https://doi.org/10.58805/itamns.V.1(32)-42

МНОГОСЦЕНАРНЫЙ СИНТЕТИЧЕСКИЙ ДРЕВЕСНО-КОЛЬЦЕВОЙ СТЕНД ДЛЯ ВНЕВЫБОРОЧНОЙ ОЦЕНКИ МОДЕЛЕЙ ДИАМЕТРАЛЬНОГО ПРИРОСТА ВИДОВ, РАСПРОСТРАНЁННЫХ В КАЗАХСТАНЕ

К.М. Акишев[1], **М.Б. Тлебаев**[2], **А. Ешмуратов**[1]✉

[1] *Казахский университет технологии и бизнеса им. К. Кулажанова, Астана, Казахстан*
[2] *Таразский университет им. М.Х. Дулати, Тараз, Казахстан*
✉ *Корреспондент- автор: assylzhanyeshmuratov@gmail.com*

Received: 23 July 2026
Accepted: 11 September 2026
Published: 30 September 2026
Citation: A. Yeshmuratov, et al. « МНОГОСЦЕНАРНЫЙ СИНТЕТИЧЕСКИЙ ДРЕВЕСНО-КОЛЬЦЕВОЙ СТЕНД ДЛЯ ВНЕВЫБОРОЧНОЙ ОЦЕНКИ МОДЕЛЕЙ ДИАМЕТРАЛЬНОГО ПРИРОСТА ВИДОВ, РАСПРОСТРАНЁННЫХ В КАЗАХСТАНЕ ». Information Technologies, Automation, and Management in Natural Sciences, vol. 1, no. 32, Sept. 2026, doi: https://doi.org/10.58805/itamns.V.1(32)-42
License CC BY 4.0

Исследование выполнено исключительно на синтетических данных: условные деревья, годичные кольца, геометрические маски и высокочастотный измерительный канал не являются полевыми наблюдениями Казахстана. Цель работы состояла в создании многосценарного стенда для проверки выбора модели и устойчивости вневыборочного вывода. Проведено 100 независимых прогонов по четырём порождающим сценариям; каждый прогон включал 120 деревьев четырёх видов, 3 120 годовых и 2 520 модельных строк. Сопоставлены AR (5), OLS, Ridge, ElasticNet, kNN, Random Forest и Gradient Boosting. При практической эквивалентности ошибок в пределах 1 % применялось правило простоты. В линейном сценарии OLS выбиралась в 64 % прогонов, тогда как в нелинейном kNN выбиралась в 96 %. Медианный выигрыш на новых деревьях составлял 25,27–37,23 %. Увеличение выборки с 40 до 120 деревьев уменьшило медианную ширину бутстрэп-интервала ΔMAE с 0,278 до 0,171 мм. Для сценария режимного сдвига 2023–2025 гг. медианный R^2 равнялся –0,089, а критерий G проходил в 24 % прогонов. Результаты показывают зависимость ранжирования моделей от механизма генерации и ограничивают вывод областью вычислительного стенда.

Ключевые слова: синтетический стенд, годичные кольца, диаметральный прирост, измерительный канал, simulation study, вневыборочная оценка, NEL1, Казахстан.

ҚАЗАҚСТАНДА ТАРАЛҒАН АҒАШ ТҮРЛЕРІНІҢ ДИАМЕТРЛІК ӨСІМ МОДЕЛЬДЕРІН ІРІКТЕМЕДЕН ТЫС БАҒАЛАУҒА АРНАЛҒАН КӨПСЦЕНАРИЙЛІ СИНТЕТИКАЛЫҚ АҒАШ-САҚИНАЛЫ СТЕНД

Қ.М. Ақышев[1], **М.Б. Тлебаев**[2], **А. Ешмұратов**[1]✉

[1] *Қ. Құлажанов атындағы Қазақ технология және бизнес университеті, Астана, Қазақстан*
[2] *М.Х. Дулати атындағы Тараз университеті, Тараз, Қазақстан*
✉ *e-mail: assylzhanyeshmuratov@gmail.com*

Зерттеу тек синтетикалық деректермен орындалды: шартты ағаштар, жылдық сақиналар, геометриялық маскалар және жоғары жиілікті өлшеу арнасы Қазақстандағы далалық бақылаулар емес. Жұмыстың мақсаты модельді таңдауды және іріктемеден тыс қорытындының тұрақтылығын тексеретін көпсценарийлі стенд құру болды. Төрт генеративтік сценарий бойынша 100 тәуелсіз іске қосу орындалды; әр іске қосуда төрт түрдің 120 ағашы, 3 120 жылдық және 2 520 модельдік жол болды. AR (5), OLS, Ridge, ElasticNet, kNN, Random Forest

және Gradient Boosting салыстырылды. Қателер 1 % шегінде практикалық баламалы болғанда қарапайымдық ережесі қолданылды. Сызықтық сценарийде OLS іске қосулардың 64 %-ында, ал сызықтық емес сценарийде kNN 96 %-ында таңдалды. Жаңа ағаштардағы медианалық ұтыс 25,27–37,23 % болды. Ағаш санын 40-тан 120-ға арттыру ΔMAE бутстрэп интервалының медианалық енін 0,278-ден 0,171 мм-ге дейін азайтты. 2023-2025 жылдардағы режимдік ығысу сценарийінде медианалық R^2 0,089 болды, ал G критерийі іске қосулардың 24 %-ында орындалды. Нәтижелер модельдер рейтингі генерация механизміне тәуелді екенін көрсетеді және қорытындыны есептеу стендімен шектейді.

Түйін сөздер: синтетикалық стенд, жылдық сақиналар, диаметрлік өсім, өлшеу арнасы, simulation study, іріктемеден тыс бағалау, NEL1, Қазақстан.

A MULTI-SCENARIO SYNTHETIC TREE-RING TESTBED FOR OUT-OF-SAMPLE EVALUATION OF DIAMETER-INCREMENT MODELS FOR SPECIES OCCURRING IN KAZAKHSTAN

K.M. Akishev[1], M.B. Tlebayev[2], A. Yeshmuratov[1]✉

[1] K. Kulazhanov Kazakh University of Technology and Business, Astana, Kazakhstan
[2] Taraz University named after M.Kh. Dulaty, Taraz, Kazakhstan
✉ *e-mail: assylzhanyeshmuratov@gmail.com*

The study uses synthetic data only: hypothetical trees, annual rings, geometric masks, and the high-frequency measurement channel are not field observations from Kazakhstan. The aim was to develop a multi-scenario testbed for model selection and out-of-sample claim evaluation. One hundred independent runs were conducted under four data-generating scenarios; each run contained 120 trees from four species, 3,120 annual rows, and 2,520 modeling rows. AR (5), OLS, Ridge, ElasticNet, kNN, Random Forest, and Gradient Boosting were compared. A prespecified simplicity rule was used when validation errors were practically equivalent within 1%. OLS was selected in 64% of linear runs, whereas kNN was selected in 96% of nonlinear runs. Median gain over Persistence on unseen trees ranged from 25.27% to 37.23%. Increasing sample size from 40 to 120 trees reduced median bootstrap interval width for ΔMAE from 0.278 to 0.171 mm. Under the 2023–2025 regime-shift scenario, median R^2 was –0.089 and gate G passed in 24% of runs. The results demonstrate that model ranking depends on the generating mechanism and support claims only within the declared computational testbed.

Keywords: synthetic testbed, tree rings, diameter increment, measurement channel, simulation study, out-of-sample evaluation, NEL1, Kazakhstan.

Introduction. Tree-growth forecasting becomes ambiguous when the biological quantity, the sensor output, and the model prediction are treated as the same object. A completed annual ring records radial increment along a particular direction. A dendrometer trace, however, combines irreversible growth with water-driven stem deformation, thermal drift, gaps, and installation effects. Unless these levels are separated explicitly, model error can be misread as sensor error or, more seriously, as evidence about growth in a natural population.

Evidence from Kazakhstan and neighboring mountain systems shows that annual-ring response changes with snow conditions, elevation, and water stress [1-4]. Its interpretation also depends on the selected climate variable and on how the growing season is delimited [5-6]. At the subdaily scale, dendrometer records reveal species-specific dynamics and nocturnal growth, while also requiring separation of stem water deficit from irreversible increment [7-10]. For that reason, the coefficients used below are not transferred to Kazakhstani forests. Species names serve only to label four contrasting synthetic groups.

The literature published in 2020-2026 represents three data layers that should not be substituted for one another. Tree-ring chronologies describe an annual retrospective response but contain no within-season deformation record [1-6, 11-13]. Dendrometers resolve subdaily change, yet their signals mix growth with reversible water-status variation [7-10, 14-19]. Cross-sectional images support inspection of ring boundaries but do not provide a forecasting sequence by themselves [20-22]. A combined workflow must therefore state which variable is forecast, which channel is observed, and which information is retained solely for geometric control.

In this work, publications establish terminology and design constraints rather than constitute proof of algorithmic effectiveness. A bounded set of sources

cannot demonstrate that no analogous system exists, nor can it validate a model. All quantitative claims are instead tied to the declared simulation study. Their evidential scope ends with the generator, parameter ranges, and tests reported here.

