Knowledge and Information Systems (2026) 68:227
https://doi.org/10.1007/s10115-026-02845-8

RESEARCH

Adaptive approaches towards fully incremental prediction interval for data stream regression

Yibin Sun[1] · Bernhard Pfahringer[2] · Heitor Murilo Gomes[1] · Albert Bifet[2,3]

Received: 19 March 2026 / Revised: 14 May 2026 / Accepted: 21 June 2026

Abstract

Prediction intervals (PIs) are a practical tool for uncertainty quantification in regression, but comparatively little work has addressed fully incremental PI generation for data streams. In streaming settings, data arrive continuously, each instance is typically processed once, and concept drift can quickly invalidate a previously well-calibrated interval. These properties make many batch PI methods and window-based adaptations difficult to apply efficiently. This paper studies Adaptive Prediction Interval (AdaPI), an online post-calibration framework that adjusts interval width according to observed coverage. We instantiate the framework with a fully incremental variant of Mean and Variance Estimation (MVE) and investigate three adaptive scaling functions. We also adopt an evaluation perspective that jointly considers coverage accuracy and interval width. Experiments on a collection of real-world and synthetic regression streams show that AdaPI can often move coverage closer to the desired confidence level while maintaining competitive interval width; under the default 95% confidence setting and coverage-heavy CING weighting, the linear variant frequently gives the strongest empirical coverage–width trade-off among the three adaptive strategies.

Keywords Data streams · regression · prediction intervals · concept drifts

Bernhard Pfahringer, Heitor Murilo Gomes and Albert Bifet have contributed equally to this work.

✉ Yibin Sun
spencer.sun@vuw.ac.nz

Bernhard Pfahringer
bernhard@waikato.ac.nz

Heitor Murilo Gomes
heitor.gomes@vuw.ac.nz

Albert Bifet
abifet@waikato.ac.nz

1 School of Engineering and Computer Science, Victoria University of Wellington, Wellington, New Zealand

2 AI Institute, The University of Waikato, Hamilton, New Zealand

3 LTCI, Télécom Paris, IP Paris, France

Published online: 24 July 2026

1 Introduction

Regression models are widely used in settings where users need more than a single predicted value. In such cases, prediction intervals (PIs) provide a practical form of uncertainty quantification by returning lower and upper bounds that are expected to contain the true target value with a specified confidence level [1, 2]. For many decision-making tasks, these intervals are more informative than point predictions alone [3, 4].

Constructing reliable PIs becomes substantially harder in data streams. In streaming regression, data arrive continuously, potentially without bound, and each instance is typically processed once [5]. Many established PI approaches, including bootstrap- and residual-based methods [6], were designed for batch settings and are awkward to deploy under these constraints. In practice, streaming adaptations often depend on sliding windows or periodic retraining [7], which introduces additional design choices and may reduce responsiveness to changing data.

The problem is further complicated by *concept drift*, where the data-generating distribution evolves over time [5]. A PI method that is well calibrated at one stage of a stream may later become too narrow, causing under-coverage, or too wide, reducing informativeness. This makes online calibration a central requirement for practical prediction intervals in streaming settings.

In our previous work [8], we introduced Adaptive Prediction Interval (AdaPI), a lightweight post-calibration strategy that rescales interval width according to observed coverage. The present paper extends that idea in three ways. First, we present AdaPI as a broader adaptive framework for streaming PI calibration. Second, we instantiate the framework with a fully incremental variant of Mean and Variance Estimation (MVE) and study three alternative scaling functions. Third, we evaluate these variants on a diverse collection of real-world and synthetic streams, including comparisons with recent online conformal methods.

The aim of this paper is not to claim that a single adaptive rule dominates all uncertainty estimation approaches. Rather, we study whether a simple, model agnostic, fully incremental post-processing mechanism can improve practical PI calibration under evolving stream conditions. In particular, we ask whether AdaPI can keep coverage closer to the desired confidence level while maintaining reasonably narrow intervals.

The main contributions of this paper are as follows:

- We present AdaPI as a general online post-calibration framework for streaming prediction intervals, designed to adapt interval width incrementally according to observed coverage.
- We develop a fully incremental MVE-based instantiation of this framework and study three alternative adaptive scaling rules: logarithmic, HLI, and linear.
- We employ an evaluation perspective that jointly considers coverage accuracy and interval width, allowing the practical trade-off between calibration and informativeness to be analysed more clearly.
- We conduct an empirical study on multiple real-world and synthetic data streams using several regression backbones and compare the proposed methods with both MVE and recent online conformal alternatives.

2 Problem setting

Consider a linear regression problem:

$$y = \sum_{k=1}^{K} \beta_k x_k + \epsilon \tag{1}$$

where K is the number of independent variables, x_k the input independent variables, y the predicted dependent variable, β_k the coefficients to be determined, ϵ the noise term. If we collect N samples, the problem can be stated in a compact matrix form such as:

$$\mathcal{Y} = \mathcal{X}\beta + \mathcal{E} \tag{2}$$

with $\mathcal{Y}, \mathcal{E}, \beta \in \mathbb{R}^N$, and $\mathcal{X} \in \mathbb{R}^{N \times K}$.

In this paper, the PI task is treated as an extension of regression. Instead of returning only a point prediction $\mathcal{Y}$, the learner outputs an interval $[\mathcal{P}_l, \mathcal{P}_u]$ that should contain the true target with probability α, where α is the desired confidence level and is denoted by $\mathcal{L}$ throughout the remainder of the paper. In other words, we seek intervals satisfying $Pr(\mathcal{P}_l \leq \mathcal{Y} \leq \mathcal{P}_u) = \alpha$.

In practice, this target is difficult to achieve exactly. Many PI methods rely on statistical assumptions that are frequently violated in real data, especially assumptions on the error distribution [9]. The challenge is even greater in data streams, where the data distribution can evolve over time and concept drift can invalidate a previously well-calibrated interval [5].

Our objective is therefore to study streaming PI methods that keep empirical coverage close to the desired confidence level while remaining as narrow as possible.

3 Related work

Prediction intervals (PIs) have been widely studied in applied domains such as renewable energy forecasting [10], wind speed analysis [11], stock market prediction [12], and tidal current forecasting [13]. Another popular area of application is autoregressive and time series analysis [14–16]. These applications illustrate the importance of PIs as a means of quantifying predictive uncertainty.

3.1 Classical PI generation methods

Xu and Xie [17] proposed a bootstrap-based methodology for dynamic time series. Their approach establishes a set of base learners $\mathcal{A}$, defines an error-based aggregation function ϵ, and resamples the time series into batches $\mathcal{B}$ for training. The resulting sequences of predictions $\mathcal{A}(x_i)$ and errors $\epsilon(x_i)$ are then combined, and the $1-\alpha$ quantile yields the final interval $[\mathcal{A}^{1-\alpha} - \epsilon^{1-\alpha}, \mathcal{A}^{1-\alpha} + \epsilon^{1-\alpha}]$.

Liu et al. [18] proposed the Interval Type-2 Fuzzy Granular Neural Network Dynamic Ensemble (IT2FGNNDEnsemble), which integrates granular neural networks with bootstrap techniques. Its seven-layer architecture features automatic generation, pruning, and weighting of learners through a "Parsimonious Ensemble" mechanism. Each learner produces a PI, and the ensemble fuses them into a final interval. While effective in capturing data tendencies, this multi-layered structure is computationally demanding, making it challenging in streaming contexts.

Another widely studied approach is Jackknife+ [19], derived from Jackknife resampling [20]. It trains models $\mathcal{A}_{-i}$ using leave-one-out residuals R^{LOO} and defines intervals as

$$[q^{-}_{n,\alpha}(\mathcal{A}_{-i} - R_i^{LOO}),\ q^{+}_{n,\alpha}(\mathcal{A}_{-i} + R_i^{LOO})],$$

where q denotes α-quantiles. Despite attractive guarantees, the method requires repeated access to the dataset and is therefore difficult to reconcile with the one-pass assumption of stream learning.

3.2 Neural and hybrid approaches

Direct interval prediction methods constitute another important family. Hadjicharalambous et al. [7] combined Bootstrap [21] and Lower-Upper Bound Estimation (LUBE) [22] into the Bootstrap-LUBE Method (BLM). This design attempts to stabilize bound prediction by generating pseudo-instances, but it also inherits the computational burden of repeated neural network training. Even when adapted to sliding or tumbling windows, such methods remain only partially incremental.

