12$, $\sigma^2 = 88$, range 1.7 m). Regression has already absorbed the spatially structured signal; what remains is near white noise, so residual kriging can only fine-tune.

2026 S.T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

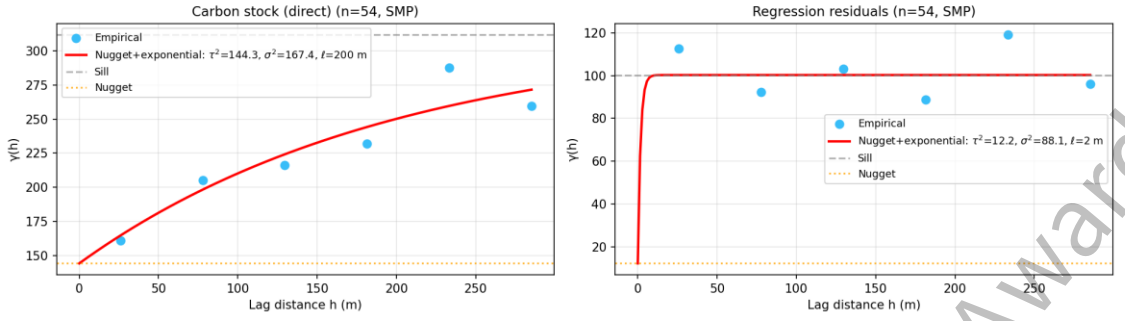


Figure 12 Variogram fits on an n = 54 SMP sample (left: carbon directly; right: regression residuals)

Figure 13 shows the same maps as 3D surfaces. The truth surface swings hard between peaks and valleys; the E4 estimate keeps the main terrain but shaves the extreme spikes; the error surface concentrates where the truth changes steeply.

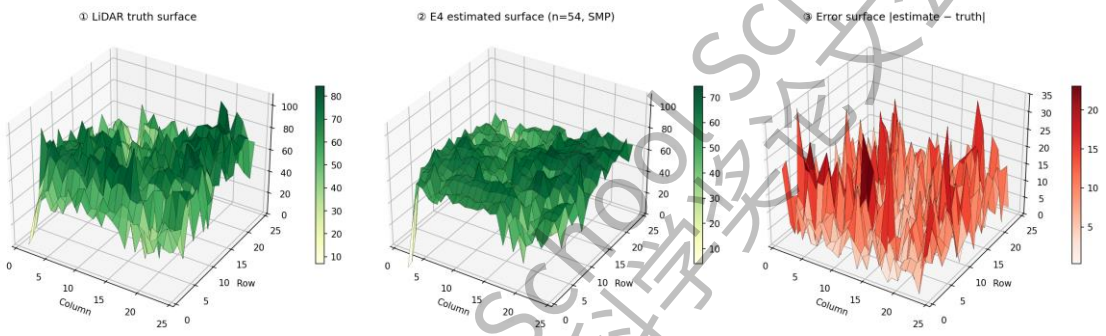


Figure 13 Truth, E4 estimate, and error as 3D surfaces (n = 54, SMP, seed 0)

The globally weighted E5/E6 compensate for weak samples. Under zone random at n = 27, E5's 14.78 sits 14% below E1's 17.25, with the LOO regression weight dropping to about 0.39: the cross-validation itself detects that regression is unreliable on this sample and hands weight to interpolation. Under SMP the same weight climbs smoothly from 0.39 to 0.63 as n grows (Figure 14), and the hybrids concede to pure regression. They hedge representativeness risk; they do not raise the best achievable accuracy.

E7 refines the hybrid family. Under SMP with n ≤ 54 it beats E5/E6 across the board (n = 27: 13.61 versus 13.83/13.95; n = 54: 12.07 versus 12.10/12.26) and takes the lowest RMSE overall for zone-systematic at n = 27. Its improvement is modest, and Figures 14 and 15 explain why. The carbon variogram has a long range (about 200 m), so the kriging variance grows slowly with distance and E7's per-pixel weights drift only from about 0.61 near plots to 0.65 far away. In Figure 15 the theoretical kriging error map (third panel) has a clean gradient, about 13.5 near plots and 15.3 in the corners, while the actual absolute-error map (fourth panel) looks random, with none of the predicted pattern. The residuals are nugget-dominated, so where errors actually land does not follow the theoretical map. The map captures the distance effect but not where the errors fall, which limits what E7 can gain.

2026 S-T Yau High School Science Award
仅用于2026丘成桐中学科学奖公开

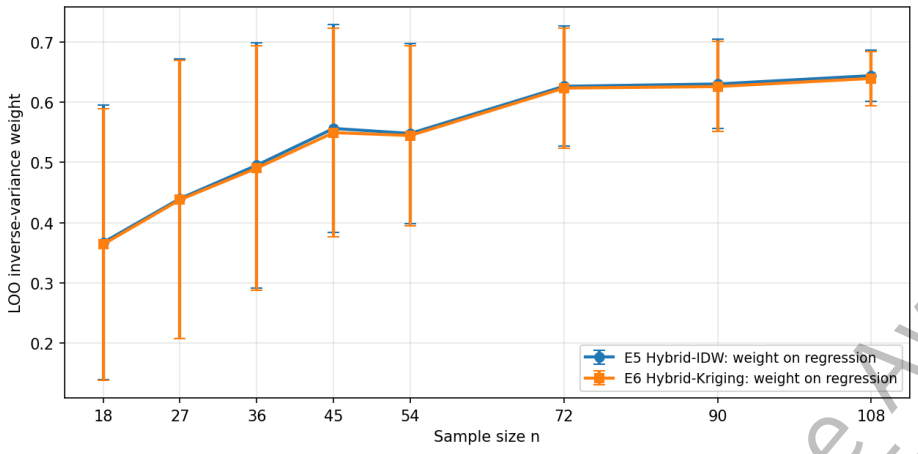


Figure 14 LOO inverse-variance regression weight of the E5/E6 hybrids versus sample size (SMP; error bars are ±1 SD across seeds)

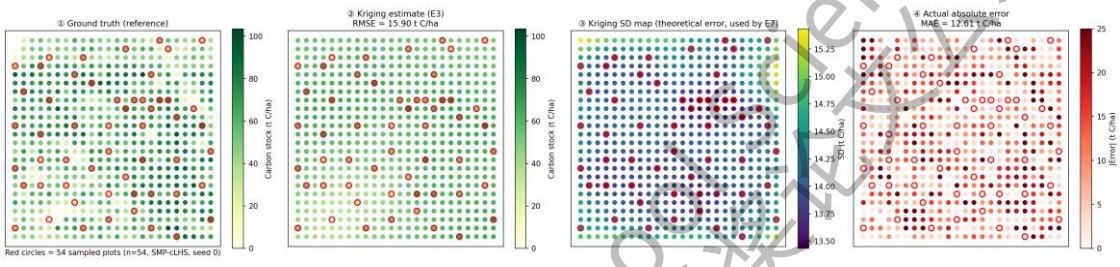


Figure 15 Kriging in four panels: truth, kriging estimate, theoretical SD, and actual absolute error (n = 54, SMP, seed 0; red circles are sampled plots)

Figure 16 compares the estimators on a radar chart of five metrics (RMSE, MAE, |Bias|, unexplained variance, error SD), each normalized independently, larger radius better. E4 rides the outer ring on four of five; E3 wins only error SD, as interpolation is naturally smooth; E2's |Bias| edge deserves caution because IDW contracts toward the sample mean by construction; E7 is outermost on |Bias|, consistent with its small-sample accuracy in the mean criterion of Section 4.5. The extra metrics leave the ranking unchanged.

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公开

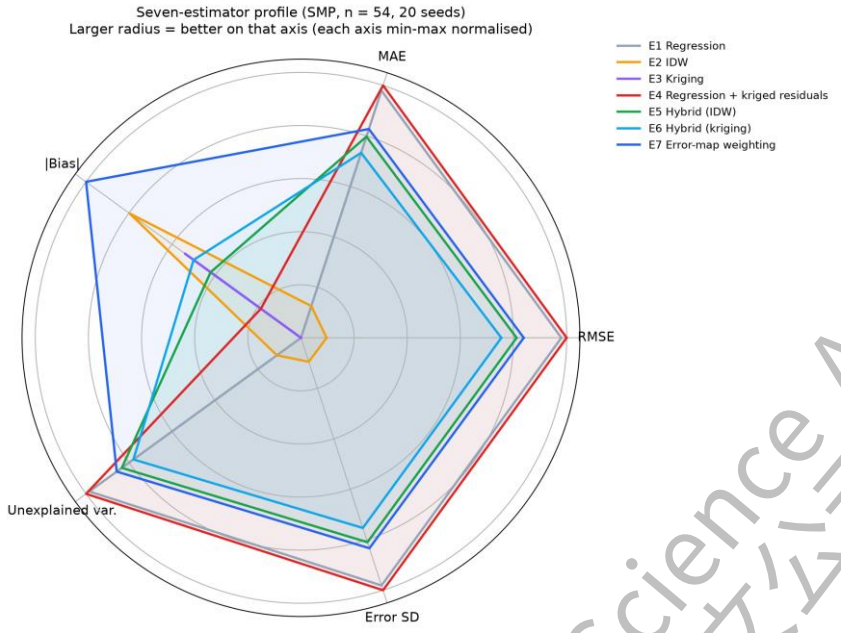


Figure 16 Five-metric radar chart of the seven estimators (each metric normalized independently, larger is better; n = 54, SMP)

4.4 Sample size versus error

Sweeping n from 18 to 108 (8 levels, 100 replicates each; Figure 17) uses the same draws as Table 2, so the values at n = 27, 54 and 90 coincide.