No single reference provides a direct counterpart to the entire protocol. Klesse et al. investigated biological memory with synthetic ring and climate series [23]. Heilman et al. forecast growth from empirical ring and inventory records while propagating uncertainty [24], and Evans et al. connected radial increment with diameter through a hierarchical formulation [25]. INTRAGRO operates on real dendrometer and climate streams [17]. Other studies contribute annotated cross-sections [20-22] or evaluation principles such as independent testing, leakage control, and publication of unsuccessful outcomes [26-29].

These strands define a methodological space rather than a ready-made solution. Heilman et al. [24] and Evans et al. [25] reconcile increment and diameter using empirical observations and domain-specific probabilistic models. Klesse et al. [23], in contrast, show what can be learned when the memory mechanism is known by construction. The testbed developed here sits between those approaches: it estimates no natural effect, but asks whether the preferred model changes when memory, nonlinearity, observation noise, and the late-period regime are controlled experimentally.

Four evaluation rules were taken from the forecasting literature: model choice and final testing must be separated; the same tree must not occur in more than one partition; chronological order must be respected; and performance relative to a baseline must be accompanied by uncertainty [26-30]. The implementation therefore keeps runs with no gain, intervals that include zero, and negative R^2. Such outcomes count as failed realizations rather than being removed from the reported pass rate.

Table 1 locates the proposed protocol and the NEL1 methodological comparator [31] within this literature by recording, for each comparator, the data origin and the presence of grouped, temporal, exclusion, and reproducibility controls.

Table 1 - Comparison of related studies with the proposed computational testbed

Source	Data	Rg;	C;	D;	G;	T;	E;	O;
Klesse et al. [23]	S/R	+	–	–	+	±	–	+
Heilman et al. [24]	R	+	–	+	+	+	–	±
Evans et al. [25]	R	+	–	+	–	–	–	±
Aryal et al. [17]	R	–	+	+	±	+	–	±
Marichal et al. [22]	R	+	–	–	–	–	–	+
Cerqueira et al. [26]	R/S	–	–	–	+	+	–	±
Stock et al. [28]	review	–	–	–	+	±	±	–
NEL1 [31]	S	+	–	+	+	+	+	+
Current study	S	+	+	+	+	+	±	±

The notation in Table 1 is as follows: R, real data; S, synthetic data; Rg, tree rings; C, a high-frequency measurement channel; D, prediction of diameter increment; G, grouped independent testing; T, temporal holdout; E, exclusion of a species or source; and O, accessible data or code. Partial coverage is marked by ±. The comparison is not presented as proof that an analogue is absent. It shows only that the reviewed studies did not combine, in one protocol, several DGPs, a synthetic high-frequency channel, repeated grouped tests, a sample-size experiment, and sensitivity analysis for a fail-closed decision rule.

The research object is an out-of-sample procedure for generating and evaluating synthetic annual diameter-increment series. The research subject is the dependence of the selected family and the stability of its result on the DGP, observation distortion, sample size, withheld trees, a late chronological segment, and exclusion of one hypothetical species group.

Accordingly, the aim was to construct a multi-scenario synthetic testbed and assess whether the selected procedure continued to outperform Persistence when the generating mechanism, object, period, group, and sample size were changed.

The experiment examined three expectations. Model rank was expected to vary across DGPs; the cluster-bootstrap interval was expected to contract as

the number of independent trees increased; and a diagnosed late-period shift was expected to block the adequacy claim even if MAE remained lower than that of Persistence. These expectations concern the computational design and are not hypotheses about forest populations.

The aim was operationalized in three tasks. First, four DGP classes and a daily measurement channel with robust annual aggregation were specified. Second, seven model families were compared over 100 independently generated runs containing 120 trees each and using repeated grouped partitions. Third, the study estimated distributions of MAE, ΔMAE, R^2, bootstrap-interval width, and the pass rate of G at gain thresholds of 2.5%, 5%, and 10%.

The contribution is a set of testable protocol distinctions rather than a claim of global priority. It combines several generators and random seeds, a practical-equivalence zone that prevents a numerical tie from being overinterpreted, and joint use of R^2, a tree-cluster bootstrap interval, and threshold sensitivity for both favorable and unfavorable tests. The stated novelty is therefore computational; it does not imply biological validation or metrological accuracy of a dendrometer.

Materials and methods. The simulation study crossed four DGPs with 25 independent seeds, yielding 100 complete runs. Every run generated 120 hypothetical trees, equally divided among Picea schrenkiana, Pinus sylvestris, Betula pendula, and Populus nigra, over the years 2000-2025. The scale parameter μs ranged from 0.85 to 1.15 of its baseline, while βm, βh, and σ(ε) ranged from 0.80 to 1.20. These ±15% and ±20% ranges are engineering stress bounds chosen to preserve parameter signs and baseline ordering while avoiding one fixed configuration; they are neither field estimates nor confidence intervals.

A complete generated run, not a resampled row set, was the replication unit. Each seed produced new tree effects, climate indices, missing values, outliers, and channel distortions. Consequently, the 25 runs within a DGP are 25 distinct synthetic populations and time series, with parameter perturbation added on top of the structural differences between DGPs.

Species labels encode contrasting scales and response coefficients only. They do not reproduce Kazakhstani populations, site zones, age distributions, or geographic ranges. Equal groups of 30 objects prevent one label from dominating training or exclusion tests; any ranking between labels must therefore be read as a property of the assigned parameters, not as a biological comparison.

Table 2 records the four prespecified mechanisms. They were fixed before errors were evaluated so that any change in model rank could be related to a known change in the generating relationship or measurement process.

Table 2 - Data-generating and measurement-process scenarios

DGP	Change in generating relationship	Measurement channel	Property tested
Linear	q(c) = 0; baseline lag and climate dependence	ρ = 0.72; missingness 1%; outliers 0.1%	favorable setting for linear candidates
Nonlinear	tanh(MH), thresholds M > 0.50 and H > 0.80, $\sin(1.5z_{t-1})$	as in the linear DGP	interactions and threshold effects
Regime shift	from 2023: M–0.90; H+1.10; AR coefficients and level changed	as in the linear DGP	quantitatively pronounced late-period shift
Noise/ missingness	σ(ε) depends on |H|; annual shocks in 2.5% of years	ρ = 0.88; missingness 6%; outliers 0.8%; intensified drift	robustness to observation distortion

The settings in Table 2 are computational stress conditions, not specifications of a physical instrument. Autocorrelation values of $\rho = 0.72$ and 0.88, missingness of 1% and 6%, and outlier rates of 0.1% and 0.8% define moderate and intensified distortion. Ring widths were clipped to 0.18–5.50 mm to constrain the exponential generator numerically. Mask eccentricity ranged from 0.02 to 0.10, but neither mask geometry nor clipping limits entered the predictor set.

The linear DGP provides a positive control for linear estimators. The nonlinear DGP introduces moisture–heat interaction, threshold behavior, and nonlinear memory. Only the regime-shift DGP changes both predictor distributions and coefficients after 2022, so a

late shift is anticipated there a priori. The noise DGP leaves the target relationship intact while increasing heteroscedasticity, autocorrelation, gaps, and outliers in the measurement channel. This separation allows specification error to be distinguished from observation error.

Five annual lags were retained as a design choice, not as an estimate of biological memory. Candidate features include lags 1–5, whereas the generator uses only lags 1, 3, and 5. The unused intermediate lags conceal the true sparsity pattern without withholding the broader memory window. Years 2000–2004 initialize that window, so modeling begins in 2005; an earlier draft's 2012 start, which had no connection to lag construction, was removed.

The latent component is defined by expression (1):

$$\begin{aligned} z_{i,t} = {} & b_1 z_{i,t-1} + b_3 z_{i,t-3} + b_5 z_{i,t-5} \\ & + \beta_m M_{i,t} - \beta_h H_{i,t} + b_i \\ & - 0.08 a_{i,t} + q(c)_{i,t} + \varepsilon_{i,t} \end{aligned} \tag{1}$$

where $z_{i,t}$ – latent component for tree i in year t;
b_1, b_3, b_5 – memory coefficients at lags 1, 3, and 5;
$M_{i,t}, H_{i,t}$ – scenario indices of moisture and heat stress;
β_m, β_h – species-specific sensitivity coefficients;
b_i – individual tree effect;
$a_{i,t}$ – normalized ring number;
$q(c)_{i,t}$ – additive term for scenario c, equal to zero in the linear DGP;
$\varepsilon_{i,t}$ – random annual disturbance.

Expressions (2) and (3) convert the latent component into ring width and diameter increment:

$$w_{i,t} = \mathrm{clip}\big(\mu_s \exp(z_{i,t}), 0.18, 5.50\big) \tag{2}$$

where $w_{i,t}$ – ring width, mm;
μ_s – scale of species s;
clip – truncation of a value at the lower and upper bounds;
0.18 and 5.50 – scenario ring-width bounds, mm.