Zhao et al. [23] proposed a tree-based PI method for time series and streaming data using Conditional Inference Trees (ctree) [24] together with Haar scaling functions [25]. Their method is more efficient than heavy neural approaches, but it still updates periodically in batches rather than in a strict instance-by-instance fashion.

3.3 Conformal prediction

Conformal Prediction (CP) [26] provides a statistically principled framework for interval construction. In regression, a common workflow is to train a base model, compute nonconformity scores on a calibration set, and then obtain interval widths from quantiles of those scores. CP has been applied successfully across many domains [27–29]. However, its practical performance depends strongly on the choice of base model and nonconformity measure, and the standard formulation is not tailored to non-stationary one-pass streams.

3.4 Recent online CP methods

Recent work has extended conformal ideas to sequential and adaptive settings. Conformal PID control [30] adjusts interval widths through feedback control. MVP [31] updates thresholds online to maintain worst-case coverage across groups. SAOCP [32] aggregates experts with different lifetimes to improve adaptation under distribution shift. ACI [33] and AgACI [34] recursively adapt miscoverage levels in order to improve long-run calibration under nonstationarity.

These methods are highly relevant to our setting and form a meaningful comparison point in this paper. At the same time, they reflect a different design tradeoff from AdaPI: They focus on adaptive sequential calibration procedures, whereas our goal is a lightweight, model agnostic, fully incremental post-calibration mechanism that can be attached to a streaming regressor.

3.5 Summary

Overall, prior work shows that interval calibration for evolving data is feasible, but the design space remains fragmented. Classical PI methods are mostly batch-based, neural, and hybrid methods are often computationally expensive, and online conformal methods address a related but not identical objective. This leaves room for simple adaptive post-processing rules that are inexpensive, incremental, and easy to combine with existing stream regressors.

4 Background

This section introduces the baseline PI model and the evaluation criteria used throughout the paper.

4.1 Mean and variance estimation (MVE)

Mean and Variance Estimation (MVE) [35] is one of the simplest and most widely used PI constructions. The method assumes that predictive errors follow a Gaussian distribution, which leads to the interval form shown in Eq. 3:

$$\mathrm{PI} \in \left(\mathcal{Y} - G^{-1}(0, \gamma) \times \sigma_\epsilon, \mathcal{Y} + G^{-1}(0, \gamma) \times \sigma_\epsilon\right) \tag{3}$$

where $G^{-1}(0, \gamma)$ is the inverse Gaussian distribution function with zero mean and probability γ and $\mathcal{Y}$ is the point prediction. According to the normality assumption, $\mathcal{E} \sim \mathcal{N}(0, \sigma_\epsilon)$. Hence, when $\gamma = 95\%$, $G^{-1}(0, 0.95) \approx 1.96$.

4.2 Evaluation metrics

Evaluating a PI requires balancing two competing goals: matching the target coverage and keeping the interval narrow. Following prior work [7], we therefore report both Coverage and Normalized Mean Prediction Interval Width (NMPIW). Coverage is defined as the proportion of ground truth targets that fall inside the predicted interval:

$$\text{Coverage} = \frac{1}{N}\sum_{i=1}^{N} I_i \tag{4}$$

where $I_i = 1$ if $y \in [\mathcal{P}_l, \mathcal{P}_u]$ and $I_i = 0$ otherwise; N is the total observation number. NMPIW is expressed in Eq. 5:

$$\text{NMPIW} = \frac{\frac{1}{N}\sum_{i=1}^{N}(\mathcal{P}_{u_i} - \mathcal{P}_{l_i})}{R} \tag{5}$$

where $\mathcal{P}_{u_i}$ and $\mathcal{P}_{l_i}$ denote the upper and lower bounds for the i_{th} observation and R is the range of observed target values. NMPIW therefore measures average interval width relative to the scale of the target variable. Throughout the remainder of the paper, coverage and confidence level are denoted by $\mathcal{C}$ and $\mathcal{L}$, respectively.

These two metrics express the main tradeoff in PI design. Wider intervals tend to improve coverage but reduce informativeness, whereas narrower intervals are useful only if they still cover the target at roughly the desired rate.

4.3 Multi-objective evaluation

Because PI quality depends on both coverage and width, comparison is naturally a multi-objective problem. The multi-objective optimization literature offers a large number of ranking and selection strategies [36, 37]. Here we borrow ideas from Non-dominated Sorting Genetic Algorithm II (NSGA II) and adapt them into an evaluation procedure for PI tasks.

4.3.1 Non-dominated sorting genetic algorithm II

First introduced by Deb et al. [38], NSGA II is a widely used method for identifying Pareto-optimal solutions in multi-objective problems [39]. In simplified form, it consists of two components: 1. non-dominated sorting; and 2. crowding distance. For the non-dominated sorting stage, we first define Pareto dominance. In the objective space, where f indexes the objective dimensions, k is the number of dimensions, and $\mathcal{O}$ denotes an observation, we say that an observation $\mathcal{O}^d$ dominates another observation $\mathcal{O}^t$ when

$$\forall i \in \{1, ..., k\} : \mathcal{O}^d_{f_i} \leq \mathcal{O}^t_{f_i} \tag{6}$$

and

$$\exists j \in \{1, ..., k\} : \mathcal{O}^d_{f_j} < \mathcal{O}^t_{f_j} \tag{7}$$

This definition is stated for minimization, where smaller values are preferred. For maximization, the inequalities would be reversed.

Observations are then separated into dominated and non-dominated groups.[1] Members of the dominated group are clearly inferior to at least one point in the non-dominated set. If multiple observations remain in the non-dominated group, NSGA II further ranks them with crowding distance.

Crowding distance measures how isolated a point is from its neighbours. For an observation $\mathcal{O}$, the contribution with respect to the i_{th} objective is defined in Eq. 8:

$$c(\mathcal{O}_{f_i}) = \mathcal{O}^R_{f_i} - \mathcal{O}^L_{f_i} \tag{8}$$

where $\mathcal{O}^R$ and $\mathcal{O}^L$ are the right and left neighbours of $\mathcal{O}$ in the sorted objective space. The total crowding distance is then:

$$C(\mathcal{O}) = \sum_{i=1}^{k} c(\mathcal{O}_{f_i}) \tag{9}$$

In words, the crowding distance is the sum of the local gaps around an observation across all objectives. In a two-dimensional setting, it is closely related to the Manhattan distance between neighbouring points.

4.3.2 Coverage interval width in non-dominated group evaluation

NSGA II is useful conceptually, but its second stage favours sparsity in order to preserve diversity along the Pareto front. That behaviour is valuable in optimization, but it is not essential for ranking PI results. We therefore define a simpler evaluation procedure for prediction intervals, Coverage Interval Width in Non-dominated Group (CING), which works as follows:

[1] General multi-objective problems may involve multiple dominance layers. In our two-objective evaluation setting, we use only the simplified dominated/non-dominated split.

1. Apply a linear Max–Min normalization step to the coverage error and NMPIW across all observations;
2. Determine the non-dominated group with the same technique as in NSGA II; and
3. Calculate scores using a weighted sum strategy within the non-dominated group.

The first step applies max–min normalization, as shown in Eq. 10:

$$x_i = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}} \qquad i \in 1, ..., k \tag{10}$$

where X_i is the raw value, $X_{\max}$ and $X_{\min}$ are the maximum and minimum values in the series, and x_i is the normalized value. We then define coverage error (CE) as the absolute difference between coverage and the target confidence level:

$$\mathrm{CE} = Abs(\mathcal{C} - \mathcal{L}) \tag{11}$$

where $\mathcal{C}$ is the observed coverage and $\mathcal{L}$ is the desired confidence level.

We use a weighted sum rather than Euclidean distance to the ideal point.[2] Euclidean distance would implicitly weight coverage error and interval width equally. In PI evaluation, however, that balance is often application dependent, and in many scenarios matching the target coverage is more important than further shrinking an already narrow interval. A weighted sum therefore provides a more transparent and adjustable ranking rule.

Under this evaluation, both CE and NMPIW are minimized, so the lower-left corner of the normalized objective space is preferred. After identifying the non-dominated group, CING assigns scores within that group according to the chosen weights. The best result is then the one with the lowest weighted score.