Across estimators, the variance of E1/E4 contracts quickly with n: ±8.8 at n = 18 down to ±0.3 at n = 108. Regression on small samples is high-variance and needs sample size or an optimized design to steady it; pure kriging (E3) has small variance but a high floor (above 15 throughout), the low-variance high-bias choice when no attribute information exists. Across designs (Table 4), contiguous sampling posts 26.2/20.6 at n = 18/27, worst of the four, 46-53% above SMP at the same n (17.9/13.5), converging only at n ≥ 90. Each zone contributes one small block of adjacent pixels, about 95% of the area lies far from any plot, the sample histogram misfits badly, and trees in one small patch grow alike. Contiguous sampling is the easiest to carry out in the field but the least representative; when dispersed placement is feasible there is little reason to choose it. The CCER mean-flat-map reference line (RMSE 17.8 to 18.2) stays almost level: more plots improve the mean but a constant map cannot express spatial heterogeneity, and E4+SMP is about 25% below it at n = 27.

2026 S-T Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

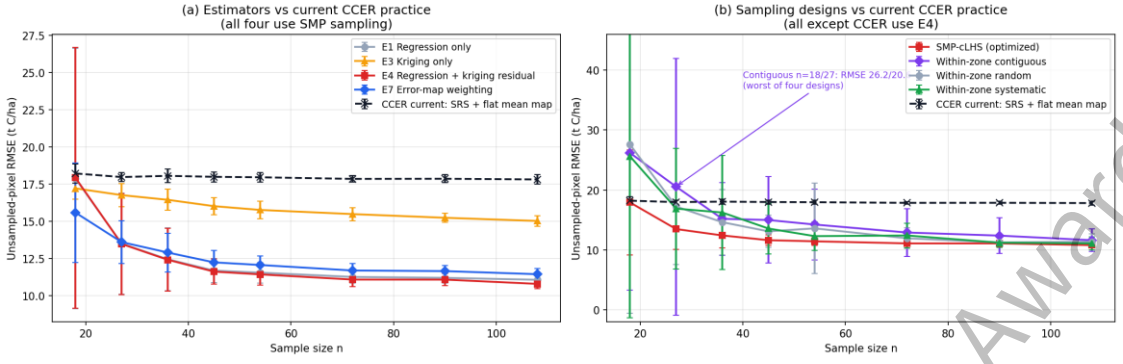


Figure 17 Mapping error versus sample size (left: estimators, all under SMP; right: designs, all with E4; 100 replicates per level; black dashed line is the CCER mean-flat-map benchmark)

Table 4 RMSE of the four sampling designs under E4 versus sample size (mean of 100 replicates, t C/ha)

n	SMP-cLHS	Zone contiguous	Zone random	Zone systematic
18	17.93	26.20	27.60	25.62
27	13.48	20.55	17.24	16.90
54	11.44	14.23	13.60	12.30
108	10.80	11.62	11.36	11.09

4.5 Comparison under the CCER mean criterion

The results so far use mapping RMSE, while the official CCER criterion scores only the mean. Table 5 switches to the relative error of the mean estimate, 100 replicates per scheme. The two SMP schemes stay ahead: E4+SMP and E7+SMP beat current CCER practice at every sample size, cutting the relative error by about half at n = 36. One detail flips relative to the mapping criterion: in the mean criterion E7 is the most accurate at small n (3.09 at n = 18 against E4's 4.15), because per-pixel weighting smooths the mean even where its pixel-wise RMSE is slightly worse.

The last column of Table 5 is the theoretical uncertainty u of the CCER Appendix F formula (half-width of the 90% interval, relative form):

$$u = t_{0.95}(n-1) \cdot \frac{S}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}} \cdot \frac{100}{\hat{\mu}} \tag{28}$$

where S is the sample standard deviation of plot carbon, N the number of pixels, and $\hat{\mu}$ the mean estimate. The theoretical u sits above the empirical mean absolute error, as their definitions differ: theory gives about 1.66 standard errors, the empirical mean absolute error about 0.8, and the predicted ratio 2.1 matches the measured (12.19/5.47 ≈ 2.2). For the stock task, theoretical u falls below 10% from n ≥ 27; the increment task on the same plots would need about n ≈ 3509, beyond the total of 576. The stock threshold is far gentler, which is why our improvement margins live in mapping rather than the total.

Table 5 Relative error of the mean estimate (%, 100 replicates) and CCER theoretical u

n	CCER current	E4+SMP	E7+SMP	E4+zone random	CCER theoretical u

2026 S.T. Yau High School Science Award 仅用于2026丘成桐中学科学奖论文公示

18	5.47	4.15	3.09	8.12	12.19
27	3.76	2.96	2.77	3.88	9.68
36	4.53	2.40	2.31	2.57	8.23
54	3.33	1.90	1.84	2.36	6.55
90	2.28	1.27	1.32	1.41	4.86
108	2.00	1.21	1.23	1.34	4.35

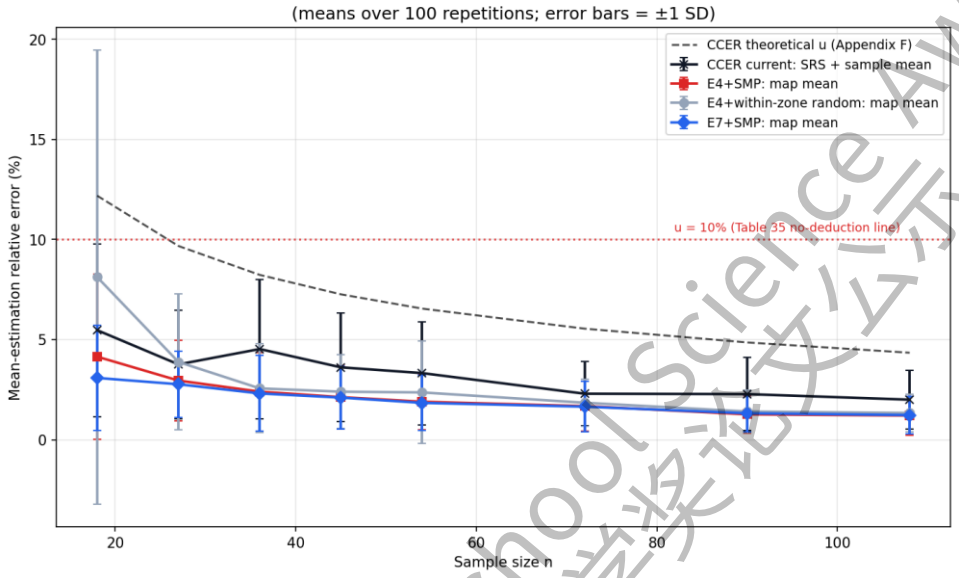


Figure 18 Relative error of the mean estimate versus sample size (100 replicates; black dashed: CCER Appendix F theoretical u; red dotted: the 10% no-deduction line)

4.6 Zone-level accuracy check

Global RMSE hides local differences. Table 6 gives per-zone mapping RMSE for each estimator (n = 27, mean of 100 draws). Under both designs the most heterogeneous zone Z1 (truth SD 19.3, containing the low-NDVI sparse patches) is taken by a hybrid estimator. Under random sampling E4 blows up to 31.5 in Z1 and E5 pulls it back to 22.2, a cut of about 30%; the lowest-carbon zone Z7 also leans hybrid. Under the weak design the hybrids lead in 5 of 9 zones; under SMP, E4 leads in 8 of 9. SMP removes most of E4's failure zones, which is why E4 needs no hedging once plots are placed well.