Expression (3) then accumulates annual diameter increment:

$$g^D_{i,t} = 2 w_{i,t}; \quad D_{i,t} = D_{i,0} + \sum_{k=1}^{t} g^D_{i,k} \tag{3}$$

where $g^D_{i,t}$ – true synthetic annual diameter increment, mm;
$D_{i,0}$ – initial diameter, mm;
$D_{i,t}$ – cumulative diameter after year t, mm;
k, t – summation index and the last year included;
2 – conversion from radial width to diameter increment in the symmetric model.

The measurement channel was generated at 365 observations per year rather than by adding a single annual Gaussian error. Its daily signal follows expression (4):

$$\begin{aligned} x_{i,t,d} = {} & g^D_{i,t} F(d) + A_{i,t} \sin\left(\frac{2\pi d}{14} + \varphi_i\right) \\ & + \gamma\big(T_d - \overline{T}\big) + v_{i,t,d} + \kappa_i d + o_{i,t,d}, \\ v_{i,t,d} = {} & \rho v_{i,t,d-1} + \xi_{i,t,d} \end{aligned} \tag{4}$$

where $x_{i,t,d}$ – observed signal on day d;
$F(d)$ – normalized seasonal cumulative function;
$A_{i,t}$ – amplitude of reversible water-related deformation;
φ_i – individual phase of the short-period component;
γ – coefficient of the temperature term;
$T_d, \overline{T}$ – daily temperature and its mean;
$v_{i,t,d}$ – autocorrelated measurement disturbance;
ρ – AR(1) coefficient, equal to 0.72 or 0.88;
$\xi_{i,t,d}$ – random innovation of the AR(1) process;
κ_i – linear channel drift;
$o_{i,t,d}$ – rare outlier; missing observations were excluded from aggregation.

Within expression (4), $F(d)$ accumulates the irreversible component monotonically through the growing season. The sinusoid is a functional surrogate for reversible short-period deformation, not a claim about one physical process. Temperature response, AR(1) error, and linear drift add correlated distortions on different time scales. Missingness is sampled by day, while rare outliers receive enlarged amplitudes; both remain documented in the daily output.

Annualization compares the medians of the first and last 21-day windows. This operator was chosen before running the experiment because a window median is less sensitive to an isolated outlier than an endpoint difference. It was not retuned for individual DGPs. The resulting annual value therefore carries the joint effects of gaps, drift, and incomplete seasonality and should not be treated as exact recovery of latent growth.

The observed annual increment is obtained from expression (5):

$$\tilde{g}^D_{i,t} = \mathrm{median}\,\{x_{i,t,d} : d = 345, \ldots, 365\} - \mathrm{median}\,\{x_{i,t,d} : d = 1, \ldots, 21\} \tag{5}$$

where $\tilde{g}^D_{i,t}$ – aggregated synthetic measurement increment, mm;
median – median of the available observations;
$1, \ldots, 21$ – initial 21-day window;
$345, \ldots, 365$ – final 21-day window.

Figure 1 makes the distinction visible for one generated series: the daily channel retains reversible oscillation, drift, missing observations, and outliers before annual aggregation.

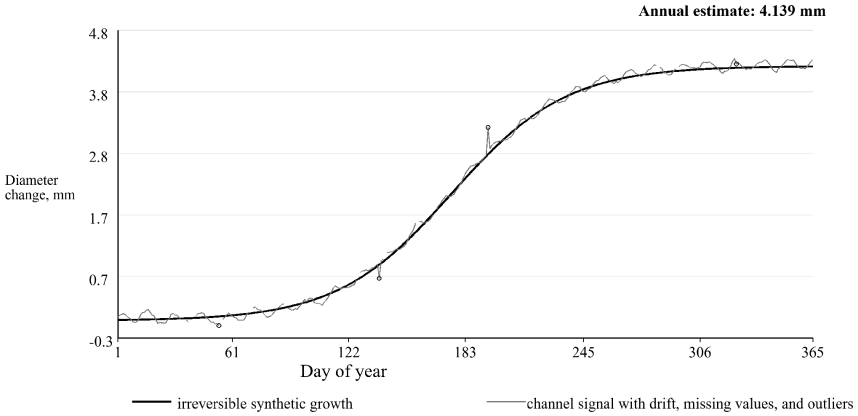


Fig. 1 - High-frequency synthetic measurement channel and robust annual aggregation

Source: authors' calculations

The trace in Figure 1 is entirely computational. Its aggregation error reflects the combined simulation of reversible motion, serial correlation, drift, and incomplete recording; it cannot be interpreted as the metrological error of any named dendrometer.

For geometric control, each annual width sequence is also converted deterministically into 26 nested boundaries. Figure 2 displays four outputs taken directly from the generator together with their identifiers.

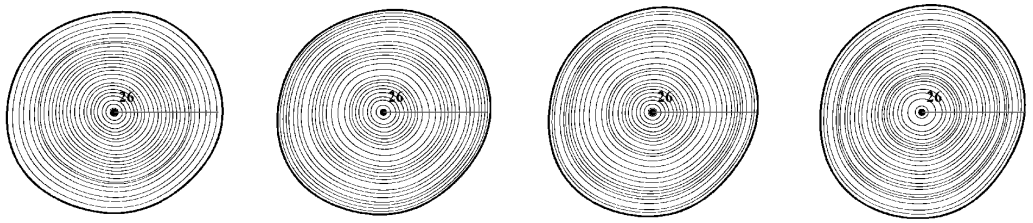


Fig. 2 - Geometric masks of four objects with identifiers and 26 boundaries

Source: authors' calculations

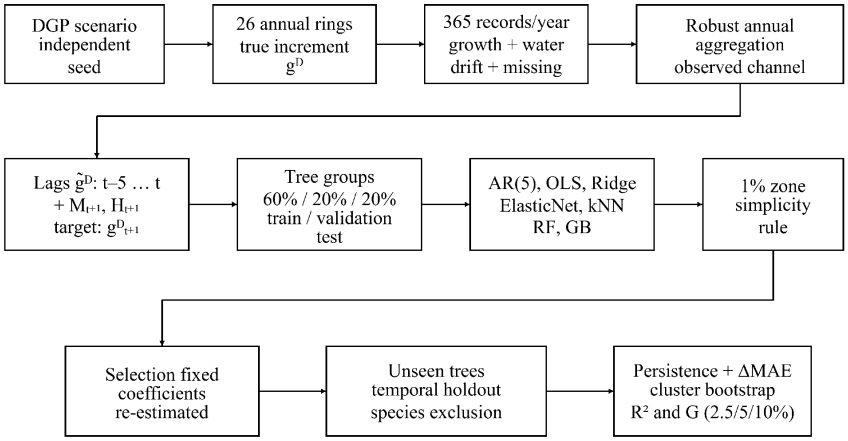


Fig. 3 - Generation, training, and out-of-sample evaluation workflow without block overlap

Source: prepared by the authors

Each mask is linked to annual records by tree_id, and its outer radius equals the accumulated widths. The masks are schematic rather than photographic, contain no wood texture, and are excluded from model training. They therefore document geometry without supporting a claim of multimodal prediction.

Figure 3 summarizes how generated records move from feature construction to partition-specific selection and closed testing.

For target year t+1, the feature vector contains the preceding five aggregated increments and the prespecified M and H indices for that target year (Figure 3). The 26-year series yields 21 model rows per tree for 2005–2025, consistent with the five-year initialization period.

Table 3 quantifies the resulting records, partitions, and provenance checks at run and experiment level.

Table 3 - Data volume, partitions, and provenance control

Metric	One run	100 runs	Control
Hypothetical trees	120; 30 per species	12,000 object-runs	tree_id unique within each seed
Annual rows	3,120	312,000	120 × 26
Modeling rows	2,520	252,000	120 × 21; 2005–2025
Daily observations	1,138,800	113,880,000	120 × 26 × 365
Grouped train/validation/test	72/24/24 trees	repeated for each seed	trees do not overlap partitions
Rows before 2023	2,160	216,000	18 years × 120 trees
temporal holdout	360 rows	36,000	2023–2025; 120 trees

Rows are repeated observations, whereas trees are the independent units represented in Table 3. Sample-size effects were evaluated in additional nested runs with n = 40, 80, and 120 trees. Forty seeds were used at each size, with ten assigned to each DGP.

The candidate set comprised AR (5), OLS, Ridge, ElasticNet, kNN, Random Forest, and Gradient Boosting; Persistence served only as the operational baseline by repeating lag 1. Ridge used $\alpha \in \{0.01, 0.1, 1, 10\}$; ElasticNet used four ($\alpha$, l1_ratio) combinations; kNN used $k \in \{5, 15, 31\}$; Random Forest used depths 4 and 6 with 20 trees; and Gradient Boosting used learning rates 0.05 and 0.10 with 30 depth-2 trees. The implementation environment was Python 3.12.13, NumPy 2.3.5, and pandas 3.0.1.

AR (5) uses only the five prior increments. OLS adds the known M and H indices without penalization; Ridge stabilizes correlated lag coefficients through an L2 penalty; and ElasticNet combines L1 and L2 penalties. Together these families span an unregularized temporal formulation and two regularized linear alternatives, for which small validation differences may represent practical equivalence rather than a meaningful ranking.