5 Adaptive prediction intervals

MVE is attractive because of its simplicity, but in streaming applications its Gaussian error assumption is often violated. When that happens, the resulting intervals may systematically miss the target confidence level. If the base regressor becomes unusually accurate, MVE may produce intervals that are too narrow; if predictive quality deteriorates, the intervals may become too wide or still fail to maintain coverage.

AdaPI addresses this problem by introducing a multiplicative factor into the MVE construction. Thus, we introduce a factor $\mathcal{F}$[3] to the MVE approach, as formulated in Eq. 12:

$$\mathrm{PI} = \big(\mathcal{Y} - \mathcal{F} \times G^{-1}(0, \gamma) \times \sigma_\epsilon, \mathcal{Y} + \mathcal{F} \times G^{-1}(0, \gamma) \times \sigma_\epsilon\big) \tag{12}$$

where $\mathcal{F}$ is determined according to the real-time coverage of the model.

For the adaptive approaches to work properly, the factor $\mathcal{F}$ must exhibit several key characteristics:

1. To ensure that the adaptive mechanism does not disrupt interval generation when the model performs as desired, $\mathcal{F}$ should equal one when the coverage equals the confidence level, i.e. $\mathcal{F} = 1 \quad \text{if} \quad \mathcal{C} = \mathcal{L}$;

[2] In our minimization setting, the ideal point is (0, 0), corresponding to zero coverage error and zero interval width.

[3] In our previous work [8], the corresponding quantity was denoted by $\mathcal{S}$; here we use $\mathcal{F}$ to emphasize its role as a multiplicative factor.

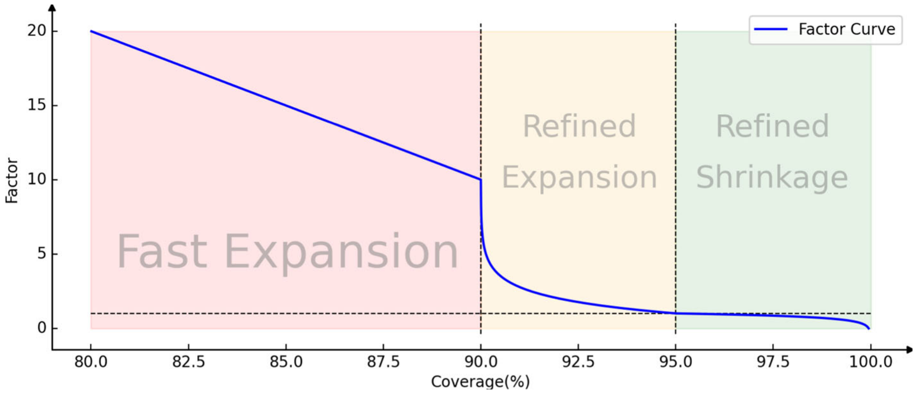


Fig. 1 Logarithm factor with a 95% confidence level ($\mathcal{C} \in [80, 100]$) and a 0.01 lower limit. Dotted lines denote significant values in the figure. The horizontal line is 1.0; the right vertical line is the desired confidence level (95%). The refined expansion and refined shrinkage regions have the same width

2. To facilitate the coverage converging to the confidence level, the interval should be expanded when coverage is insufficient, thus allowing the modified interval to include more ground truth values. Conversely, when the coverage is excessively high, the interval should be contracted to maintain an acceptable uncertainty level. Therefore, $\mathcal{F} > 1 \quad \text{if} \quad \mathcal{C} < \mathcal{L}$, and $\mathcal{F} < 1 \quad \text{if} \quad \mathcal{L} < \mathcal{C} < 100\%$;
3. A too narrow prediction interval is also not informative as it might frequently miss the ground truth. Therefore, a reasonably conservative lower limit for $\mathcal{F}$ can be beneficial, i.e. $\mathcal{F} = \text{limit} \quad \text{if} \quad \mathcal{C} = 100\%$.

Based on these characteristics, we outline the following generating curves for our adaptive approaches. Note that coverage is expressed as a percentage and therefore ranges over [0, 100].

5.1 Logarithm approach

The first approach defines $\mathcal{F}_{\text{Log}}$ as in Eq. 13:

$$\mathcal{F}_{\text{Log}} = \begin{cases} 100 - \mathcal{C} & \text{if } \mathcal{C} < 2\mathcal{L} - 100 \\ (\mathcal{L} - 100) \log_{\mathcal{L}} \frac{100 + \mathcal{C} - 2\mathcal{L}}{100 - \mathcal{L}} + 1 & \text{if } 2\mathcal{L} - 100 \leq \mathcal{C} < \mathcal{L} \\ \log_{\mathcal{L}} \frac{100 - \mathcal{C}}{100 - \mathcal{L}} + 1 & \text{if } \mathcal{L} \leq \mathcal{C} \end{cases} \tag{13}$$

where $\mathcal{C}$ is the current coverage and $\mathcal{L}$ is the target confidence level. The exact shape depends on the chosen confidence level. Figure 1 illustrates the curve for $\mathcal{L} = 95$.

Figure 1 reveals three operating regions. When coverage is far below the target, the curve enters a fast expansion regime that increases interval width aggressively. Closer to the target, the refined expansion regime still enlarges the interval, but does so more gradually to avoid oscillation. When coverage exceeds the target, the refined shrinkage regime contracts the interval while keeping the transition smooth near $\mathcal{L}$.

This design has three practical advantages: 1. Radical adjustments occur for substantial coverage confidence disparities; 2. Modifications are moderate when coverage approaches

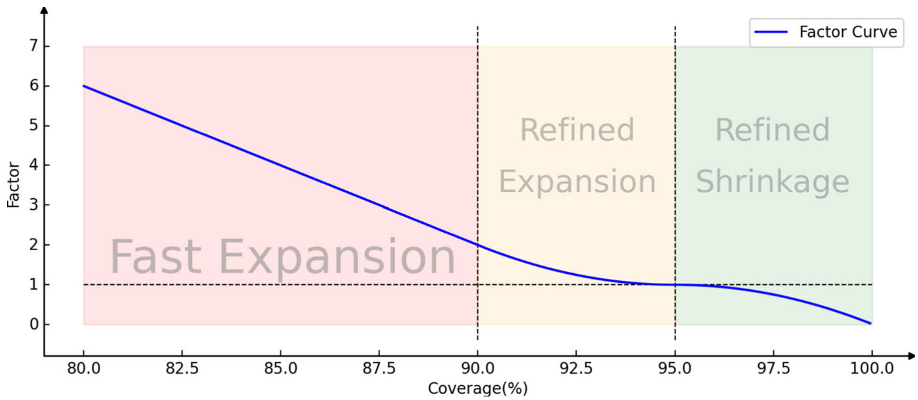


Fig. 2 HLI factor with a 95% confidence level ($\mathcal{C} \in [80, 100]$) and a 0.01 lower limit. Dotted lines denote significant values in the figure. The horizontal line is 1.0; the right vertical line is the desired confidence level (95%). The refined expansion and refined shrinkage regions have the same width

the confidence level, preventing overreactions; and 3. Auto cessation at the desired confidence level is achieved without adding extra hyperparameters.

5.2 Huber loss inspired (HLI)

The logarithmic curve contains a visible inflection point. To obtain a smoother transition, we derive a second scaling function from the Huber loss [40], originally introduced as a robust alternative to quadratic loss:

$$L_\delta = \begin{cases} \frac{1}{2}e^2 & \text{if } |e| < \delta \\ \delta(e - \frac{1}{2}\delta) & \text{otherwise} \end{cases} \tag{14}$$

where e is the predictive error and δ controls the width of the quadratic region. The appeal of the Huber function in our setting is its smooth transition between linear and quadratic behaviour.

For AdaPI, we modify this curve so that it satisfies the requirements stated earlier. Specifically, we align the symmetry line with the confidence level, discard the right linear branch, and rescale the remaining shape so that the right boundary matches the lower limit constraint at 100% coverage. The resulting HLI factor is defined in Eq. 15:

$$\mathcal{F}_{\text{HLI}} = \begin{cases} \frac{(\mathcal{M}-1)(\mathcal{C}-\mathcal{L})^2}{(100-\mathcal{L})^2} + 1 & \text{if } \mathcal{C} \geq \mathcal{L} \\ 2(\frac{\mathcal{C}-2\mathcal{L}+100}{\mathcal{L}-100} + 1) & \text{if } \mathcal{C} < 2\mathcal{L} - 100 \\ (\frac{\mathcal{C}-\mathcal{L}}{100-\mathcal{L}})^2 + 1 & \text{otherwise} \end{cases} \tag{15}$$

where $\mathcal{M}$ is the lower limit and the remaining notation is the same as in Eq. 13.