Table 6 Per-zone mapping RMSE (n = 27, mean of 100 draws; SMP and zone-random designs, t C/ha)

Zone	Truth mean/SD	SMP: E4	E5	E6	E7	Random: E4	E5	E6	E7
Z1	59.3 / 19.3	19.0	16.1	17.1	16.9	31.5	22.2	22.4	23.3
Z2	62.0 / 15.0	11.4	12.9	13.0	12.8	11.9	12.5	12.6	12.5
Z3	51.6 / 17.9	12.0	13.8	13.7	13.1	13.9	13.5	13.5	13.5
Z4	61.3 / 15.9	12.4	13.4	13.6	13.5	12.7	12.9	13.0	13.1
Z5	61.9 / 12.2	10.6	11.1	11.0	10.9	11.1	10.7	10.7	10.7
Z6	67.5 / 16.5	12.8	14.1	14.0	13.7	13.4	13.5	13.6	13.5
Z7	42.8 / 16.0	14.0	14.6	14.7	14.1	17.1	14.9	15.1	15.4
Z8	52.4 / 14.3	11.7	12.8	12.7	12.4	12.5	12.4	12.4	12.4
Z9	70.8 / 13.6	12.8	13.7	13.6	13.3	12.9	13.1	13.2	13.1

2026 Siyuan High School Science Award 仅用于2026年思源中学科学奖论文公示

4.7 Supplementary experiments

Sampling-attribute control. A natural simplification is histogram matching on NDVI alone. The control (100 draws, E4) shows a price: 13.48 for bivariate against 15.66 for univariate matching at n = 27, a gap of 2.18 (about 14%); the gap narrows at n = 54 and closes at n = 90 (Table 7). The univariate dif is itself near perfect (0.000 to 0.006), yet mapping degrades, which shows that representativeness in NDVI does not buy representativeness in NIRv. The regression uses both dimensions; covering one leaves blind spots in the other, and small samples pay for it in generalization. Sampling representativeness should align with the feature space the downstream model actually uses: with the two indices correlated at 0.82, the NIRv dimension is not one to drop under the current bivariate regression.

Table 7 Univariate versus bivariate sampling attributes (E4 mapping RMSE, t C/ha)

n	NDVI only	NDVI+NIRv bivariate
27	15.66	13.48
54	11.77	11.44
90	11.03	11.09

Two-stage protocol. Under SMP at n = 54, adding 200 calibration plots (the calibration set fits the prior and covariance parameters once, disjoint from the monitoring set) moves RMSE from 11.37 ± 0.56 to 10.64 ± 0.17, a 6.4% gain with smaller variance. Single-stage is already workable for mapping; an independent calibration set pays when the budget allows pursuing precision or stability.

Random-forest prior variant. Swapping E1's quadratic polynomial for a random forest [12] (same two 2024 indices, conservative hyperparameters, depth 3) raises RMSE from 13.50 to 13.87 at n = 27, from 11.55 to 12.63 at n = 54 and from 11.21 to 11.91 at n = 90 (100 draws). In the parallel increment analysis the forest won outright. For the stock, the NDVI-carbon relation is smooth and strong, and the polynomial extrapolates better than a conservative tree model. The relative performance of a prior is scene-dependent.

Reference-map error. All RMSE values are measured against the LiDAR reference map, which has not been validated cell by cell for carbon. We therefore perturbed the reference in four ways and repeated the SMP versus zone-random comparison with E4 (50 draws; plot locations are unchanged because the sampler sees only satellite indices). A uniform allometric bias (reference times 1.15) and a saturation-type shrinkage of extremes towards the mean (0.8 times the deviation) leave the relative gain unchanged, 20.0% at 27 plots and 6.1% at 54; E4 rescales with the reference, so this is expected. Independent cell-level noise with a 15% coefficient of variation lowers the gain to 16.0% at 27 plots and raises it to 9.4% at 54. A bias proportional to NDVI, the least favourable case because the design itself uses NDVI, gives 20.6% and 6.7%. The ranking of the

2026 Siyou Hi&D School Science Award
仅用于2026年思源高中科学奖论文公示

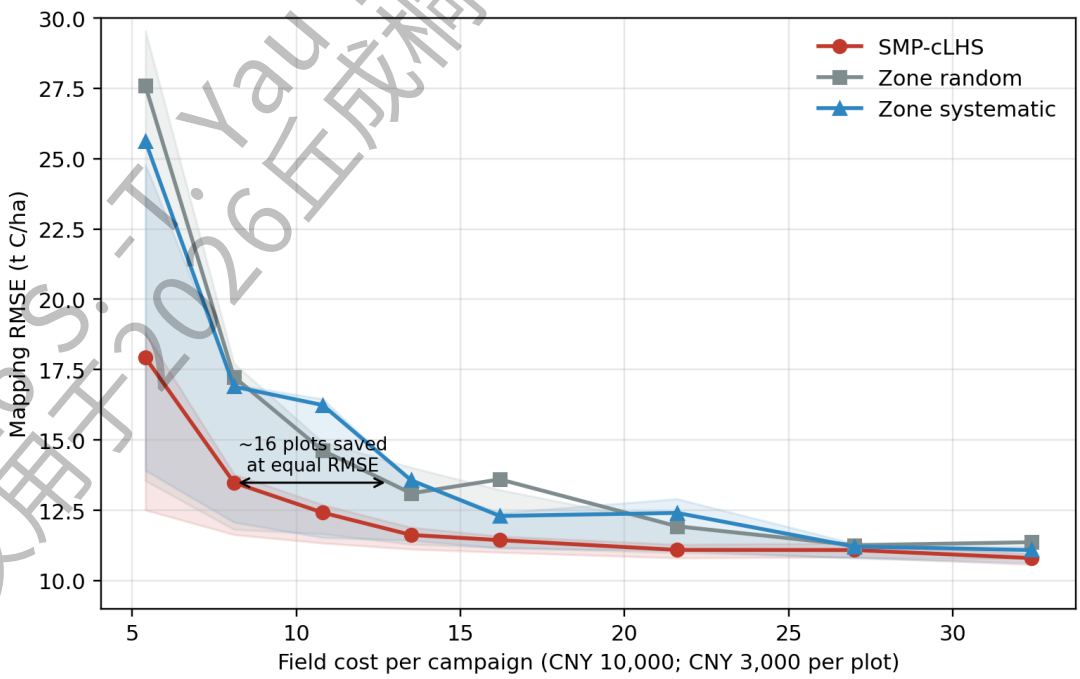
designs survives all four perturbations, although absolute RMSE moves with the reference. The test cannot exclude large reference errors that are unrelated to anything the satellite sees; only field-measured plot carbon could do that.

5 Economic analysis: from statistical gains to verification costs

This section evaluates the configurations of Section 4 against the economic framework of Section 3.7. To keep the evidence levels apart, every number carries one of three labels: measured (taken directly from the repeated experiments), calculated (an interpolation or a conversion of a measured result with a stated unit cost), or scenario (dependent on assumed prices, project sizes or adoption). Map results use estimator E4 and 100 repeated draws per configuration; project-mean results use 300 draws.

5.1 Cost-effectiveness frontier for the map target

Figure 19 plots mapping RMSE against field cost for each sampling design. At every cost level the optimized design lies below the random and systematic designs, so among the designs tested it is the cheapest way to reach a given map accuracy. The comparison covers only the placement rules tested. Table 8 converts the frontier into plots. Setting the target A at the RMSE that 27 optimized plots reach (13.48 t C/ha), zone-random sampling needs about 43 plots and zone-systematic sampling about 46 (measured curves, calculated by linear interpolation). Repeating the comparison at each sample size from 18 to 54 gives savings of 32-46% against random placement. The saving is a small-sample result: at 90 plots random placement comes within 0.2 t C/ha of the optimized design, the paired gap is not significant, and no saving is reported (Table 8).



2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学生数学奖公示

Figure 19 Cost-effectiveness frontier: mapping RMSE against field cost per campaign (E4, 100 draws; lines are means, bands the interquartile range; CNY 3,000 per plot). Equal plot counts imply equal budgets.

Table 8 Equal-accuracy plot requirement and paired comparison for the map target (E4, 100 draws). "—": random placement does not reach the target within 108 plots.

SMP plots	RMSE SMP	RMSE random	Random plots, equal RMSE	Plot saving	Paired gap [95% CI]	SMP better
18	17.93	27.60	26	32%	9.67 [3.9, 15.5]	59%
27	13.48	17.24	43	37%	3.76 [1.9, 5.6]	63%
36	12.42	14.63	67	46%	2.21 [1.1, 3.3]	65%
54	11.44	13.60	85	37%	2.16 [0.7, 3.7]	64%
90	11.09	11.26	—	—	0.17 [-0.0, 0.4]	44%

5.2 Diminishing returns and the shadow price of map accuracy

Diminishing returns appear along the optimized design's own curve. Each extra plot lowers RMSE by 0.49 t C/ha between 18 and 27 plots, 0.12 between 27 and 36, 0.054 between 36 and 54, and 0.012 between 54 and 108 (measured). The value of optimization relative to random placement also falls with sample size, which is a separate fact: the paired RMSE gap is 9.67 t C/ha at 18 plots, 3.76 at 27, 2.16 at 54 and 0.17 at 90, where its 95% confidence interval includes zero (Table 8).