The nonlinear candidates cover different mechanisms. kNN estimates from nearby standardized feature vectors, Random Forest averages shallow randomized trees, and Gradient Boosting fits residual structure sequentially with small trees. Depth and ensemble size were deliberately limited because only 72 objects were available for training in a full run. The comparison therefore evaluates specified algorithms on a bounded grid, not an unspecified general-purpose model.

Means and standard deviations were fitted on training data and then carried unchanged to validation and test partitions. Although scaling is irrelevant to tree split locations, the same preprocessing interface was retained for every family. Species labels were encoded by fixed indicators. During exclusion tests, neither the excluded group's rows nor its statistics contributed to preprocessing. All transformations followed the grouped split.

Within each seed and species, trees were allocated 60%/20%/20% to training, validation, and test sets. Hyperparameters were chosen on validation data only. If several configurations fell within 1% of the minimum validation MAE, the fixed simplicity order AR (5), OLS, Ridge, ElasticNet, kNN, Random Forest, Gradient Boosting was applied. Family and configuration were then locked. Scaling and estimable coefficients were refitted only with data permitted by each closed test: training trees for the unseen-tree test, 2005–2022 for temporal holdout, and the three retained groups for species exclusion.

Three forms of controlled generalization were assessed. The unseen-tree partition contains 24 objects not used for selection. Temporal holdout evaluates all 120 trees in 2023–2025 after refitting on 2005–2022. In four exclusion tests, one hypothetical species group is withheld in turn while the remaining three are used for fitting. Here, out-of-sample denotes those explicit partitions and should not be confused with an independent field campaign.

The 1% band is a rule of practical equivalence, not a significance test. Added complexity was not accepted when its validation error improved on the minimum by less than one hundredth. Because the ordering was fixed before the closed test was viewed, it is especially useful for OLS, Ridge, and ElasticNet, whose forecasts can differ numerically under weak penalization without differing materially.

Prediction error and paired improvement over Persistence were calculated from expression (6):

$$\mathrm{MAE} = n^{-1}\sum_{r=1}^{n}|y_r - \hat{y}_r|; \quad \Delta\ \mathrm{MAE} = \mathrm{MAE}_{\mathrm{model}} - \mathrm{MAE}_{\mathrm{persistence}} \tag{6}$$

where n – number of test rows;

$y_r, \hat{y}_r$ – true increment and forecast for row r;

Δ MAE – paired error difference, for which a negative value is favorable;

$\mathrm{MAE}_{\mathrm{persistence}}$ – error of the Persistence baseline.

RMSE, bias, and R^2 supplemented MAE. For ΔMAE, uncertainty was estimated by 500 paired cluster-bootstrap resamples at tree level. Expression (7) combines the resulting interval with practical gain and adequacy.

Each bootstrap draw uses identical tree clusters for the selected model and Persistence, preserving dependence among the 21 annual rows belonging to one object. Rows were not resampled individually because they are not independent replicates. U_{95} was first obtained within a run; only then were run-level results summarized across seeds.

$$G_\tau = I\left[\left(1 - \frac{\mathrm{MAE}_{\mathrm{model}}}{\mathrm{MAE}_{\mathrm{persistence}}}\right)\cdot 100 \geq \tau \wedge U_{95}(\Delta\ \mathrm{MAE}) < 0 \wedge R^2 \geq 0\right] \tag{7}$$

where G_τ – indicator that the test passes at threshold τ;

$I[\cdot]$ – indicator that all conditions hold simultaneously;

τ – gain threshold of 2.5%, 5%, or 10%;

$U_{95}(\Delta\ \mathrm{MAE})$ – upper bound of the 95% cluster-bootstrap interval;

R^2 – coefficient of determination.

The main gain threshold was τ = 5%, with 2.5% and 10% retained for sensitivity analysis. This choice was recorded in the internal computational protocol before closed-test calculation, although no external preregistration was made. Seeds followed 2026082400 + 1000·j + r, where j indexes DGP and r indexes run. Late-period change between 2005–2022 and 2023–2025 was described by W_1, KS, and PSI.

Traceability was enforced through the seed formula, structural checks, and retained analysis records. Row counts, unique keys, admissible ranges, and tree_id separation were verified for every run; selected families, hyperparameters, candidate errors, and individual G conditions were stored before aggregation. These controls permit an internal audit, but open reproducibility still depends on deposition of the replication package.

All candidates shared the same feature and partition pipeline in Python 3.12.13 with NumPy 2.3.5 and pandas 3.0.1. Training was serial to keep random-operation order deterministic. Figures were rendered from stored run-level and summary CSV files, while image files themselves never entered the feature matrix.

Results and discussion. Structural validation succeeded for all 100 runs: expected record counts were obtained, each tree_id-year key was unique, and grouped partitions had no object overlap. Five lagged features generated 21 eligible years per tree without carrying information across objects. The robust aggregator returned annual values under the imposed gaps, although error rose under intensified noise. Stored ring widths and diameter increments also satisfied $g^D = 2w$ to numerical precision.

The provenance audit found no unseen-tree identifier in the corresponding training or validation set. Records

from 2023-2025 were excluded from family and hyperparameter selection in temporal holdout, and an excluded species supplied no scaling statistics. These results verify the internal table construction; they do not amount to replication by an independent generator.

A single run produced 1,138,800 daily channel observations. Missingness and outliers were fixed at 1% and 0.1% for the linear, nonlinear, and regime-shift DGPs, and at 6% and 0.8% for the noise DGP. The paired-window median still yielded one annual observation whenever both windows contained data. Models were therefore trained on an aggregated channel output rather than on the latent increment used to generate it.

Figure 4 presents all trajectories from the first stored linear run (seed = 2026082401), selected by run order rather than by its performance. A shared ordinate makes the four groups comparable: individual trees are thin lines, the group median is shown by the heavy line, and the interquartile range by the gray envelope.

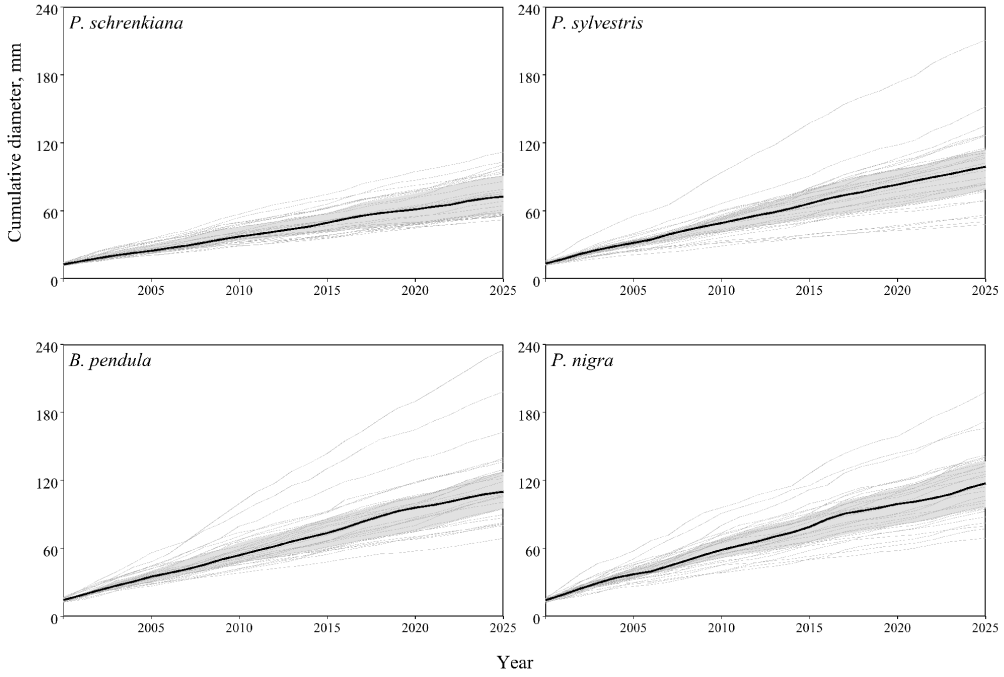


Fig. 4 – Trajectories of all 120 trees in the first linear run (seed = 2026082401) on a common diameter scale
Source: authors' calculations

Table 4 – MAE on unseen trees and model-selection frequency by scenario

Model	Linear	Nonlinear	Regime shift	Noise/missingness
AR(5)	0.823 [0.797–0.864]; 0%	1.177 [1.103–1.283]; 0%	0.843 [0.796–0.875]; 0%	0.866 [0.782–0.905]; 0%
OLS	0.687 [0.658–0.735]; 64%	0.958 [0.919–1.042]; 0%	0.698 [0.662–0.757]; 68%	0.716 [0.659–0.749]; 84%
Ridge	0.687 [0.658–0.735]; 4%	0.958 [0.918–1.041]; 0%	0.699 [0.662–0.757]; 0%	0.715 [0.658–0.749]; 0%
ElasticNet	0.687 [0.658–0.738]; 8%	0.958 [0.917–1.050]; 0%	0.698 [0.653–0.756]; 16%	0.715 [0.656–0.749]; 12%
kNN	0.704 [0.658–0.765]; 20%	0.830 [0.767–0.902]; 96%	0.710 [0.667–0.761]; 12%	0.715 [0.673–0.779]; 4%
Random Forest	0.721 [0.690–0.776]; 4%	0.862 [0.798–0.942]; 4%	0.721 [0.689–0.774]; 4%	0.718 [0.688–0.799]; 0%
Gradient Boosting	0.741 [0.706–0.825]; 0%	0.979 [0.929–1.072]; 0%	0.746 [0.697–0.823]; 0%	0.770 [0.711–0.823]; 0%

The between-group spread in Figure 4 is a designed property of the generator and should not be interpreted as realistic growth-rate separation. One extreme trajectory is left in place because it arose from the declared seed; no manual cleaning was applied after inspection.