Compared with the logarithmic variant, HLI provides a smoother transition across the expansion and shrinkage regions, as shown in Fig. 2.

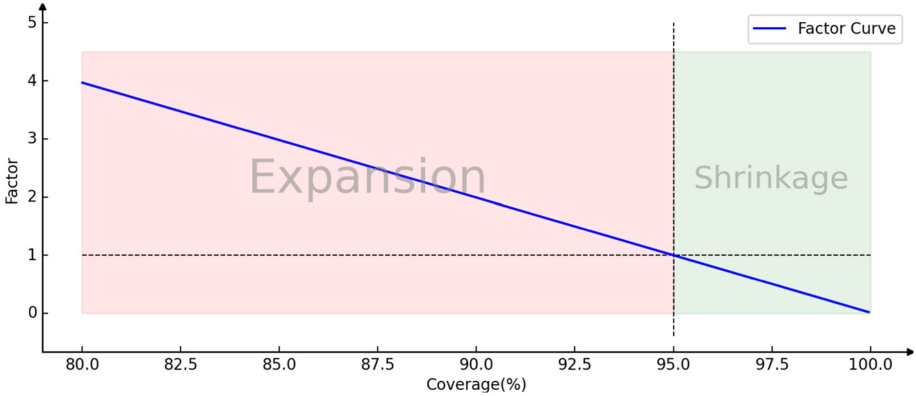


Fig. 3 Linear lever factor with a 95% confidence level ($\mathcal{C} \in [80, 100]$) and a 0.01 lower limit. Dotted lines denote significant values in the figure. The horizontal line is 1.0, and the vertical line is the desired confidence level (95%)

5.3 Linear lever

Our third variant is deliberately simple: a linear scaling rule. To satisfy the AdaPI requirements, the line must pass through two points: $f(\mathcal{L}) = 1$ and $f(100) = \mathcal{M}$. This yields the factor in Eq. 16:

$$\mathcal{F}_{\mathrm{L}} = \frac{1-\mathcal{M}}{\mathcal{L}-100}\,\mathcal{C} + \frac{\mathcal{M}\mathcal{L}-100}{\mathcal{L}-100} \tag{16}$$

where $\mathcal{F}_{\mathrm{L}}$ is the linear lever factor and the remaining notation is shared with Eqs. 13 and 15.

The linear variant has only two regimes. Coverage below the confidence level ($\mathcal{C} < \mathcal{L}$) expands the interval; coverage above it ($\mathcal{C} \geq \mathcal{L}$) contracts the interval. Its simplicity makes it easy to interpret and, as the experiments later show, often surprisingly competitive. This competitiveness should be understood as a property of the empirical tradeoff rather than as a universal dominance claim: The direct correction can recover target coverage quickly when MVE is biased, but it can also be more sensitive at very low or very high confidence levels (Fig. 3).

5.4 Convergence sketch under i.i.d. assumptions

We provide an informal convergence argument to clarify the behaviour of AdaPI under stationary conditions. The analysis is deliberately simplified: we work under i.i.d. data and treat the adaptive factor as *approximately* fixed over short windows. A fully rigorous treatment of the coupled process $(F_t, \hat{C}_t)$ is left for future work.

Setup

Let $(X_t, Y_t)_{t\geq 1}$ be an i.i.d. stream and let the base prediction interval before adaptation be

$$I_t(F) = \big[\hat{y}_t - F\, z_\ell\, \hat{\sigma}_t,\ \ \hat{y}_t + F\, z_\ell\, \hat{\sigma}_t\big],$$

where $z_\ell = G^{-1}(0, \ell)$, $F > 0$ is the adaptive factor, $\ell \in (0, 1)$ is the target coverage probability, and $\mathcal{L} = 100\ell$ is the corresponding percentage used by the scaling rules. Define the *population coverage* induced by a fixed factor F as

$$p(F) = \Pr\{Y_t \in I_t(F)\}.$$

Assumptions
We impose three conditions.

1. *(Monotonicity)* $p(F)$ is continuous and strictly increasing in a neighbourhood of $F^\star$. This states that widening the interval locally increases coverage, which holds whenever the error distribution has positive density around the interval boundary.
2. *(Attainability)* There exists a unique $F^\star > 0$ such that $p(F^\star) = \ell$. This requires the target coverage to be achievable by multiplicative rescaling of the base MVE interval; it may fail for very heavy-tailed error distributions.
3. *(Stable base learner)* The residual estimates $\hat{\sigma}_t$ are stable in the sense that $p(F)$ does not vary with t under the i.i.d. assumption.

Step 1: consistency of the coverage estimate
At time t, AdaPI estimates the coverage probability by the empirical running average

$$\hat{c}_t(F) = \frac{1}{t}\sum_{i=1}^{t} \mathbf{1}\{Y_i \in I_i(F)\}.$$

For any *fixed* F, the indicators are i.i.d. Bernoulli$(p(F))$, so the strong law of large numbers gives

$$\hat{c}_t(F) \xrightarrow{\text{a.s.}} p(F).$$

We note an important caveat: in practice F changes at every step, so $\hat{c}_t$ mixes coverage information from all past factor values, not only the current one. The formal analysis of the coupled process $(F_t, \hat{C}_t)$ requires tools beyond the SLLN and is outside the scope of this paper. The result above is therefore used as motivation: once F_t is near $F^\star$, the running average is a consistent estimate of $p(F^\star) = \ell$.
Step 2: negative-feedback structure
AdaPI updates the factor through a continuous scaling rule

$$F_{t+1} = g\left(\hat{C}_t\right),$$

where g satisfies the three design conditions introduced in Sect. 5:

$$g(\mathcal{L}) = 1, \qquad g(C) > 1 \ \text{ if } \ C < \mathcal{L}, \qquad g(C) < 1 \ \text{ if } \ C > \mathcal{L}.$$

Here $\hat{C}_t = 100\hat{c}_t$ is the percentage form of empirical coverage. Define the coverage error $e_t = \hat{C}_t - \mathcal{L}$. When $e_t < 0$, the rule selects $F_{t+1} > 1$, widening the interval and increasing future coverage by monotonicity; when $e_t > 0$, it selects $F_{t+1} < 1$, narrowing the interval and reducing future coverage. The *sign* of the expected update therefore opposes the sign of the error:

$$\mathbb{E}[e_{t+1} - e_t \mid e_t] < 0 \quad \text{when } e_t > 0, \qquad \mathbb{E}[e_{t+1} - e_t \mid e_t] > 0 \quad \text{when } e_t < 0.$$

This establishes that $\mathcal{L}$ is a *sign-stable* fixed point of the adaptive process. We emphasize that sign-stability does not by itself exclude persistent oscillation; bounding the magnitude of updates (e.g. via a Lyapunov argument on e_t^2) would be required for a full convergence proof.
Step 3: properties of the linear rule For the linear lever (Eq. 16), g is affine in C, so its derivative g' is constant everywhere. Two consequences follow.

- *No local over-reaction* Because $|g'|$ is the same near $\mathcal{L}$ as far from it, the correction does not amplify small errors. Curved rules (logarithmic, HLI) have larger $|g'|$ far from $\mathcal{L}$

and smaller $|g'|$ near $\mathcal{L}$; this trades faster recovery from large deviations for a gentler approach to the target.

- *Bounded step sizes* For any $C \in [0, 100]$, the linear rule produces a factor increment $|F_{t+1} - 1|$ bounded by $|C - \mathcal{L}| \cdot \frac{1-\mathcal{M}}{100-\mathcal{L}}$, where $\mathcal{M}$ is the lower limit. This makes the step size explicitly proportional to the coverage error, a standard sufficient condition for stability in linear feedback systems.

These properties explain why the linear rule is empirically competitive despite its simplicity, and motivate further theoretical study of convergence rates.