Table 9 converts the same curve into the discrete shadow price of Eq. (26), an incremental cost-effectiveness ratio: extra field cost per unit of accuracy gained. At CNY 3,000 per plot, moving from 27 to 54 optimized plots lowers RMSE by 2.04 t C/ha and pays off only if a map user values accuracy at more than CNY 39,600 per t C/ha of RMSE. Over the 23.04 ha block that is CNY 1,719 per tonne of mis-mapped carbon, about 6.6 times the carbon price p_C = CNY 259 per t C. Moving from 54 to 108 plots requires CNY 11,036 per tonne, about 43 times the carbon price. We do not observe what any user would pay, so these are break-even values, not estimates of willingness to pay (calculated).

Continuous frontier. The discrete differences use few points, so we also fit a smooth frontier to the eight measured points of the optimized design (n = 18 to 108):

$$\mathrm{RMSE}(n) = a + bn^{-\gamma} \tag{29}$$

with a = 10.82 t C/ha, b = 5,441 and γ = 2.30 (largest residual 0.16 t C/ha). The constant a is the error floor that more plots cannot remove; with these estimators it is set by the satellite prior. The frontier is convex, so the continuous shadow price $\eta(n) = c/(b\gamma n^{-\gamma-1})$ rises steeply: CNY 548, 1,415 and 5,388 per tonne of mis-mapped carbon at 27, 36 and 54 plots. The discrete value of Table 9 between 27 and 36 plots (CNY 1,100 per tonne) lies between the two continuous values, as it should. Setting the marginal plot

2026 S.T. Yau High School Science Award 仅用于2026丘成桐中学科学奖论文公示

cost equal to the marginal value of accuracy, $c = \lambda_C A_m b\gamma n^{-\gamma-1}$, gives the map-only optimum in closed form:

$$n^*_{\text{map}} = \left(\frac{\kappa p_C A_m b\gamma}{c}\right)^{1/(\gamma+1)} \tag{30}$$

For κ = 1, 2 and 5 this gives about 22, 27 and 35 optimized plots. The same fit for zone-random placement gives 28, 34 and 45 plots, so the optimized design reaches its optimum with 6 to 10 fewer plots and a lower RMSE. The fit for random placement is looser (largest residual 1.0 t C/ha), and both optima rest on an unobserved κ.

Table 9 Discrete shadow price of map accuracy for SMP, Eq. (26), at CNY 3,000 per plot: the value of a 1 t C/ha RMSE reduction above which adding plots pays off, and the same value per tonne of mis-mapped carbon over A_m = 23.04 ha, as a multiple of p_C = CNY 259 per t C.

Plots	**RMSE reduction (t C/ha)**	**Extra field cost (CNY)**	**Break-even (CNY per t C/ha)**	Break-even (CNY per t C; × p_C)
18 → 27	4.44	27,000	6,075	264 (× 1.0)
27 → 36	1.07	27,000	25,348	1,100 (× 4.2)
27 → 54	2.04	81,000	39,614	1,719 (× 6.6)
54 → 108	0.64	162,000	254,280	11,036 (× 42.5)

5.3 The project-mean target and threshold risk

A developer is exposed to one campaign, not to an average over campaigns. Table 10 reports the share of 300 draws in which the error of the project mean stays within 5%, the target of Eq. (25). The optimized design reaches an 80% chance with about 25 plots and a 95% chance with about 44. Zone-random sampling needs about 32 plots for 80%, and does not reach 95% within the 45 plots tested, so the comparison can only be made at 80%: a saving of 23%, with a 95% bootstrap interval of 11-32% (calculated from the measured shares). Whole-area random sampling with the sample mean, the current practice, reaches only 70% at 45 plots, so against current practice the saving is at least 44% (calculated, lower bound).

Two cautions apply. The shares come from 300 draws, so neighbouring sample sizes can differ by a few points by chance, and the plots needed are interpolated between the tested sizes of 18, 27, 36 and 45. The saving for the mean is also smaller than for the map: satellite information and plot placement mainly sharpen the spatial pattern, while the project total is already estimated fairly well by a sample mean once a few dozen plots are measured.

2026 S.T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

Table 10 Share of 300 draws with project-mean error within 5%, and plots needed for 80% and 95% assurance (linear interpolation). "—": not reached within 45 plots.

Design + estimator	n = 18	n = 27	n = 36	n = 45	Plots for 80%	Plots for 95%
SMP + E4 (this paper)	71%	83%	90%	95%	25	44
Zone random + E4	49%	71%	86%	85%	32	—
Simple random + sample mean (current CCER practice)	55%	65%	67%	70%	—	—

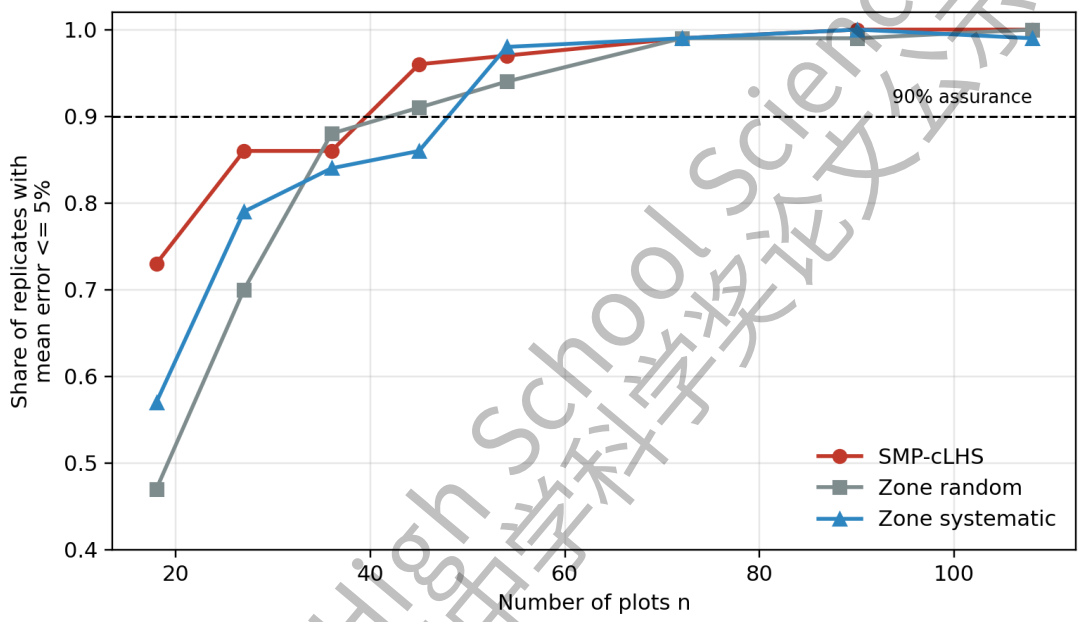


Figure 20 Share of draws in which the project-mean error stays within 5% (E4, 100 draws per point). The dashed line marks 90% assurance; Table 10 gives the 300-draw values.

5.4 Expected total loss: deduction risk and the optimal campaign

Eq. (23) can now be solved with all three terms. For every draw we compute the uncertainty u that the Appendix F formula reports from the plot values, and read the deduction share from Table 35 of the CCER methodology [3]. Table 11 gives the share of draws above the 10% line (measured). At 18 plots most draws of every design are above it. At 27 plots the optimized design is above it in 60% of draws, against 49% for zone-random and 41% for current practice; at 36 plots the order reverses, 5% against 10% and 14%.

The order at 27 plots and its reversal at 36 have one cause. The Appendix F formula scores the spread of the plot values. A representative design includes low- and high-carbon cells, so its average reported uncertainty is slightly higher: 10.3% at 27 plots against 9.6-9.9% for the other designs (Table 14). It also varies much less from one campaign to the next, because histogram matching fixes the mix of cells in the sample.

2026 S. T. Yau High School Science Award 仅用于2026丘成桐中学科学奖论文公示

Its standard deviation across draws is 1.3 percentage points at 27 plots against 1.7-2.1 for the other designs, and 0.9 against 1.3-1.5 at 36 plots. When the average sits on the 10% line, as at 27 plots, every design crosses it in 41-60% of draws. Once the average is below the line, the most stable design crosses it least often: at 36 plots in 5% of draws, against 10% for zone-random and 14% for current practice. The developer's optimal campaign lies in this range, so near its own optimum the rule rewards the optimized design. That design is also the most accurate: at 27 plots the project mean is within 5% of the reference in 83% of draws, against 71% for zone-random and 65% for current practice (Table 10). The next result states the mechanism.