Table 4 reports, for every candidate and DGP, the median unseen-tree MAE, its interquartile range in brackets, and the proportion of the 25 runs in which that family was selected.

Selection changed markedly with the generating mechanism (Table 4). OLS accounted for 64% of linear runs, 68% under regime shift, and 84% under intensified noise; its remaining shares were distributed among ElasticNet, Ridge, kNN, and Random Forest. The nonlinear DGP produced a different result: kNN was selected in 96% of runs and Random Forest in 4%. Thus, the experiment supplied no basis for naming one family as the general winner.

In the linear DGP, OLS, Ridge, and ElasticNet each rounded to a median MAE of 0.687 mm and had almost coincident IQRs. Their unequal selection frequencies are therefore attributable to seed-level validation variation and the equivalence rule, not to a stable absolute advantage for OLS. AR (5) reached 0.823 mm and was never chosen. The two ensemble methods likewise failed to gain enough validation accuracy to overcome their later position in the simplicity order. Because the three linear medians coincide after rounding, their selection shares describe the full validation procedure rather than an ordering inferred from the closed-test table.

Nonlinear generation separated the candidates more clearly. Median MAE was 0.830 mm for kNN, compared with 0.958 mm for OLS and Ridge, 0.862 mm for Random Forest, and 0.979 mm for Gradient Boosting. The 96% kNN selection rate consequently reflects a material gap rather than rounding noise. By contrast, regime shift and channel degradation again produced similar errors among the linear candidates and did not create a universal advantage for a more complex family on unseen trees. Added flexibility was useful when it matched the generator's interaction structure, but complexity alone did not protect against either coefficient change or channel distortion.

An exploratory check related unseen-tree gain to the run-specific parameter multipliers through Spearman correlation within each DGP, after averaging a multiplier across the four groups. The largest absolute correlations were 0.282 for β_m in the linear DGP, 0.555 for $\sigma(\varepsilon)$ in the nonlinear DGP, 0.614 for β_m under regime shift, and 0.553 for $\sigma(\varepsilon)$ under noise. With only 25 jointly varied runs per DGP, these coefficients describe local association and cannot isolate a causal parameter effect. Because the multipliers were neither orthogonal nor held constant one at a time, a large rank correlation may also reflect covariation with another perturbed parameter.

Selection rate and test MAE summarize different stages of the protocol. The former records the family that survived validation and the 1% equivalence rule for each seed. The latter is computed for every candidate, whether selected or not. A family can therefore show a similar cross-run median while being selected rarely if its validation advantage is unstable or falls inside the equivalence band. Keeping both summaries avoids retrospectively replacing the selection rule with closed-test ranking.

Across all DGPs, AR (5) had the largest median error, from 0.823 to 1.177 mm. In this generator, lag history alone was inadequate once M and H were available. AR (5) and Persistence nevertheless play different roles: AR (5) is fitted and competes for selection, whereas Persistence is fixed before every test. Retaining that fixed denominator makes gain comparable across scenarios and prevents baseline choice from following the observed result.

Random Forest medians were 0.721, 0.862, 0.721, and 0.718 mm for the linear, nonlinear, regime-shift, and noise DGPs. It improved on the linear candidates in the nonlinear case but remained behind kNN and was selected in only one run. Gradient Boosting was never selected and produced medians between 0.741 and 0.979 mm. These findings apply to the shallow trees and small ensembles prespecified for the available training size; they are not evidence against ensemble methods in general.

The IQR in Table 4 measures variability between independently generated runs, not uncertainty for one forecast. Under noisy observation, kNN and Random Forest show broader intervals than the penalized linear methods, which is compatible with local and split-based rules responding more strongly to distorted features. For the nonlinear DGP, however, the complete kNN interval lies at lower error. Distributional graphics are therefore read alongside selection frequency rather than used as a substitute for it.

For three DGPs in Figure 5, the linear-family medians differ only at approximately 10^{-6}–10^{-4} mm and are not treated as substantive separation. In the nonlinear panel,

the kNN interval is visibly displaced toward lower error relative to OLS, Ridge, and ElasticNet.

A separate nested experiment was used to examine how the number of trees affected uncertainty (Table 5).

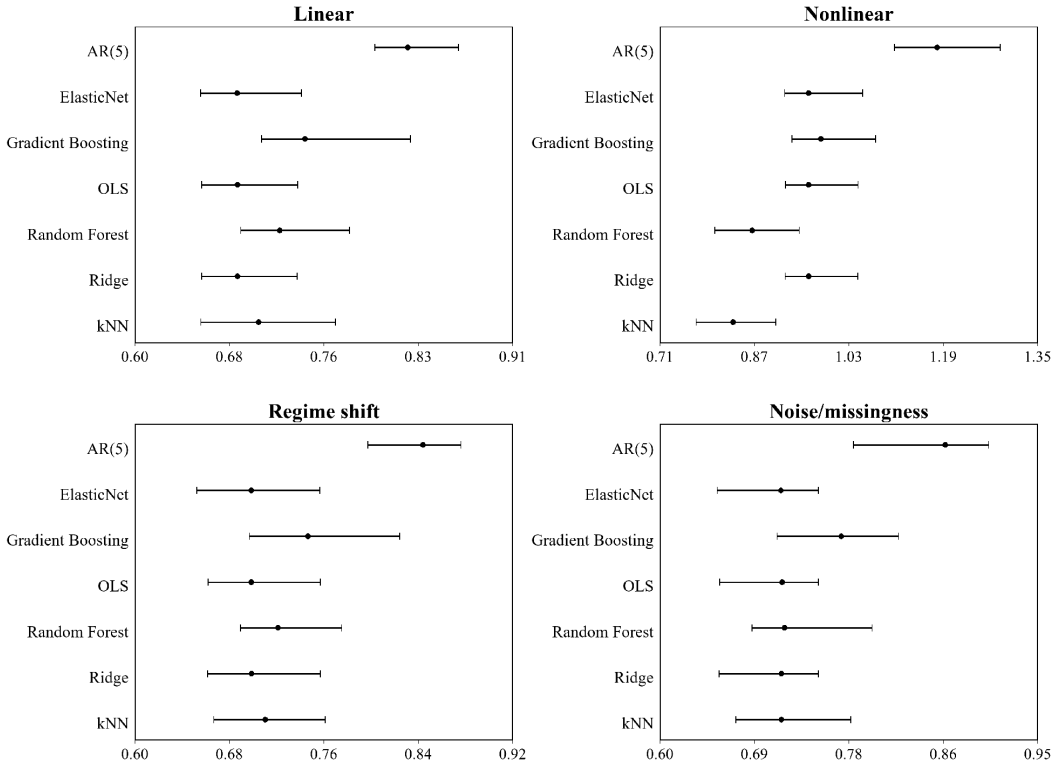


Fig. 5 - Median and IQR of MAE for seven models on unseen trees across 25 runs of each DGP *Source: authors' calculations*

Table 5 - Sample-size analysis based on 40 independent repetitions

n, trees	Number of runs	Width of 95% CI for ΔMAE, mm	Median gain, %	Frequency of the two leading families
40	40	0.278 [0.228–0.360]	25.85	OLS: 42.5%; kNN: 30.0%
80	40	0.212 [0.190–0.259]	27.63	OLS: 50.0%; kNN: 27.5%
120	40	0.171 [0.147–0.189]	26.98	OLS: 47.5%; kNN: 40.0%

Raising n from 40 to 120 reduced the median 95% interval width for ΔMAE from 0.278 to 0.171 mm, a decrease of 38.5%. OLS was selected in 42.5%, 50.0%, and 47.5% of runs at n = 40, 80, and 120; the corresponding kNN shares were 30.0%, 27.5%, and 40.0%. More trees therefore improved precision without producing monotonic convergence on one family across four mixed DGPs. The contraction follows from additional independent clusters, not from the 500 bootstrap resamples themselves. The 0.107 mm decrease in median interval width therefore supports collecting more independent trees before increasing resampling intensity within an unchanged dataset.

Median gain was 25.85%, 27.63%, and 26.98% at n = 40, 80, and 120. Because each size used newly generated seeds and repeated model selection, this modest non-monotonicity is not interpreted as an effect of sample size on mean performance. The supported sample-size result is the change in interval width. Table 6 next separates chronological testing from a diagnosed late-period regime change.