Summary

Under the three assumptions above and for fixed F, the coverage estimate is consistent and the update rule forms a negative-feedback loop around $\mathcal{L}$. Taken together, these properties support the expectation that AdaPI will stabilize near

$$\lim_{t\to\infty} \hat{C}_t \approx \mathcal{L},$$

provided that the factor dynamics are well behaved and that the target coverage is attainable by multiplicative rescaling. Because the i.i.d. assumption does not hold in streaming settings with concept drift, this argument is best understood as a characterization of the *stationary regime*. In non-stationary settings, AdaPI instead responds reactively: drift reduces coverage, the factor increases, and coverage is partially restored—a behaviour confirmed empirically in Sect. 7.3.

6 Experiments

This section summarizes the experimental design used to evaluate AdaPI and the comparison methods.

6.1 Datasets

We evaluate the proposed methods on a mixture of real-world and synthetic regression streams. Table 1 summarizes the datasets, including their size, source, and known drift characteristics.

- Abalone [41]: aims at predicting the age of abalones based on the physical measurements;
- CPU Activity*[4]: predicts the portion of time that CPUs run in user mode;
- Naval Propulsion Plant [42]: predicts the GT Turbine decay state coefficient of naval propulsion plant based on other features;
- Ailerons [43]: artificial data address a control problem for flying a F16 aircraft;
- Miami Housing*: predicts sale price of single family house in Miami;
- Elevators [43]: also predicts a control problem for flying a F16 aircraft;
- Bike [44]: predicts the number of shared-bike rentals, including both casual and registered users;
- California Housing [45]: predicts the median house price in California, US, in 1990s;
- Superconductivity [46]: predicts the critical temperature of superconductors based on other features;
- Health Insurance*: predicts the working hours of married women based on their insurance information;

[4] Datasets with * can be found at OpenML.

Table 1 Datasets overview drift types—A: Abrupt; G: Gradual; N: No drift; ?: Unknown

Dataset	#Feature	#Instance	Source	Drift type
Abalone	8	4,177	Real	?
CPU activity	22	8,192	Real	?
Naval propulsion plant	15	11,934	Real	?
Ailerons	40	13,750	Synth	N
Miami housing	16	13,932	Real	?
Elevators	18	16,599	Synth	N
Bike	12	17,379	Real	?
California housing	10	20,640	Real	?
Superconductivity	82	21,263	Real	?
Health insurance	12	22,272	Real	?
House8L	8	22,784	Real	?
NZ energy price	11	26,121	Real	?
MetroTraffic	7	48,204	Real	?
Diamonds	10	53,940	Real	?
Video transcoding	18	68,784	Real	?
Wave energy	49	71,993	Real	?
FriedmanGra	10	100,000	Synth	A
FriedmanGsg	10	100,000	Synth	G
FriedmanLea	10	100,000	Synth	A
Hyper_a	10	500,000	Synth	A

Bold values indicate the best performance for each dataset

- House8L*: predicts the house price in 1990s, US;
- NZ Energy Price: provided by the Electricity Market Information website (EMI) for different locations (points of connection), with the aim of predicting electricity prices 24 h ahead in real time. The Hamilton region is used in this work, and the data are available in a GitHub repository;
- MetroTraffic*: predicting the impacts on the traffic volume from weather information;
- Diamonds [47]: predicts the price of diamonds based on the physical attributes;
- Video Transcoding[5] [48]: predicts video transcoding time according to input information and output format;
- Wave Energy*: consists of positions and absorbed power outputs of wave energy converters and the goal is to predict the total energy converted;
- FriedmanGra/Gsg/Lea: Friedman synthetic dataset family proposed in [49], with a publicly accessible generator in the river machine learning platform. *Gra*—Global Recurring Abrupt drift; *Gsg*—Global and Slow Gradual drift; and *Lea*—Local Expanding Abrupt drift;
- Hyper_a [50]: simulated dataset with three abrupt drifts, predicting the distance from a random point to a hyperplane in the feature space. The generator is available in CapyMOA [51];

[5] The feature "ID" was in string format and was therefore removed from our experiments.

6.2 Regression models

Because both MVE and AdaPI operate on top of a point predictor, the quality of the underlying regressor strongly affects the resulting intervals. We therefore evaluate each PI method with three different streaming regressors: k-Nearest Neighbours (kNN), Adaptive Random Forest Regression (ARF-Reg) [52], and Self-Optimizing k-Nearest Leaves (SOKNL) [53].

6.2.1 K-Nearest neighbours

kNN is a simple and well-known nonparametric learner. In streaming settings it is typically paired with a sliding window so that prediction depends only on recent instances. For regression, the output is usually the average or median target value of the nearest neighbours in the current window.

6.2.2 Adaptive random forest regression

ARF-Reg extends Adaptive Random Forest [54] to regression. It uses FIMT-DD [49] as the base tree learner, injects diversity via bagging [55], employs ADWIN [56] for drift detection, and averages the predictions of the ensemble members.

6.2.3 Self-optimizing k-nearest leaves

SOKNL combines ideas from kNN and ARF-Reg. It augments each tree node with summary statistics that define a centroid-like representation. At prediction time, the learner compares the incoming instance with these node representations and aggregates information from the most relevant leaves. The value of k is selected dynamically according to past performance.

6.3 Experimental setting

This subsection reports the parameter settings used throughout the experiments.

At the base learner level, the parameters are set as follows:

- KNN: sliding window size = 1000, k = 10;
- ARF-Reg: ensemble size = 100, feature subspace size = 60%, lambda value for Poisson distribution = 6, base learner = FIMT-DD, base learner grace period = 50, base learner split confidence = 0.01;
- SOKNL: ensemble size = 100, feature subspace size = 60%, lambda value for Poisson distribution = 6, base learner = SOKNL-BT (Base Tree),[6] base learner grace period = 50, base learner split confidence = 0.01;

At the prediction interval level, we evaluate the following settings:

- $\mathcal{C} \in \{80\%, 90\%, 95\%, 99\%\}$; and
- $\mathcal{M} \in \{0.05, 0.1, 0.5\}$.[7]

All experiments include a warm-up phase of 100 examples so that the initial MVE estimates are based on a non-trivial amount of residual information before adaptive scaling begins.

[6] A modification of FIMT-DD that stores necessary information for the centroid.

[7] Limit $\mathcal{M}$ is not applicable to MVE.

7 Results and discussion

Experimental results and relevant discussion are presented in this section.

7.1 Raw results

Tables 2 and 3 report representative raw results for confidence level $\mathcal{L} = 95\%$ and lower limit $\mathcal{M} = 0.1$. We show this setting in the main text because it sits between the most conservative and most aggressive lower limit choices, while the full set of results is provided in "Appendix A". Within each dataset block, bold values identify the result closest to the target confidence level in the coverage columns and the smallest interval width in the NMPIW columns.

Several broad patterns are visible in these tables. First, the adaptive methods usually improve calibration relative to plain MVE, although the magnitude of that improvement depends heavily on the dataset and the base regressor. Second, at the 95% setting reported in the main tables, the linear rule most often provides the best balance between matching the target coverage and limiting interval width. This should not be interpreted as a confidence level-independent ranking: at lower confidence levels the same aggressive correction can lead to under-coverage, while at very high confidence levels it can widen intervals beyond MVE. Third, the base learner matters substantially: for example, on MetroTraffic the same adaptive rule can produce noticeably different NMPIW values across ARF-Reg, SOKNL, and kNN even when coverage is similar.

This dependence on the base regressor is expected. According to Eq. 12, AdaPI rescales an interval generated from the underlying predictive residuals, so stronger point predictors can indirectly lead to narrower final intervals. In our experiments, SOKNL often correlates with lower interval width for exactly this reason.

The Naval Propulsion Plant results illustrate an important caveat. For that dataset, some combinations of kNN or SOKNL with the adaptive methods achieve both high coverage and very small NMPIW. While this may appear ideal, the behaviour is likely influenced by properties of the dataset itself, including a very narrow target range and an ordering that makes point prediction unusually easy. We therefore treat these results as informative but atypical rather than as decisive evidence of general superiority.

Because the raw tables contain many settings and many datasets, they are useful mainly for detailed inspection. To obtain a clearer global picture of the tradeoff between coverage and width, we next turn to the CING evaluation.

7.2 Results from CING evaluation

As discussed in Sect. 4.3.2, CING summarizes each PI run using coverage error and interval width. This makes it possible to compare multiple hyperparameter settings and confidence levels within a single visual summary for each dataset.