Proposition 4 (Stable reported uncertainty lowers deduction risk). Suppose the reported uncertainty u of a design with n plots is approximately normal across campaigns, with mean m and standard deviation s. Then the probability of crossing the 10% line is $\Pr(u > 10) = \Phi((m - 10)/s)$. If m < 10 this probability rises with s, and if m > 10 it falls with s. **Proof.** With $z = (m - 10)/s$, $\partial \Pr(u > 10)/\partial s = \phi(z)(10 - m)/s^2$, which has the sign of 10 − m. This completes the proof. At 27 and 36 plots the normal approximation reproduces the shares in Table 11 to within 2 percentage points. The optimized design lowers the developer's deduction risk because its reported uncertainty is predictable; its average level is no lower. A lower variance of the outcome is worth more to a risk-averse developer (Eq. (31)).

Table 12 adds the money. Credits are a flow, the verified net removal in a monitoring period against a baseline, while our map measures a stock. At the block's mean density the farm's 74,750 ha of forest hold about 4.40 million t C, or 16.1 million t CO2e (1 t C = 3.67 t CO2e). We take an illustrative issuance of Q = 500,000 t CO2e, about 3% of that stock, at the 2025 CCER average price of CNY 70.76 [23]. Q is a scenario, since a single period of data cannot give the baseline, and the scale analysis below shows how the result depends on it. At 27 plots expected deductions are CNY 1,274, 1,040 and 870 thousand for the three designs, against a field cost of CNY 81 thousand. Deductions dominate, so the loss-minimizing campaign is the one that clears the 10% line, about 45 plots, for every design (calculated; the optimum is located on the tested grid of 18, 27, 36, 45 and 54 plots).

The designs then differ through the third term. With κ = 0, that is if nobody values the map, the optimized design and current practice reach the same minimum loss of CNY 135 thousand at CNY 3,000 per plot, and zone-random placement costs CNY 156 thousand. Once a mis-mapped tonne has any value, the optimized design has the lowest expected total loss in every cost and κ scenario of Table 12. At CNY 3,000 per plot it saves CNY 38 thousand per campaign against current practice when a mis-mapped tonne is worth its market price (κ = 1), CNY 76 thousand at κ = 2 and CNY 190 thousand at κ

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公示

= 5 (calculated; κ is a scenario). On the two observable terms alone the optimized design is not cheaper than current practice. Its value is elsewhere: it reaches a given accuracy with fewer plots (Section 5.1), and it leaves less over-issue for buyers to bear (Section 5.8). κ enters only as a sensitivity.

Project scale. The deduction term grows with Q while field cost does not, so verification has economies of scale. The direction of the effect follows from a general result.

Proposition 5 (Campaign size rises with project scale). If the expected deduction E[D(u)] is non-increasing in the number of plots n, the loss-minimizing campaign n*(Q) of Eq. (23) is non-decreasing in Q. **Proof.** For $n' > n$ the change in loss, $c(n' - n) + pQ(\mathbb{E}[D_{n'}] - \mathbb{E}[D_n])$, is non-increasing in Q, so the objective has decreasing differences in (n, Q). By the monotone comparative statics theorem of Milgrom and Shannon [31] its minimizer is non-decreasing in Q. This completes the proof. The condition holds for all three designs at every tested n (Table 11).

Solving Eq. (23) with κ = 0 for Q from 1,000 to 1,000,000 t CO2e gives three regimes. Below about 16,000 t CO2e field cost dominates and about 18 plots minimize the loss for every design. Between that level and a design-specific threshold Q*, a middle campaign of 27 to 36 plots is optimal. Above Q* the 10% line governs and the campaign rises to 45 plots: Q* is about 50,000 t CO2e for current practice, 71,000 for zone-random and 141,000 for the optimized design. The optimized design has the highest threshold because at 36 plots its draws stay below the 10% line 95% of the time (Table 11, Proposition 4), so for mid-sized projects it can stop at 36 plots where the other designs need 45. All thresholds are computed at p = CNY 70.76 per t CO2e and CNY 3,000 per plot, and scale with the ratio of plot cost to carbon price (scenario).

Project area and the 23 ha block. Q refers to a whole project, while the plots come from one 23 ha block, so the scaling needs a word. The number of plots needed to estimate a mean depends on the variability of the stratum, not on its area: for a large stratum the finite-population factor in Eq. (28) tends to one and drops out. Treating the block as a sample of one large, homogeneous stratum (that factor set to one) leaves the loss-minimizing campaign at 45 plots for every design. What moves it is heterogeneity. With the block's coefficient of variation of 30% the answer is 45 plots; at 20% it falls to 27 and at 40% it rises to 72 (calculated). A heterogeneous project would therefore be split into strata, each with its own campaign, which is what the CCER methodology already allows. The map term of Eq. (23) is different: it scales with the mapped area A_m, because map accuracy depends on plot density.

Risk aversion. A deduction is a lumpy loss: in most draws nothing is lost, in some 6% of the issuance is. A risk-averse developer weighs that lottery above its mean. With

2026 Siyuan High School Science Award 仅用于2026年思源高级中学科学奖公开赛

constant absolute risk aversion ρ the deduction term in Eq. (23) is replaced by its certainty equivalent, computed from the draws of each design:

$$\mathrm{CE}_{\rho}(S) = \frac{1}{\rho}\ln\mathbb{E}[e^{\rho pQD(u(S))}] \tag{31}$$

We hold ρ fixed across project sizes and set the risk tolerance 1/ρ at CNY 2.12 million, the value of one 6% deduction on the reference issuance of 500,000 t CO2e. Q* then falls from 50,000 to about 45,000 t CO2e for current practice and from 141,000 to about 126,000 for the optimized design, and stays at 71,000 for zone-random (calculated). A risk-averse developer moves to the 45-plot campaign at smaller project sizes. At no tested scale does risk aversion lower the campaign below the risk-neutral choice. That is a result of these draws, not a general law: the certainty equivalent is at least the expected loss, but the optimal n depends on how it changes with n [30].

Repeated campaigns. Plots in a CCER project are permanent and are re-measured in each monitoring period. The placement is computed once, but the saving recurs. With five campaigns at five-year intervals over a 25-year crediting period, the present value of the per-campaign saving is 3.8 times that saving at a 3% discount rate and 3.0 times at 6%. The block saving of about 16 plots (CNY 47 thousand at CNY 3,000 per plot, Section 5.6) is then worth about CNY 143-180 thousand in present value. This assumes the same plots stay representative across periods, which repeated data would have to confirm.

Table 11 Share of 100 draws in which the Appendix F uncertainty exceeds the 10% deduction line (computed from plot values; independent of the estimator).

Design	n = 18	n = 27	n = 36	n = 45	n = 54
SMP (this paper)	90%	60%	5%	0%	0%
Zone random	86%	49%	10%	1%	0%
Current CCER practice	83%	41%	14%	0%	0%

Table 12 Loss-minimizing number of plots and minimum expected total loss (CNY thousand) per campaign, Eq. (23). p = CNY 70.76 per t CO2e, Q = 500,000 t CO2e (scenario); λ_C = κ·p_C with p_C = CNY 259 per t C and A_m = 23.04 ha (κ is a scenario). Optimum located on the grid n = 18, 27, 36, 45, 54.

Plot cost	κ	SMP: n* / loss	Zone random: n* / loss	Current practice: n* / loss
CNY 1,500	0	45 / 68	54 / 81	45 / 68
CNY 1,500	1	45 / 137	54 / 162	45 / 175
CNY 1,500	2	45 / 206	54 / 244	45 / 283
CNY 1,500	5	45 / 415	45 / 480	45 / 605
CNY 3,000	0	45 / 135	45 / 156	45 / 135
CNY 3,000	1	45 / 204	45 / 235	45 / 243
CNY 3,000	2	45 / 274	45 / 313	45 / 350
CNY 3,000	5	45 / 482	45 / 548	45 / 673
CNY 4,500	0	45 / 202	45 / 224	45 / 202
CNY 4,500	1	45 / 272	45 / 302	45 / 310
CNY 4,500	2	45 / 341	45 / 380	45 / 418
CNY 4,500	5	45 / 550	45 / 615	45 / 740

2026 S. T. Yau High School Science Award 仅用于2026年成师中学征文公示

5.5 Field logistics: does dispersion raise the cost per plot?

Problem (24) assumes equal plot costs, but dispersed plots may need more walking between them. We measured the route each design requires, entering the block at one corner and visiting all plots in a near-shortest order. At 27 plots the route is 2,513 m for optimized placement, 2,116 m for random and 1,543 m for contiguous placement. Optimized placement adds about 19% to the walking distance, roughly 0.2 hours for a crew at 2 km/h in forest terrain, far less than the time needed to measure the 16 extra plots that random placement requires. Contiguous placement is the cheapest to walk but has the worst accuracy (20.55 t C/ha at 27 plots against 13.48 for optimized placement), so its travel saving does not offset the extra plots it needs (measured). The routes are straight-line distances on the study block and ignore terrain, roads and access points.