A pronounced target change appeared only where it had been built into the regime-shift DGP: median W_1 was 1.532 mm, KS was 0.454, and target PSI was 1.502. Target PSI remained at or below 0.035 in the other three DGPs. Their 2023–2025 observations are consequently described as a temporal holdout, not as evidence of source shift. This wording prevents a chronological split from being promoted to a distributional event solely because it occurs at the end of the record.

Table 6 - Diagnostics of differences between 2023-2025 and 2005-2022

Scenario	Target W_1	Target KS	Target PSI	Max. feature PSI	Interpretation
Linear	0.135	0.056	0.033	0.238	temporal holdout without a specified shift
Nonlinear	0.280	0.062	0.035	0.297	temporal holdout without a specified shift
Regime shift	1.532	0.454	1.502	1.631	shift confirmed
Noise/ missingness	0.156	0.051	0.033	0.230	temporal holdout without a specified shift

Maximum feature PSI rose to 1.631 under regime shift but did not exceed a median of 0.297 elsewhere. W_1, KS, and PSI were considered jointly as descriptive evidence of the prespecified change; no post hoc significance threshold was introduced. The late block was retained in stationary DGPs because chronological extrapolation remains informative even when the generator itself does not shift.

Table 7 consolidates all six closed tests. Entries are medians over 25 runs; U_{95} is the median run-specific upper cluster-bootstrap bound, and the last column gives the proportion satisfying G at $\tau = 5\%$. The label Shift identifies the DGP with the prespecified regime change.

Table 7 - Out-of-sample tests, R^2, and stability of criterion G

DGP	Test	n clusters	ΔMAE, mm	Gain, %	R^2;	U_{95}, mm	$G_5 = 1$, %
Lin.	Unseen trees	24	–0.233	25.27 [23.04–27.41]	0.710	–0.167	100
Lin.	2023–2025	120	–0.282	28.39 [25.68–30.74]	0.739	–0.196	100
Lin.	Excl. P. schrenkiana	30	–0.158	25.57 [19.32–29.87]	0.723	–0.088	88
Lin.	Excl. P. sylvestris	30	–0.176	24.52 [21.06–26.73]	0.693	–0.114	92
Lin.	Excl. B. pendula	30	–0.186	18.58 [15.65–21.14]	0.599	–0.107	100
Lin.	Excl. P. nigra	30	–0.358	24.87 [23.38–27.47]	0.574	–0.306	100
Nonlin.	Unseen trees	24	–0.471	37.23 [32.66–38.76]	0.787	–0.354	96
Nonlin.	2023–2025	120	–0.370	30.93 [28.44–33.11]	0.812	–0.257	100
Nonlin.	Excl. P. schrenkiana	30	–0.373	39.13 [34.08–40.49]	0.818	–0.277	100
Nonlin.	Excl. P. sylvestris	30	–0.437	38.75 [36.69–40.71]	0.803	–0.344	100
Nonlin.	Excl. B. pendula	30	–0.416	32.89 [29.70–35.31]	0.760	–0.321	100
Nonlin.	Excl. P. nigra	30	–0.455	28.05 [23.39–30.05]	0.729	–0.322	100
Shift	Unseen trees	24	–0.222	25.33 [20.53–27.02]	0.704	–0.148	100
Shift	2023–2025	120	–0.233	24.85 [21.90–26.93]	–0.089	–0.159	24
Shift	Excl. P. schrenkiana	30	–0.186	28.60 [24.07–35.69]	0.753	–0.114	88
Shift	Excl. P. sylvestris	30	–0.150	23.51 [16.32–26.12]	0.686	–0.090	84
Shift	Excl. B. pendula	30	–0.134	16.16 [13.26–18.49]	0.628	–0.073	100

DGP	Test	n clusters	ΔMAE, mm	Gain, %	R^2;	U_{95}, mm	G_5 = 1, %
Shift	Excl. P. nigra	30	–0.363	23.77 [22.26–26.18]	0.593	–0.283	100
Noise	Unseen trees	24	–0.238	25.41 [23.41–28.37]	0.704	–0.163	100
Noise	2023–2025	120	–0.268	27.72 [25.64–29.35]	0.722	–0.189	100
Noise	Excl. P. schrenkiana	30	–0.225	28.65 [19.83–33.83]	0.632	–0.139	84
Noise	Excl. P. sylvestris	30	–0.210	25.82 [22.21–29.40]	0.647	–0.151	100
Noise	Excl. B. pendula	30	–0.162	17.89 [15.52–21.68]	0.564	–0.083	96
Noise	Excl. P. nigra	30	–0.365	25.23 [23.83–26.70]	0.609	–0.282	100

For unseen trees, median gains over Persistence were 25.27% for the linear DGP, 37.23% for the nonlinear DGP, 25.33% under regime shift, and 25.41% under noise. G_5 succeeded in every run except one nonlinear realization, giving 96% there. The late block of the regime-shift DGP produced a different conclusion: MAE was still 24.85% lower than Persistence, but median R^2 fell to –0.089 and only 24% of runs passed G_5. A relative error gains alone therefore did not support adequacy. This unfavorable result is informative: beating a weak continuation rule does not guarantee that the forecast explains variation in the shifted target.

In the stationary DGPs, temporal holdout gains were 28.39% with R^2 = 0.739, 30.93% with R^2 = 0.812, and 27.72% with R^2 = 0.722 for linear, nonlinear, and noise generation. Every run met G_5, although shift diagnostics remained low. These rows show robustness to a later chronological segment within the same generator; they do not demonstrate adaptation to a changed source.

When one hypothetical species was excluded, every median gain remained positive, while G pass rates ranged from 84% to 100%. The lowest rate was not tied to one label across all DGPs. It resulted from the interaction of assigned group parameters, selected family, and seed. Reporting all four exclusions avoids drawing the conclusion from favorable labels alone.

These exclusion tests remain internal to a single synthetic family. Several DGP–group combinations yielded an upper ΔMAE bound above zero or a negative R^2, and those failures were retained. None of the exclusion rows establishes transfer to a natural population bearing the corresponding species name.

Figure 6 evaluates how the overall decision and the regime-shift temporal holdout respond to the three values of τ.

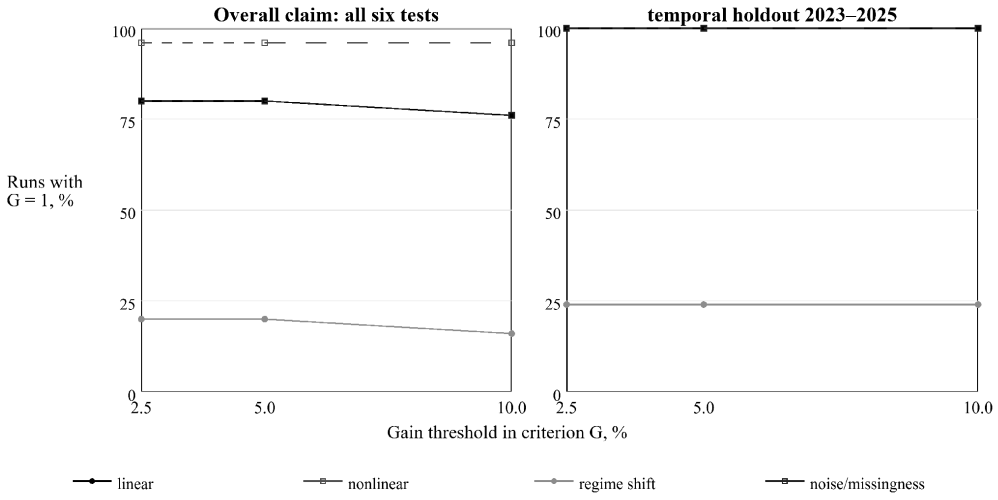


Fig. 6 - Proportion of runs with G = 1 at gain thresholds of 2.5%, 5%, and 10%.
Source: authors' calculations

For the six-test overall decision, the linear DGP passed in 88%, 88%, and 84% of runs at τ = 2.5%,

5%, and 10%. Nonlinear and noise generation passed in 100%, whereas the regime-shift DGP passed in 32%. Its temporal holdout result remained 24% at every τ because negative R^2, rather than the gain threshold, was the binding condition.

Coincident curves in Figure 6 should not be read as general insensitivity to required gain. In those tests, either the interval or R^2 had already blocked G, so changing τ could not alter the result. The linear overall decision did respond: raising τ to 10% lowered its pass rate from 88% to 84%, showing that some runs lay between the practical thresholds.

The central result concerns conditional model ranking, not algorithmic supremacy. Linear generation placed OLS, Ridge, and ElasticNet close together, whereas interaction and threshold structure favored kNN. Introducing four DGPs therefore removes the circular implication of a single linear generator, under which proximity among linear models was largely predetermined. The change in selection is thus tied to a controlled structural difference rather than attributed to an algorithm label in isolation.

This interpretation is compatible with Klesse et al. [23]: synthetic series can expose method behavior under a known memory process but cannot replace evaluation in a natural system. Heilman et al. [24] and Evans et al. [25] reach biological conclusions because they analyze empirical ring and inventory records. INTRAGRO [17] likewise uses measured subdaily dendrometer signals. The daily channel used here reproduces selected distortion patterns only and is not an empirical counterpart to those datasets.