Figure 4 shows representative CING plots; the remaining datasets appear in Figs. 16 and 17 in appendix. In each plot, translucent blue markers indicate dominated solutions, while the coloured markers form the non-dominated group. The red frame identifies the best non-dominated solution under the weighting used in this paper, namely 0.8 for coverage error and 0.2 for NMPIW. This choice reflects our emphasis on calibrating coverage first and then preferring narrower intervals among the well-calibrated candidates.

Table 2 Coverage (%) and NMPIW (%) results: confidence level $\mathcal{L}$ = 95% and limit $\mathcal{M}$ = 0.1 (Part I)

	ARF-Reg		SOKNL		KNN	
	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW
Logarithm	94.88	35.12	95.3	30.92	95.26	34.36
HLI	94.72	36.11	95.36	32.34	**94.98**	35.21
Linear	94.82	35.17	95.1	**30.69**	95.2	34.3
MVE	94.68	34.98	95.62	31.9	94.96	33.81
Abalone						
Logarithm	97.81	33.85	97.6	23.75	98.0	24.49
HLI	97.25	33.15	97.23	23.17	97.51	22.78
Linear	**96.14**	31.17	96.36	20.48	96.67	**19.19**
MVE	99.71	40.54	98.44	27.0	98.68	30.34
CPU activity						
Logarithm	96.53	11.42	99.99	1.2	99.92	1.31
HLI	96.37	10.91	99.99	**0.68**	99.85	0.75
Linear	**95.44**	10.46	99.99	**0.68**	99.72	0.74
MVE	97.1	13.24	99.99	6.78	99.99	6.82
Naval propulsion plant						
Logarithm	95.2	30.34	95.4	28.97	94.92	23.4
HLI	95.05	30.43	95.37	29.18	94.71	23.3
Linear	**95.03**	30.0	95.11	28.31	94.88	23.34
MVE	94.8	29.62	95.16	28.71	94.41	**22.8**
Ailerons						
Logarithm	97.28	29.58	97.37	27.42	97.39	24.52
HLI	96.95	28.51	97.02	26.83	97.06	24.06
Linear	97.08	26.96	97.19	24.93	97.1	**22.53**
MVE	97.03	29.52	**96.91**	27.59	97.17	24.99
Miami housing						
Logarithm	96.86	4.54	97.27	3.88	96.8	4.01
HLI	96.89	4.53	97.22	3.83	96.85	4.03
Linear	96.57	4.19	96.72	**3.46**	**96.49**	3.73
MVE	96.95	4.72	97.45	4.2	96.99	4.2
NZ energy price						
Logarithm	97.7	5.64	97.85	5.61	97.4	6.57
HLI	97.42	5.42	97.51	5.37	97.14	6.38
Linear	96.62	4.83	96.64	**4.76**	**96.31**	5.79
MVE	98.88	8.3	98.17	6.34	97.66	7.17
Elevators						
Logarithm	93.85	39.56	**93.97**	33.39	93.76	53.74
HLI	92.58	35.61	92.69	30.48	92.32	49.37
Linear	93.3	37.46	93.47	32.01	93.05	51.45
MVE	89.53	28.88	89.56	**25.86**	88.16	39.1

Table 2 continued

	ARF-Reg		SOKNL		KNN	
	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW
Bike						
Logarithm	94.55	51.59	**94.62**	44.56	94.53	49.55
HLI	93.71	49.29	93.89	42.42	93.65	47.28
Linear	94.33	51.38	94.48	44.15	94.34	49.34
MVE	92.65	45.52	93.2	**39.92**	92.75	44.23
California housing						
Logarithm	96.31	46.29	96.34	44.5	96.55	41.09
HLI	96.34	45.37	96.35	44.5	96.32	39.84
Linear	95.91	42.83	**95.9**	41.85	96.16	**37.82**
MVE	96.55	46.72	96.54	45.82	96.12	39.69
Superconductivity						

Bold values indicate the best performance for each dataset

Three representative patterns appear in Fig. 4. In Abalone, several adaptive runs belong to the non-dominated group, suggesting that no single variant dominates across all parameter settings and that different scaling rules offer slightly different tradeoffs. In Bike, by contrast, a single solution dominates the rest, so the weighted ranking becomes irrelevant because all alternatives are clearly worse in at least one objective. In Diamonds, an MVE solution emerges as the top-ranked point, illustrating that adaptation is not universally beneficial: when the adaptive correction changes width more than it improves coverage, plain MVE can remain competitive or even preferable under the chosen weights.

The Diamonds result is also consistent with the stable stream case in which adaptive calibration is least necessary. The task predicts diamond price from physical measurements and related attributes (Sect. 6.1), so the underlying relationship is comparatively stationary and does not contain the kind of temporal shift that motivates online recalibration. In this setting, the base regressors can capture much of the systematic structure, leaving residuals that are closer to the Gaussian assumption used by MVE. When MVE is already close to the desired coverage, AdaPI has little systematic bias to correct; small factor updates may then add width without producing a meaningful coverage gain. Practitioners should therefore avoid applying adaptive calibration automatically on stable regression streams where a validation prefix or rolling diagnostic shows that MVE already maintains the target coverage with well-behaved residuals. In such cases, the unadapted MVE interval can be the simpler and more efficient choice.

7.3 Dealing with concept drifts

In this subsection, we examine how AdaPI reacts when the stream distribution changes over time.

Because ground truth drift locations are rarely available in real streams, synthetic data remain useful for controlled analysis. We therefore focus on HyperA, a commonly used synthetic stream with three abrupt drifts [50, 52, 53].

Figure 5 illustrates the variations over time in the windowed coverage and the corresponding adjustment factors generated by our adaptive prediction interval methods.

Table 3 Coverage (%) and NMPIW (%) results: confidence level $\mathcal{L}$ = 95% and limit $\mathcal{M}$ = 0.1 (Part II)

	ARF-Reg		SOKNL		KNN	
	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW
Logarithm	97.15	68.97	96.5	70.56	95.43	67.76
HLI	96.86	68.38	96.39	70.57	95.54	68.6
Linear	95.86	**65.89**	95.53	68.87	**95.16**	67.33
MVE	98.85	77.9	98.63	76.5	95.75	69.03
Health insurance						
Logarithm	95.95	28.56	96.19	27.13	95.37	31.46
HLI	95.98	28.87	96.17	27.27	95.43	31.94
Linear	95.45	26.71	95.56	**25.08**	**95.15**	30.91
MVE	96.32	29.99	96.61	28.83	95.46	32.01
House8L						
Logarithm	95.65	90.11	95.71	86.26	95.79	102.34
HLI	95.77	90.79	95.83	86.99	95.91	102.82
Linear	**95.24**	88.54	95.26	**84.71**	**95.24**	100.55
MVE	96.22	92.77	96.32	88.96	96.88	106.66
MetroTraffic						
Logarithm	96.23	29.23	96.08	27.48	96.12	26.9
HLI	95.35	27.21	95.16	25.11	95.55	23.57
Linear	95.66	28.21	95.39	26.22	95.78	24.93
MVE	93.06	22.49	93.43	**19.75**	**94.89**	20.92
Diamonds						
Logarithm	95.15	19.3	95.33	11.54	95.17	11.59
HLI	95.1	19.18	95.37	11.6	95.09	11.42
Linear	95.08	19.18	95.13	11.3	95.12	11.54
MVE	94.96	19.04	95.36	11.62	**94.98**	**11.26**
Video transcoding						
Logarithm	96.38	20.66	**94.74**	6.17	94.6	5.93
HLI	96.34	20.58	94.28	5.96	93.94	**5.68**
Linear	95.54	19.96	94.63	5.98	94.44	5.75
MVE	98.28	22.84	94.75	6.47	94.27	6.14
Wave energy						
Logarithm	95.26	36.76	95.32	26.99	95.01	33.38
HLI	95.38	36.99	95.4	27.13	95.01	33.41
Linear	95.08	36.38	95.11	**26.7**	**95.0**	33.36
MVE	95.73	37.64	95.92	27.96	95.09	33.5
FriedmanGra						
Logarithm	95.17	36.78	95.15	27.12	95.02	33.48
HLI	95.27	36.98	95.17	27.17	94.93	33.37
Linear	95.05	36.48	95.05	**26.93**	**95.0**	33.46
MVE	95.59	37.57	95.62	27.92	**95.0**	33.44

Table 3 continued

	ARF-Reg		SOKNL		KNN	
	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW
FriedmanGsg						
Logarithm	94.86	30.0	94.78	22.0	94.2	29.72
HLI	94.6	29.76	94.56	**21.79**	93.35	28.45
Linear	94.77	29.85	94.63	21.81	93.99	29.38
MVE	94.86	30.17	**95.04**	22.48	92.93	27.86
FriedmanLea						
Logarithm	94.9	28.19	94.68	17.11	94.64	18.27
HLI	94.86	28.14	94.18	**16.63**	94.04	17.7
Linear	94.84	28.09	94.62	17.02	94.57	18.19
MVE	**95.03**	28.41	94.17	16.68	93.7	17.39
Hyper_a						

Bold values indicate the best performance for each dataset

AdaPI responds to drift indirectly through coverage. When a drift occurs, point prediction quality typically deteriorates first, which reduces interval coverage. The adaptive factor then increases, widening the interval until coverage recovers. In this sense, AdaPI acts as a passive but immediate calibration layer on top of the base regressor.