5.6 Cost and carbon-price scenarios

Table 13 lists every economic parameter with its source. We have no market quotation for a 20 m verification plot, so we build the cost from the forestry survey fee guideline [25], which prices field work at CNY 1,500 per worker-day (Article 11). A three-person crew therefore costs CNY 4,500 a day. At three plots a day, which suits accessible plantation blocks, a plot costs CNY 1,500; at one to two plots a day, typical where crews walk between scattered or remote plots, it costs CNY 2,250-4,500. We use CNY 1,500, 3,000 and 4,500 per plot as low, central and high scenarios.

Study block. Matching the map accuracy of 27 optimized plots with random placement needs about 43 plots, so one campaign saves about 16 plots, or CNY 24, 47 and 71 thousand in the three cost scenarios (calculated). The percentage saving does not depend on the unit cost; only the money amount does.

Host farm. The Saihanba farm holds 74,750.30 ha of forest. A wall-to-wall map at the accuracy of 27 optimized plots, with the experimental plot density projected to the whole farm, would need field work worth about CNY 131-263 million with optimized placement against CNY 208-416 million with random placement, at CNY 1,500-3,000 per plot. This applies only to a user who wants the full map at that accuracy: map accuracy depends on local plot density, whereas the plots needed for the project mean do not grow with area (Section 5.3). The farm also contains stands that differ from the tested block in age, species and terrain. The figure is a mechanical area-scaling illustration of the method, not a project budget or an estimate (scenario).

National inventory. China's continuous forest inventory re-measures about 415,000 fixed plots in each five-year cycle [26]. If the measured map saving carried over, the cycle would need 132-191 thousand fewer plots, about CNY 79-115 million a year at CNY 3,000 per plot. This extrapolates one 23 ha plantation block to forest types, terrain and

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学生数学奖论文公示

designs we have not tested; the inventory has other aims than carbon verification and its plots are already systematic. The number shows the order of magnitude at stake and should not be read as an estimate (scenario).

Carbon prices and deductions. Two dated prices anchor the scenarios: the 2025 average CCER trading price of CNY 70.76 per t CO2e [23] and the end-2024 closing price of national ETS allowances, CNY 97.49 per t [24]. For an illustrative issuance of 500,000 t CO2e, a 6% deduction is worth CNY 2.1-2.9 million. Whether optimized sampling reduces this exposure depends on how verifiers compute uncertainty: the Appendix F value does not respond to the estimator, while stratifying by satellite indices under Appendices E and F could lower it. That would require confirmation by the verification body, so the deduction channel stays a scenario and is not added to the savings (scenario).

Table 13 Economic parameters, values and sources.

Parameter	Value	Source / status
Plot cost	CNY 1,500 / 3,000 / 4,500	Fee guideline [25]: CNY 1,500 per worker-day, 3-person crew, 1-3 plots a day
Block saving per campaign	CNY 24k / CNY 47k / CNY 71k	Calculated: 16 plots x cost
CCER price	CNY 70.76 / t CO2e	2025 average trading price [23]
ETS allowance price	CNY 97.49 / t	End-2024 closing price [24]
Illustrative issuance	500,000 t CO2e	Assumption (scenario)
Carbon to CO2e	1 t C = 3.67 t CO2e	Molecular-weight ratio 44/12
Value of mis-mapped carbon λ_C	$\kappa \cdot p_C$, κ = 0 / 1 / 2 / 5; p_C = CNY 259 per t C	Scenario anchored to [23]
Mapped area A_m	23.04 ha	Study block (measured)
Farm carbon stock	4.40 million t C = 16.1 million t CO2e	Calculated: 74,750 ha × 58.85 t C/ha (scenario)
Risk tolerance 1/ρ	CNY 2.12 million (one 6% deduction on 500,000 t CO2e), fixed	Assumption (scenario)
Discount rate	3% / 6% a year, five campaigns over 25 years	Assumption (scenario)
National inventory	415,000 plots / 5 years	NFGA [26]; national saving CNY 79-115 million a year (scenario)

5.7 From carbon stock to creditable removals

Our data measure carbon stock in one year. Credits are issued for the change in stock against a baseline, after leakage and additionality are addressed. The experiment shows that optimized placement lowers the cost of measuring stock; it does not show that credit revenue rises in proportion. The paper's own figures show why the step to stock change is hard: under the Appendix F formula the increment task on the same plots would need about 3,509 plots, more than the 576 available (Section 4.5), and the 2018 table in the dataset was back-derived from 2024 (Section 3.5), so it cannot serve as an independent second period. Extending the analysis requires two independently measured periods, a baseline, and propagation of uncertainty through the difference. That uncertainty

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公示

combines the errors of two inventories and is usually larger than for one stock, so the deduction risks of Section 5.4, computed from single-period stock plots, are a lower bound on the exposure of a real issuance. The Saihanba block is used as a test site for the method; whether a given stand would qualify as a new CCER afforestation project is a separate question that we do not address.

5.8 Who pays, incentives, and alternatives

Who pays and who benefits. The developer pays for the field campaign. Part of the benefit of a more accurate estimate goes to buyers and to the market as a whole, because buyers cannot check carbon claims themselves, a standard case of quality uncertainty [21]. Studies of forest offset programmes elsewhere have found systematic over-crediting [22]. A developer therefore has a weaker private reason to pay for accuracy than the market as a whole would have.

Incentives and manipulation. A developer gains from a high estimate and a verifier from a cautious one. The optimized placement is computed from satellite indices only, before any field work, so given the imagery, the code and the random seed, a verifier can regenerate the plot set and confirm that plots were not steered towards high-carbon stands. The seed can be drawn by the verifier. A caution remains: SMP is an optimized rather than a probability design, so the unbiasedness of the mean rests on the model, not on random selection. A simple remedy is to keep the optimized strata but draw plots at random within them.

The uncertainty rule as an incentive device. The deduction rule sets up a principal-agent problem. The verifier, acting for buyers and the market, cares about the true error of the project mean. The developer alone chooses where the plots go, and it cares about field cost and about the reported uncertainty u that triggers deductions. The verifier sees u but not the true error; checking the true error would need an independent field inventory, a problem of costly state verification [28]. Given the rule, the developer picks the design and n that minimize its own loss in Eq. (23); the verifier's choice is the rule. Table 14 shows what the current rule does to that choice.

Table 14 Reported uncertainty and true accuracy at n = 27 under the current Appendix F rule and a residual-based (GREG) rule, Eq. (32). Accuracy is for the sample mean (Appendix F) and the GREG estimate; Table 10 uses E4 and 300 draws, so its shares differ (100 draws per design; "within 5%": share of draws whose project-mean error is at most 5% of the reference mean).

Design	Appendix F: mean u (SD) / share > 10%	Within 5% (sample mean)	GREG: mean u / share > 10%	Within 5% (GREG)
SMP (this paper)	10.3% (1.3) / 60%	89%	6.1% / 0%	84%
Zone random	9.9% (1.7) / 49%	68%	5.8% / 0%	71%
Zone contiguous	9.9% (2.1) / 50%	57%	6.0% / 1%	71%
Simple random (current CCER practice)	9.6% (1.7) / 41%	72%	6.0% / 1%	77%

2026 S. T. Yau High School Science Award 仅用于2026丘成桐中学科学奖论文公示

Under the current rule the optimized design is the more accurate choice, and for projects of roughly 18,000 to 128,000 t CO2e also the cheaper one (κ = 0, CNY 3,000 per plot); above that range all designs need 45 plots and it ties with current practice. At 27 plots it is within 5% of the reference mean in 89% of draws, against 57-72% for the other designs. At 36 plots its stable reported uncertainty keeps it below the 10% line in 95% of draws (Proposition 4), which is why its scale threshold Q* is the highest of the three designs. A developer that minimizes its own expected loss therefore has a private reason to adopt it. The private and the public interest point the same way here, but only by accident: the rule scores the stability of the spread of plot values, not the error, so it would reward any design that keeps that spread stable, accurate or not. The verifier can check where the plots are, since the placement can be regenerated from the seed, but not how far the estimate is from the truth. This is an information problem of the kind behind the over-crediting found in other offset programmes [22].