The comparison with Klesse et al. [23] is particularly informative for lag interpretation. Their synthetic setting separates inherited autocorrelation from a biological claim. Here, the generator uses three of the five offered lags but does not disclose that sparsity to the candidates. Performance therefore measures how a family handles a partially redundant window; it does not establish five-year memory in any named species.

Uncertainty has a narrower meaning here than in Heilman et al. [24], who address new data and regional source fusion: our interval covers only the paired error difference within the declared synthetic test. The link between ring width and diameter is also simpler than the hierarchy of Evans et al. [25]. Deterministic doubling in expression (3) is auditable, but it omits eccentric stem geometry and cannot stand in for a measured diameter process.

The equivalence band changes what selection means. When OLS and Ridge produced nearly identical MAE, neither numerical minimum was promoted to substantive superiority. The simpler family was preferred only inside the 1% band; outside it, validation MAE governed selection. Because the rule precedes closed testing, it distinguishes a practical tie without reusing the test outcome. Without this band, minute seed-dependent differences among linear candidates would appear as frequent winner changes despite nearly identical test distributions.

The sample-size experiment also corrects the earlier reliance on only eight test trees. At $n = 40$, paired ΔMAE intervals were appreciably wider and selected-family frequencies varied more. Increasing bootstrap repetitions cannot compensate for missing trees because uncertainty is clustered at object level. Even the $n = 120$ interval remains conditional on the simulated DGPs and parameter ranges.

The non-monotonic family shares at $n = 40$, 80, and 120 do not conflict with narrower intervals. Interval width concerns a within-run paired contrast against Persistence, whereas selection frequency pools outcomes from four structurally different DGPs. More objects sharpen a contrast within a scenario but cannot remove between-scenario disagreement. Stable selection of one family would require a larger, DGP-specific sample-size experiment.

Late-period analysis shows why temporal holdout and regime shift must remain distinct terms. In the shifted DGP, W_1, KS, and PSI quantify the constructed change, and negative R^2 prevents an adequacy claim despite lower error than Persistence. In stationary DGPs, the same years form only a chronological test segment. Calling every late block, a source shift would therefore exceed the diagnostic evidence. Operationally, the distinction determines whether the result is described as ordinary late-period generalization or as performance under a changed data source.

Positive gain and negative R^2 are not contradictory. Persistence is a weak local comparator and can be worse than a forecast that still fails to explain variation in the new regime. R^2 adds comparison with the target mean, exposing that loss of explanatory value. The fail-closed rule requires all three conditions: practical gain, $U_{95}(\Delta\text{MAE})$ below zero, and nonnegative R^2. One failed condition is sufficient to withhold the claim.

Excluding a species label changes group composition without leaving the common generator. Its G rate therefore measures robustness to an omitted parametric

category, not transfer between natural species. A field test would require independent sites, populations, instruments, and seasons because spatial splitting and cross-source transfer alter performance estimates [32-33]. Likewise, the masks provide geometry checks only and are absent from the model features.

Operationally, exclusion asks whether the software remains stable when one encoded category is unavailable. Success supports robustness only within the assigned coefficient family; failure identifies where external evidence is needed. Since every group shares the same functional equations, a label such as P. nigra or B. pendula cannot by itself justify inference to the corresponding real species.

NEL1 [31] is adapted here as a protocol for limiting claims, not inserted as another forecasting algorithm. The retained elements are separated selection and testing, a fixed Persistence comparator, grouped provenance units, paired errors, and a fail-closed conclusion. This logic is related to reject-option methods [34]. The present implementation adds four DGPs, practical equivalence, a daily synthetic channel, shift diagnostics, and τ sensitivity, while also following recommendations on temporal evaluation, leakage, and ecological forecasting [26-29]. In this adaptation, NEL1 supplies the conditions for withholding a conclusion; it does not predetermine which candidate family should be selected.

Its scientific value is consequently bounded: the testbed can reveal DGP-dependent conclusions, partition leakage, and disagreement between evaluation metrics. Its practical role is preliminary verification of a software workflow before field deployment. It neither estimates tree age nor measures actual stem diameter, and it is not suitable for forestry decisions or validation of a physical sensor.

A subsequent field study would require a stricter data contract: persistent tree identifiers, initial diameters, daily sensor values with maintenance and missingness flags, weather covariates, completed annual increments, and an external ring-width reference. Trees and seasons must be partitioned before tuning. Cross-sections or cores should carry expert annotations and provenance, rather than being replaced by synthetic boundary masks.

Internal validity remains limited by hand-specified DGP forms, a restricted hyperparameter grid, and common code for generation and evaluation. Independent seeds, checksums, a fixed selection order, and retention of failed runs reduce but do not remove these risks. External validity is weaker still because real climate, instrument calibration, spatial site structure, and hardware differences are absent. The supported conclusion is therefore about protocol behavior, not a completed telemetry system. Common generation and evaluation code may reproduce the same unnoticed assumption in both stages, so independent reimplementation remains necessary.

Persistence is also the only operational baseline used by G. It transparently tests improvement over the latest increment but does not represent all reasonable forest forecasts. AR (5) provides a stronger temporal candidate inside selection, while Persistence remains fixed for comparability. A field protocol should declare several seasonal or climatological baselines in advance and should not replace them after inspecting the closed period.

The main limitations are the synthetic data origin, simplified seasonality, lack of dendrometer calibration, and absence of a persistent public identifier for the replication package. Main DGP parameters were not varied independently: their multipliers changed jointly, and 25 runs per DGP do not support stable separation of effects. The reported correlations are therefore exploratory and cannot replace a factorial sensitivity experiment over ranges justified by field measurements. A stronger design would vary one factor at a time or use an orthogonal scheme, repeat each configuration across seeds, and report interaction effects separately from main effects.

Conclusion. The completed experiment produced a multi-scenario testbed comprising 100 independent runs, 12,000 object-runs, 312,000 annual records, and 252,000 modeling rows. Robust annual increments were derived from 113.88 million daily synthetic observations containing water-related variation, drift, gaps, and outliers. The resulting records preserve a traceable distinction between latent increment, daily channel output, and the annual value supplied to the models.

Model choice was conditional on the generator. OLS was selected in 64-84% of runs for the linear, regime-shift, and noise DGPs, whereas kNN accounted for 96% of nonlinear runs. No family was universally optimal, and the 1% equivalence rule prevented near-ties among linear candidates from being reported as meaningful superiority.

Increasing the number of trees from 40 to 120 narrowed the median ΔMAE interval from 0.278 to 0.171 mm. For the diagnosed 2023–2025 regime shift, median R^2 was negative and only 24% of runs met

G_5 despite lower MAE than Persistence. Joint metrics therefore changed the conclusion that would have followed from relative error alone. This negative test forms part of the result rather than an exception removed from reporting.

These findings remain limited by the absence of field ring and dendrometer measurements from Kazakhstan, internal rather than external species exclusion, and the lack of a persistent public link for the replication package. The next phase should preregister the protocol and combine a calibrated dendrometer stream, independent diameter measurement, expert-annotated real cross-sections, and a genuinely closed evaluation season.