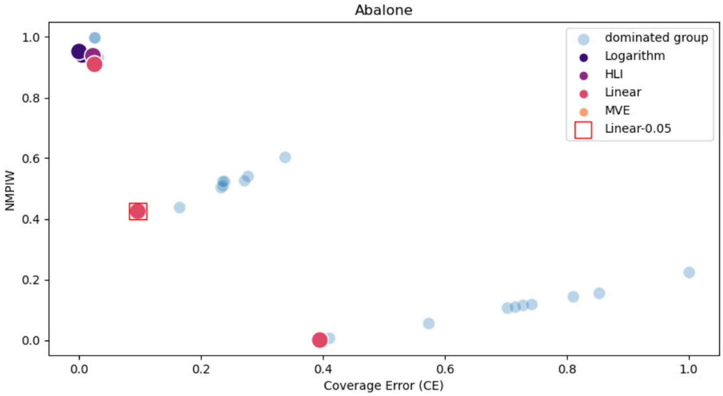
This behaviour is visible in Fig. 5. The three abrupt drifts in HyperA occur at instances 125,000, 250,000, and 375,000, and each drift is followed by a drop in coverage and a corresponding increase in the adaptive factor. Coverage then returns towards the target level. The figure therefore supports the main intuition behind AdaPI: when the stream becomes harder, the method expands the interval; when the stream stabilizes again, the interval contracts.

The three adaptive rules react with different intensity. HLI tends to adjust more gently, whereas the linear rule changes more aggressively. This difference helps explain why the linear variant often performs well empirically under the default 95% setting: stronger corrections can recover target coverage faster, although they may also be more sensitive in borderline or extreme confidence level settings.

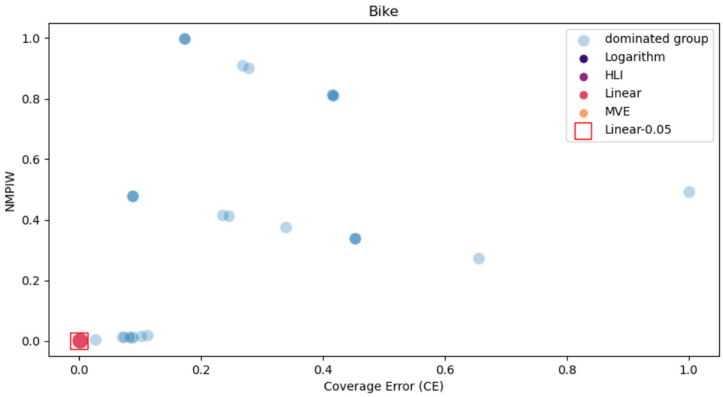
7.4 Behavioural visualization

Figure 6 provides a more local view of how the adaptive mechanism behaves in practice. Although the three scaling functions differ analytically, their qualitative behaviour is similar: they shrink MVE when coverage is higher than desired and expand it when coverage falls below the target. To keep the visual discussion focused, we therefore illustrate only the logarithmic variant with $\mathcal{L} = 95\%$ and $\mathcal{M} = 0.1$.

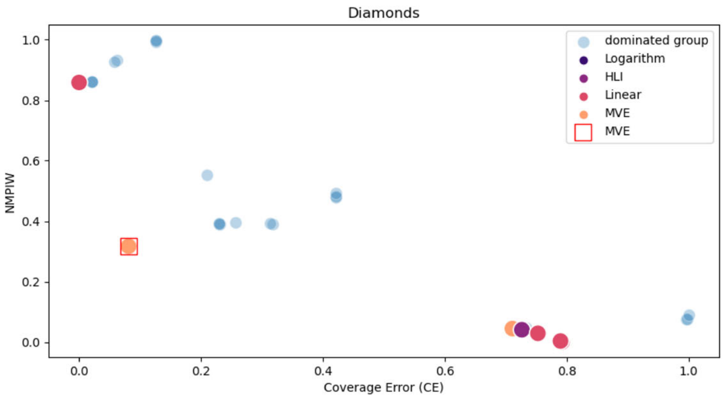
The three panels correspond to three characteristic situations. In Fig. 6a, the underlying MVE intervals are slightly too conservative, so the adaptive rule mostly shrinks them. In Fig. 6b, the opposite happens: MVE undercovers and the adaptive rule expands the interval to capture more targets. Figure 6c shows a mixed scenario on the NZ Energy Price stream, where the method initially shrinks MVE and then switches to expansion after a change in the target pattern reduces coverage. This panel is useful because it highlights the mechanism that motivates AdaPI: the interval width is not fixed but continuously adjusted according to recent empirical behaviour.



(a) Abalone



(b) Bike



(c) Diamonds

Fig. 4 Partial CING evaluation results for different datasets

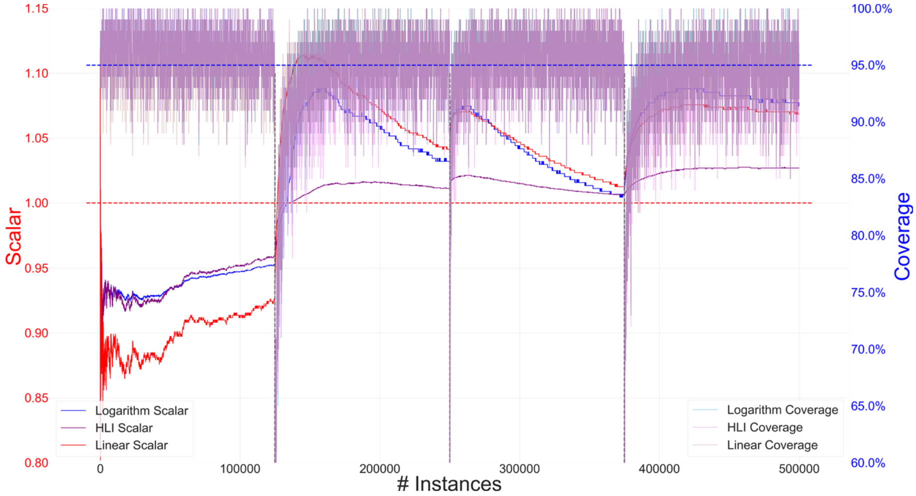


Fig. 5 Factors from different adaptive approaches and the correlated coverage for SOKNL on HyperA dataset over time. The black vertical lines highlight positions of concept drifts; the red horizontal line marks the value 1.0; and the blue horizontal line emphasizes 95% confidence level (colour figure online)

7.5 Coverage results comparison

Coverage remains the primary calibration objective of a PI method, which is also why CE receives the higher weight in CING. Nevertheless, it is informative to examine coverage alone and ask whether AdaPI actually moves MVE towards the desired confidence level.