By the informativeness principle [29], a rule should condition on any statistic that carries information about the true error beyond u. The residual standard deviation of a satellite-assisted estimator is such a statistic, and a rule built on it would reward the optimized design for its accuracy rather than for its stability.

A model-assisted rule rewards accuracy directly. The generalized regression (GREG) estimator uses the wall-to-wall satellite prediction as auxiliary information and the plots only to estimate the residual, so its uncertainty is computed from the residual standard deviation s_e (p = 6 model terms):

$$u_G = \frac{t_{0.95,n-1}}{\hat{\mu}_G}\frac{s_e}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}, s_e^2 = \frac{1}{n-p}\sum_{j\in S}(Y_j - \hat{R}_j)^2 \tag{32}$$

Because the satellite indices explain about 60% of the variance, u_G is about 40% below the Appendix F value at every sample size. At 27 plots no design exceeds the 10% line in more than 1% of draws, and the optimized design is still the most accurate (84% within 5%). Re-solving Eq. (23) at CNY 3,000 per plot with κ = 0, the cost-minimizing campaign falls from 45 to 27 plots, and the field bill of the optimized design from CNY 135 thousand to CNY 81 thousand (calculated). Neither campaign incurs a deduction, so the whole saving is field cost. Two conditions apply. GREG is design-unbiased only under probability sampling, so the optimized strata should be kept and plots drawn at random within them, with the seed set by the verifier. The regression model must be fixed before field work. Both would need the agreement of the methodology's custodians.

Private and social optimum. The developer bears the field cost and the deductions, but not the cost of credits issued without carbon behind them; over-issue raises its revenue. A planner would count that cost. Valuing an over-issued credit at the carbon price, the

2026 S. T. Yau High School Science Award 仅用于2026丘成桐中学科学奖公开赛

social loss of a campaign is

$$L_S(S) = \sum_{j \in S} c_j + pQ\mathbb{E}[\frac{(\hat{\mu}-\mu)^+}{\mu}] \tag{33}$$

where $(\cdot)^+$ keeps only overestimates. The asymmetry is deliberate. An over-issued credit is sold and used to offset an emission elsewhere, so the atmosphere bears its cost; an underestimate means the developer sells fewer credits than its forest earned, a loss it bears itself and already weighs. At the developer's own optimum under the current rule (45 plots, CNY 135 thousand of field cost), expected over-issue is worth CNY 691 thousand for current practice (standard error 92), 533 thousand for zone-random and 361 thousand for the optimized design (standard error 52). The external cost is 2.7 to 5.1 times the field bill, which is the economic content of the quality-uncertainty argument above [21]. Against current practice, optimized placement roughly halves it at no extra field cost; the difference from current practice is about three standard errors. Under the GREG rule the private campaign falls to 27 plots and expected over-issue is 408 thousand for the optimized design, 499 thousand for whole-area random and 799 thousand for zone-random placement. The rule change lowers the developer's cost but does not by itself lower the external cost; it needs representative placement alongside it. With 100 draws per design the standard errors are about 15% of the means, too wide to locate the planner's optimal n, so we report the comparison at the private optimum only (calculated; Q and p are scenarios).

Why the market does not solve this alone. The methodology already permits stratification, and near the optimal campaign the current formula already favours the optimized design, yet the over-issue that buyers bear is not priced anywhere in the developer's loss. The private return to better placement covers deduction risk and field cost, not the external cost. This makes the problem one of verification rules as much as of technique, and Table 14 measures how large the rule effect is.

Alternatives and costs not counted. Wall-to-wall airborne LiDAR could replace much of the field work, but we have no price quotation and LiDAR itself needs field calibration, so we do not compare the two in money. The optimized design also uses satellite imagery and computation that random placement does not need. Free imagery (for example 10 m Sentinel-2) can stand in for the commercial PlanetScope data used here, and the placement step runs in minutes on a laptop, but these costs are not included in the savings above.

5.9 Implications for project developers and verification rules

For developers, the results give a simple rule: when the field budget is small, spend effort on where plots go before adding plots, because that is where the return per plot is highest. Once several dozen plots are affordable, random and optimized placement

2026 S.T. Yau High School Science Award 仅用于2026丘成桐中学科学奖论文公示

converge. For a large project the field bill is a small share of credit revenue, so the saving matters most for small projects, where verification is a larger share of costs.

For verification rules, the results suggest that allowing satellite-index strata within the stratified formulas of Appendices E and F would let the reported uncertainty reflect information the project already holds. This is a question for the methodology's custodians; the paper offers evidence on its potential value, not a proposal to change the 10% line. In our data the largest gain comes from that combination: a residual-based uncertainty and plots drawn at random within satellite strata (Section 5.8). The two parts are complements. The residual-based rule lowers the developer's field cost, and representative placement lowers the over-issue that buyers bear; adopting the rule without the placement requirement lowers only the first.

6 Discussion

6.1 Why optimized sampling improves small-sample mapping

The relative importance of sampling design and estimator choice varied with sample size. At $n = 27$, the difference between SMP-cLHS and random sampling for a given estimator was larger than the differences among several estimators under the optimized design, indicating that sample placement was the dominant source of improvement under the most restrictive sampling budget. As sample size increased, however, the performance gap between optimized and random designs narrowed substantially and was small by $n = 90$.

This pattern can be explained by coverage of the auxiliary-variable space. At small sample sizes, random sampling may underrepresent parts of the NDVI-NIRv distribution, leading to a mismatch between the training sample and the full population. The optimized design explicitly constrains this mismatch while maintaining spatial dispersion, thereby improving the representativeness of the calibration sample and reducing prediction error. With larger samples, random designs increasingly reproduce the population distribution without explicit optimization, and the additional benefit of optimized placement therefore diminishes. The results indicate that sampling optimization is most valuable when field observations are scarce rather than when sample coverage is already dense.

6.2 Estimator performance under different sampling designs

Estimator performance also depended on the quality of the sampling design. Regression kriging (E4) showed the most consistent performance, producing the lowest RMSE in five of the nine sampling-design × sample-size combinations, including all three under SMP. Its advantage arises from combining the broad-scale relationship captured by remote-sensing covariates with spatial interpolation of the remaining residual structure.

2026 S.T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公示

The inverse-variance fusion estimators were comparatively more robust when sampling designs were less representative, although their advantage diminished as sampling quality improved. This suggests that model fusion can provide some protection against deficiencies in sample placement, but it does not consistently outperform a well-specified regression-kriging approach when the sampled plots adequately represent both the attribute and spatial domains. Direct kriging was less stable at small sample sizes, most likely because reliable variogram estimation becomes difficult when only a limited number of observations are available.

These results suggest a practical hierarchy rather than a single universally optimal estimator. When optimized sampling is feasible and sufficient auxiliary information is available, regression kriging provides the most reliable default choice in this experiment. Under less favorable sampling configurations, fusion approaches may improve robustness, whereas purely spatial estimators should be used cautiously when the sample size is too small to estimate spatial covariance reliably.

6.3 Implications for repeated carbon monitoring

The present analysis focuses on single-period carbon stock rather than temporal carbon-stock change. The two tasks may differ in the relative importance of spectral covariates and spatial residual structure because temporal increments often exhibit a lower signal-to-noise ratio than standing stock. Nevertheless, the same sampling framework can in principle be extended to repeated inventories by redefining the response variable as carbon-stock change and updating both the auxiliary-variable model and residual covariance structure. Such an extension would require independent multi-temporal reference data and should therefore be evaluated separately.

6.4 Implications for CCER-oriented carbon accounting

The CCER framework is primarily concerned with project-level carbon accounting and associated uncertainty rather than the production of spatially continuous carbon-stock maps. Our results therefore should not be interpreted as a direct replacement for the current CCER estimator. Instead, they show how wall-to-wall remote-sensing information and optimized field sampling could complement project-level accounting when spatially explicit information is also required.

For comparison with conventional mean-based estimation, we represented the estimated project mean as a spatially constant benchmark map. Relative to this benchmark, the recommended combination of optimized sampling and regression kriging reduced mapping RMSE by 25-39% across the evaluated sample sizes, with larger improvements at higher sampling effort. These gains are relevant to applications that require spatial information in addition to a project total, including targeted field verification, stand-level

2026 S. T. Yau High School Science Award
仅用于2026丘成桐中学科学奖论文公开

management, and spatial risk assessment. The results therefore support the use of remote-sensing-assisted spatial estimation as a complementary component of forest-carbon MRV rather than as a substitute for the accounting requirements specified by CCER.