References

1. Utebekova A., Maisupova B., Mambetov B., Dosmanbetov D., Kelgenbayev N., Abzhanov T. Analysis of Tree-Ring Chronologies in Coniferous Forests of Southeastern Kazakhstan // OnLine Journal of Biological Sciences. - 2021. -Vol. 21(1). - P. 41–47. DOI 10.3844/ojbsci.2021.41.47.
2. Qin L., Bolatov K., Yuan Y. et al. The Spatially Inhomogeneous Influence of Snow on the Radial Growth of Schrenk Spruce (Picea schrenkiana Fisch. et Mey.) in the Ili-Balkhash Basin, Central Asia // Forests. −2022. -Vol. 13(1):44. DOI 10.3390/f13010044.
3. Li T., Liu Y., Cai Q., Duan X., Li P., Ren M., Ye Y. Water Stress-Induced Divergence Growth of Picea schrenkiana in the Western Tianshan and Its Forcing Mechanisms \// Forests. - 2023. -Vol. 14(2):354. DOI 10.3390/f14020354.
4. He J., Shen Z., Ning C., Zhang W., Halik Ü. Elevational Effects of Climate Warming on Tree Growth in a Picea schrenkiana Forest in the Eastern Tianshan Mountains \// Forests. - 2024.- Vol.15(12): 2052. DOI 10.3390/f15122052.
5. Dannenberg M.P. Modeling tree radial growth in a warming climate: where, when, and how much do potential evapotranspiration models matter? // Environmental Research Letters. – 2021.- Vol.16(8):084017. DOI 10.1088/1748-9326/ac1292.
6. Wolkovich E.M., Ettinger A.K., Chin A.R. et al. Why longer seasons with climate change may? not increase tree growth // Nature Climate Change. - 2025. -Vol. 15(12). -P. 1283–1292. DOI 10.1038/s41558-025-02476-x.
7. Wang X., Zhang Q. Interspecific Differences of Stem Diameter Variations in Response to Water Conditions for Six Tree Species in Northeast China // Forests. - 2023. Vol.- 14(4): 805. DOI 10.3390/f14040805.
8. Zhang Q., Lyu L., Wang Y. Patterns of daily stem growth in different tree species in a warm- temperate forest in northern China // Dendrochronologia. - 2022.- Vol. 72:125934. DOI 10.1016/j.dendro.2022.125934.
9. Zweifel R., Sterck F., Braun S. et al. Why trees grow at night // New Phytologist. - 2021. -Vol. 231(6)-. P. 2174-2185. DOI 10.1111/nph.17552.
10. Peters R.L., Basler D., Zweifel R. Et al. Normalized tree water deficit: an automated dendrometer signal to quantify drought stress in trees // New Phytologist. - 2025.- Vol. 247(3). - P. 1186-1198. DOI 10.1111/nph.70266.
11. Martinez-Garcia J., Stelzner I., Stelzner J., Gwerder D., Schuetz P. Automated 3D tree-ring detection and measurement from X-ray computed tomography // Dendrochronologia. - 2021.- Vol.69:125877. DOI 10.1016/j.dendro.2021.125877.
12. Zhang W., Zhao T., Su X. et al. A Tree Ring Measurement Method Based on Error Correction in Digital Image of Stem Analysis Disk // Forests. - 2021. -Vol. 12(4): 464. DOI 10.3390/f12040464.
13. Constantz B.M., Port A.A., Senock R.S. Comparing automatically generated and manually measured tree-ring transects of growth trends with Hawaiian sandalwood as an example species \// Dendrochronologia. - 2021. -Vol.68:125831. DOI 10.1016/j.dendro.2021.125831.
14. Rose T., Ali N., Dong Y. Design and development of an IoT-based dendrometer system for real-time trunk diameter monitoring of Christmas trees // Smart Agricultural Technology. - 2025. -Vol. 10:100765. DOI 10.1016/j.atech.2024.100765.
15. Clonch C., Huynh M., Goto B. Et al. High precision zero-friction magnetic dendrometer // HardwareX. - 2021.- Vol. 10: e00248. DOI 10.1016/j.ohx.2021.e00248.
16. Miller T.W., Stangler D.F., Larysch E. et al. A methodological framework to optimize models predicting critical dates of xylem phenology based on dendrometer data \// Dendrochronologia. - 2022. Vol. Art. 125940. DOI 10.1016/j.dendro.2022.125940.

17. Aryal S., Grießinger J., Dyola N. et al. INTRAGRO: A machine learning approach to predict future growth of trees under climate change // Ecology and Evolution. 2023.- Vol. 13(10): e10626. DOI 10.1002/ece3.10626.
18. Yrttimaa T., Junttila S., Luoma V. Et al. Capturing seasonal radial growth of boreal trees with terrestrial laser scanning // Forest Ecology and Management. - 2023.- Vol. 529:120733. DOI 10.1016/j.foreco.2022.120733.
19. Tumajer J., Braun S., Burger A. et al. Dendrometers challenge the 'moon wood concept' by elucidating the absence of lunar cycles in tree stem radius oscillation \// Scientific Reports. - 2023.- Vol.13(1):19904. DOI 10.1038/s41598-023-47013-y.
20. Wu F., Huang Y., Benes B., Warner C.C., Gazo R. Automated tree ring detection of common Indiana hardwood species through deep learning: Introducing a new dataset of annotated images \// Information Processing in Agriculture. −2024. -Vol. 11(4). - P. 552-558. DOI 10.1016/j.inpa.2023.10.002.
21. Zambrano-Suárez J.D., Pérez-Martín J., Muñoz-Torrero Manchado A., Ballesteros Cánovas J.A. Tree ring segmentation performance in highly disturbed trees using deep learning // PLOS One. −2026.- Vol.21(6): e0321841. DOI 10.1371/journal.pone.0321841.
22. Marichal H., Passarella D., Lucas C. et al. UruDendro, a public dataset of 64 cross-section images and manual annual ring delineations of Pinus taeda L. // Annals of Forest Science. - 2025.- Vol. 82:25. DOI 10.1186/s13595-025-01296-5.
23. Klesse S., Babst F., Evans M.E.K., Hurley A., Pappas C., Peters R.L. Legacy effects in radial tree growth is rarely significant after accounting for biological memory \// Journal of Ecology. –2023.- Vol. 111(6). - P. 1188-1202. DOI 10.1111/1365-2745.14045.
24. Heilman K.A., Dietze M.C., Arizpe A.A. et al. Ecological forecasting of tree growth: Regional fusion of tree-ring and forest inventory data to quantify drivers and characterize uncertainty // Global Change Biology. - 2022. -Vol. 28(7). - P. 2442–2460. DOI 10.1111/gcb.16038.
25. Evans M.E.K., Falk D.A., Arizpe A. et al. Fusing tree-ring and forest inventory data to infer influences on tree growth // Ecosphere. 2017. Vol. 8(7). Art. e01889. DOI 10.1002/ecs2.1889.
26. Cerqueira V., Torgo L., Mozetič I. Evaluating time series forecasting models: an empirical study on performance estimation methods // Machine Learning. - 2020.- Vol. 109(11). - P. 1997- 2028. DOI 10.1007/s10994-020-05910-7.
27. Hewamalage H., Ackermann K., Bergmeir C. Forecast evaluation for data scientists: common pitfalls and best practices // Data Mining and Knowledge Discovery. 2023.- Vol. 37.- P. 788-832. DOI 10.1007/s10618-022-00894-5.
28. Stock A., Gregr E.J., Chan K.M.A. Data leakage jeopardizes ecological applications of machine learning // Nature Ecology & Evolution. −2023.- Vol. 7(11). - P.1743–1745. DOI 10.1038/s41559-023-02162-1.
29. Lewis A.S.L., Woelmer W.M., Wander H.L. et al. Increased adoption of best practices in ecological forecasting enables comparisons of forecastability // Ecological Applications. - 2022.-Vol.- 32(2): e02500. DOI 10.1002/eap.2500.
30. Chernozhukov V., Wüthrich K., Zhu Y. Distributional conformal prediction // Proceedings of the National Academy of Sciences. - 2021. -Vol. 118(48): e2107794118. DOI 10.1073/pnas.2107794118.
31. Teslia N.E. NEL1: Claim-Gated Tree-Ring Forecasting under Source Shift and Operational Constraints: From Leakage Audit to Fail-Closed Operational Closure. Version 1.0 [Electronic resource]. Zenodo.-2026.- 72 p. DOI 10.5281/zenodo.21812690.
32. Wadoux A.M.J.-C., Heuvelink G.B.M., de Bruin S., Brus D.J. Spatial cross-validation is not the right way to evaluate map accuracy // Ecological Modelling. 2021.- Vol. 457:109692. DOI 10.1016/j.ecolmodel.2021.109692.
33. Rau E.-P., Fischer F., Joetzjer E. Et al. Transferability of an individual- and trait-based forest dynamics model: A test case across the tropics // Ecological Modelling. - 2022.- Vol. 463:109801. DOI 10.1016/j.ecolmodel.2021.109801.
34. Hendrickx K., Perini L., Van der Plas D., Meert W., Davis J. Machine learning with a reject option: a survey // Machine Learning. −2024. -Vol. −113(5). - P. 3073-3110. DOI 10.1007/s10994024-06534-x.

Информация об авторах

Акишев К.М.- кандидат технических наук, ассоциированный профессор, Казахский университет технологии и бизнеса им. К. Кулажанова, Астана, Казахстан; e-mail: q.akishev@kaztbu.edu.kz;

Тлебаев М.Б.- доктор технических наук, профессор, профессор, Таразский университет им. М.Х. Дулати, Тараз, Казахстан; e-mail: mb.tlebayev@dulaty.kz;

Ешмуратов А.- магистрант, Казахский университет технологии и бизнеса им. К. Кулажанова, Астана, Казахстан, e-mail: assylzhanyeshmuratov@gmail.com.

Information about the authors

Akishev K. M.- candidate of Technical Sciences, Associate Professor, K. Kulazhanov Kazakh University of Technology and Business, Astana, Kazakhstan, e-mail: q.akishev@kaztbu.edu.kz;

Tlebayev M.B.- doctor of Technical Sciences, Professor, Professor, M.Kh. Dulaty Taraz University, Taraz, Kazakhstan, e-mail: mb.tlebayev@dulaty.kz;

Yeshmuratov A.A.- master's student, K. Kulazhanov Kazakh University of Technology and Business, Astana, Kazakhstan, e-mail: assylzhanyeshmuratov@gmail.com.

Ответственный редактор: Оспанова М. К.
Верстка: Ундасынов Р. Е.

Подписано в печать: 29.09.2026

Издание: АО «Казахский университет технологии и бизнеса им. К. Кулажанова», г. Астана,
ул. Кайым Мухамедханова, здание 37 А
Рабочий телефон: +7(7172)72-58-12 (134)
E-mail: itamns@kaztbu.edu.kz