6.5 Limitations

(1) Reference map. The benchmark is a LiDAR-derived carbon map, not field-measured carbon in each cell. Tree diameters were checked against 101 measured trees (rRMSE 6.5%), but only 39.6% of those trees fall inside the block and the step from diameter to carbon was not checked. Section 4.7 shows that the ranking of designs survives four kinds of reference error; it cannot rule out errors unrelated to anything the satellite observes. The simulated plots also carry no field measurement error, which would reduce the gains somewhat.

(2) One block. The study uses a single 480 m × 480 m plantation block, one date and one scale. Natural or mixed forests, steeper terrain and larger areas would need the prior and the variogram recalibrated, and the farm- and national-scale figures of Section 5.6 illustrate the method rather than test it.

(3) Statistical simplifications. The hybrid weights ignore the correlation between the two errors, simulated annealing may miss a better layout, and no independent prior-period data were used. These affect the ranking within the hybrid family and the size of the gains, not the comparison between designs.

(4) Economic model. Unit costs come from a fee guideline, carbon prices from 2024-2025 market data, and the issuance Q is a scenario. The value of map accuracy enters only through the scenario multiple κ. Deduction risks are computed from single-period stock plots, while credits depend on a change in stock, so they are a lower bound (Section 5.7). The welfare comparison values over-issue at the carbon price, and its standard errors are too wide to locate the planner's optimal campaign. Risk aversion is a single CARA parameter; a dynamic model of investment timing and a game in which the verifier chooses audit effort are left for future work.

7 Conclusion

This paper treated forest-carbon verification as a purchase of information by a project developer. On 576 LiDAR-derived reference cells at Saihanba, where the plots go matters most when plots are few. At 27 plots optimized placement cut mapping RMSE by 22% against random placement, and between 18 and 54 plots it reached the same map accuracy with 32-46% fewer plots; by 90 plots the advantage is gone. For the project mean it gave an 80% chance of an error within 5% with about 23% fewer plots than zone-random sampling. The plots earn their value mainly by calibrating the satellite regression,

2026 S.-T. (Siyu High School Science Award 仅用于2026江成南中学科学奖论文公示

so they should cover the range of the satellite indices.

The economic results follow from the cost structure of verification. The cost frontier is convex, so the shadow price of map accuracy rises steeply with the number of plots. Field cost is paid per plot while the deduction grows with the project, so verification has economies of scale: above about 50,000-140,000 t CO2e of creditable removals, at the carbon price and plot cost used here, the 10% deduction line rather than field cost sets the campaign, and the campaign rises with project scale (Proposition 5). The Appendix F rule scores the spread of plot values. Optimized placement keeps that spread stable, so near the developer's optimal campaign it faces the lowest deduction risk and is also the most accurate design (Proposition 4). The credits issued without carbon behind them are a cost to buyers that the developer does not bear; at its optimum they are worth 2.7 to 5.1 times the field bill under the scenario prices, and optimized placement roughly halves them against current practice.

For CCER practice the results point to two complementary changes. A residual-based uncertainty rule, with plots drawn at random within satellite strata and a seed set by the verifier, would cut the cost-minimizing campaign from 45 to 27 plots, but it lowers the external cost only together with representative placement. Both changes keep the 10% line and the official accounting rules; they change where the plots go and which statistic the verifier reads.

References

[1] Zeng, Y., Hao, D., Hu, B., et al. (2015). An optimal sampling design for observing and validating long-term leaf area index with temporal variations in spatial heterogeneities. Remote Sensing, 7(2), 1300-1319.

[2] National standard, Validation of remote sensing products of leaf area index, and the accompanying cLHS/SMP programs (MATLAB: cLHS.m, LHSobj.m).

[3] Ministry of Ecology and Environment (2023). Methodology for greenhouse gas voluntary emission reduction projects: Afforestation (CCER-01-001-A01). Appendix E (sampling design), Appendix F (uncertainty accounting), Table 35 (precision deduction).

[4] Minasny, B., & McBratney, A. B. (2006). A conditioned Latin hypercube method for sampling in the presence of ancillary information. Computers & Geosciences, 32(9), 1378-1388.

[5] Cochran, W. G. (1977). Sampling Techniques (3rd ed.). New York: Wiley.

[6] Matheron, G. (1963). Principles of geostatistics. Economic Geology, 58(8), 1246-1266.

[7] Journel, A. G., & Huijbregts, C. J. (1978). Mining Geostatistics. London: Academic Press.

[8] Odeh, I. O. A., McBratney, A. B., & Chittleborough, D. J. (1995). Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression-kriging. Geoderma, 67(3-4), 215-226.

[9] Hengl, T., Heuvelink, G. B. M., & Rossiter, D. G. (2007). About regression-kriging: From equations to case studies. Computers & Geosciences, 33(10), 1301-1315.

2026 S. T. Yau High School Science Award 仅用于2026丘成桐中学科学奖论文公示

[10] Cressie, N. (1993). Statistics for Spatial Data (Revised ed.). New York: Wiley.

[11] Moran, P. A. P. (1950). Notes on continuous stochastic phenomena. Biometrika, 37(1/2), 17-23.

[12] Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32.

[13] Webster, R., & Oliver, M. A. (2007). Geostatistics for Environmental Scientists (2nd ed.). Chichester: Wiley.

[14] Goovaerts, P. (1997). Geostatistics for Natural Resources Evaluation. New York: Oxford University Press.

[15] Gibbs, H. K., Brown, S., Niles, J. O., & Foley, J. A. (2007). Monitoring and estimating tropical forest carbon stocks: making REDD a reality. Environmental Research Letters, 2(4), 045023.

[16] Saatchi, S. S., Harris, N. L., Brown, S., et al. (2011). Benchmark map of forest carbon stocks in tropical regions across three continents. Proceedings of the National Academy of Sciences, 108(24), 9899-9904.

[17] Särndal, C.-E., Swensson, B., & Wretman, J. (1992). Model Assisted Survey Sampling. New York: Springer.

[18] Gregoire, T. G., & Valentine, H. T. (2008). Sampling Strategies for Natural Resources and the Environment. Boca Raton: Chapman & Hall/CRC.

[19] Howard, R. A. (1966). Information value theory. IEEE Transactions on Systems Science and Cybernetics, 2(1), 22-26.

[20] Raiffa, H., & Schlaifer, R. (1961). Applied Statistical Decision Theory. Boston: Graduate School of Business Administration, Harvard University.

[21] Akerlof, G. A. (1970). The market for "lemons": Quality uncertainty and the market mechanism. Quarterly Journal of Economics, 84(3), 488-500.

[22] Badgley, G., Freeman, J., Hamman, J. J., et al. (2022). Systematic over-crediting in California's forest carbon offsets program. Global Change Biology, 28(4), 1433-1445.

[23] National Voluntary Greenhouse Gas Emission Reduction Trading System (2026). Overview of the national voluntary emission reduction trading market in 2025. https://www.ccer.com.cn

[24] Ministry of Ecology and Environment (2025). Allowance trading and compliance of the national carbon emission trading market for 2024 completed. Press release, 5 January 2025.

[25] Forestry survey and planning project fee guideline (industry guidance document), Article 11. https://www.dzxzch.cn/zcjx/tdsy/202005/P020200508346589264495.pdf

[26] National Forestry and Grassland Administration. How is forest coverage surveyed? (Ninth national continuous forest inventory, 415,000 fixed plots).

2026 S.T. Yau High School Science Award 仅用于2026丘成桐中学科学奖公示

https://www.forestry.gov.cn/c/www/gkzcjd/84085.jhtml

[27] Stigler, G. J. (1961). The economics of information. Journal of Political Economy, 69(3), 213-225.

[28] Townsend, R. M. (1979). Optimal contracts and competitive markets with costly state verification. Journal of Economic Theory, 21(2), 265-293.

[29] Holmström, B. (1979). Moral hazard and observability. Bell Journal of Economics, 10(1), 74-91.

[30] Pratt, J. W. (1964). Risk aversion in the small and in the large. Econometrica, 32(1/2), 122-136.

[31] Milgrom, P., & Shannon, C. (1994). Monotone comparative statics. Econometrica, 62(1), 157-180.

[32] Blackwell, D. (1953). Equivalent comparisons of experiments. Annals of Mathematical Statistics, 24(2), 265-272.

[33] Mason, C. F., & Plantinga, A. J. (2013). The additionality problem with offsets: Optimal contracts for carbon sequestration in forests. Journal of Environmental Economics and Management, 66(1), 1-14.

[34] Duflo, E., Greenstone, M., Pande, R., & Ryan, N. (2013). Truth-telling by third-party auditors and the response of polluting firms: Experimental evidence from India. Quarterly Journal of Economics, 128(4), 1499-1545.

2026 S.-T. Yau High School Science Award
仅用于2026丘成桐中学科学奖公示