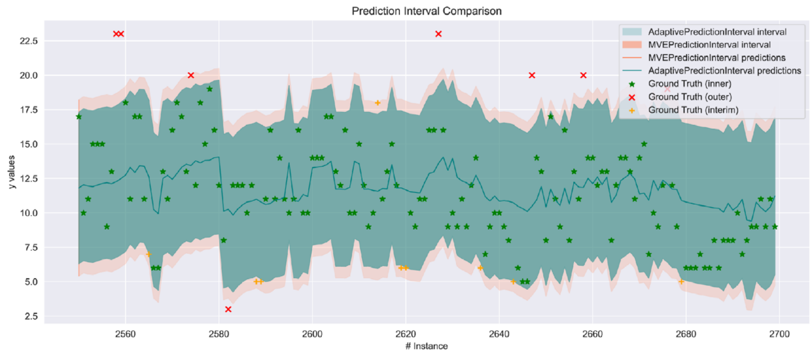
Figure 7 compares the coverage of MVE with the coverage of all adaptive runs at confidence level 95%.

The black lines mark the desired confidence level. The boundary of the green region corresponds to equal coverage error for AdaPI and MVE, that is, $\mathrm{CE}_{\mathrm{MVE}} = \mathrm{CE}_{\mathrm{AdaPI}}$. Points inside the green region therefore indicate cases where AdaPI is closer to the target confidence level than plain MVE.

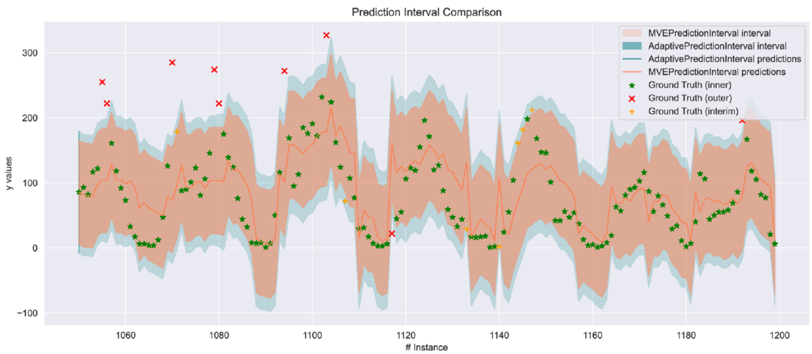
Most points fall inside this region, which supports the main empirical claim of the paper: AdaPI usually pushes MVE towards the target coverage. The exceptions tend to cluster near the intersection of the guide lines, where MVE is already close to the target and the adaptive mechanism alternates between expansion and shrinkage. In such borderline cases, small corrections can occasionally make the adaptive result slightly worse. Even so, the overall trend remains favourable to AdaPI. Corresponding plots for the other confidence levels are included in "Appendix C".

7.6 AdaPI vs. conformal prediction

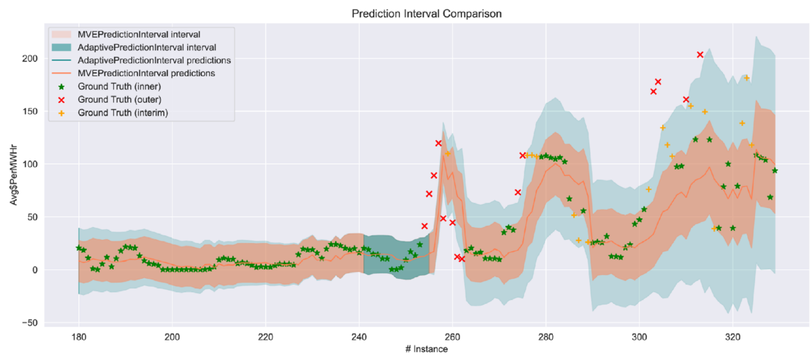
We compare AdaPI with MVE and several recent conformal methods introduced in Sect. 3, namely PID, MVP, SAOCP, ACI, and AgACI. All methods are evaluated at a confidence level of 95% for consistency. As most conformal baselines are not strictly incremental in the stream learning setting, we conduct two types of comparison experiments: (1) using global calibration statistics (global) and (2) using a sliding window of 500 instances for calibration (500-Windowed).



(a) Abalone Data: Adaptive PI constantly smaller than MVE



(b) Bike Data: Adaptive PI constantly larger than MVE



(c) NZ Energy Price Data: Adaptive PI fluctuates between different modes

Fig. 6 MVE Area (coral) and Adaptive PI approach (green), ground truths covered by both areas (green "★"), ground truths covered by neither areas (red "×"), and ground truths covered by wider area but not narrower area (orange "+") (colour figure online)

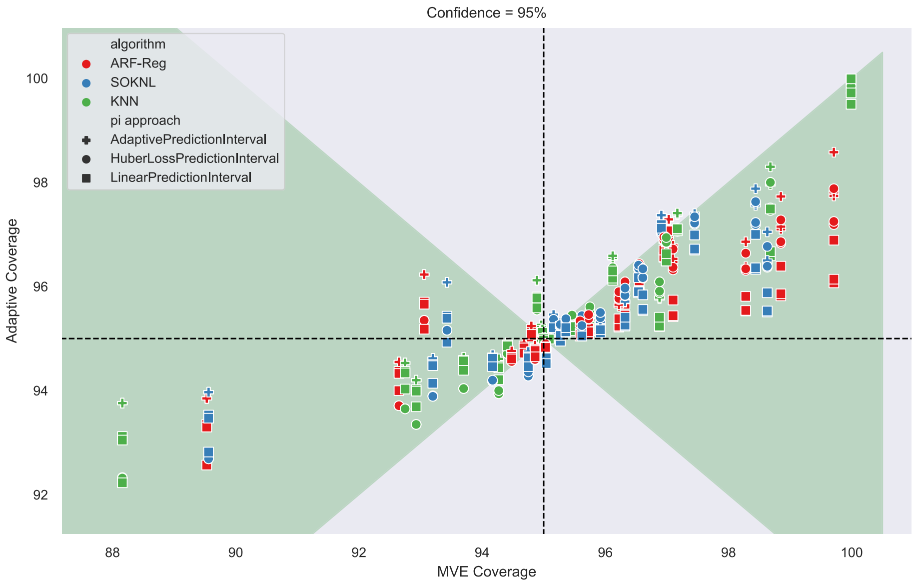


Fig. 7 Scattergram for $\mathcal{L} = 95\%$. The x-axis is the coverage value for MVE, and the y-axis shows the coverage value for adaptive approaches, black lines denote the 95% confidence level for both axes, green area highlights the space in which the adaptive approaches outperform MVE in terms of coverage, different colours represent different base learners, and different shapes distinguish different adaptive approaches (colour figure online)

In the global setting, calibration information for both AdaPI and conformal methods is accumulated from the start of the stream up to the current time step. For AdaPI, this means that the empirical coverage used by the factor generator is computed globally. In contrast, the 500-windowed setting restricts the calibration signal for all compared methods to the most recent 500 instances; for AdaPI, the adaptive factor is therefore generated from windowed coverage rather than from cumulative coverage. Additionally, a warm-up phase of 1000 instances is applied to conformal methods to stabilize their statistics, which may provide them with a slight advantage over AdaPI. The SOKNL is used as the common base regressor, and datasets containing unsupported nominal attributes are excluded from this comparison.

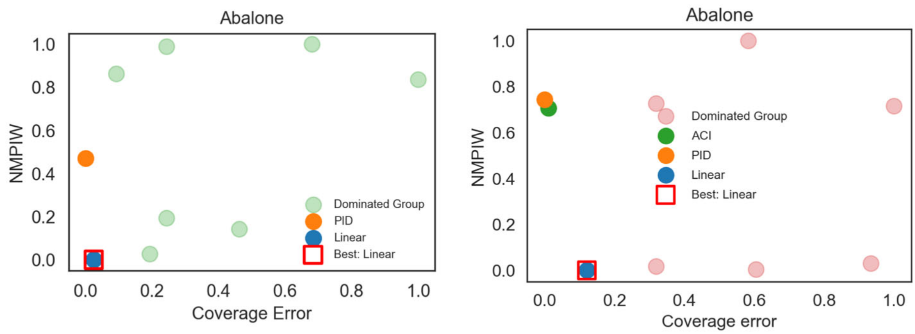
7.6.1 CING results

Figures 8 and 9 present a subset of the CING results, providing a visual comparison between the global and windowed experimental settings.[8] A fixed weighting of 0.8/0.2 is applied throughout.

Figure 8 shows the results on the Abalone dataset. In the global setting, the linear variant of AdaPI is selected as the best point under the 0.8/0.2 CING weighting, while PID also appears in the non-dominated set. Under the 500-instance windowed setting, linear remains the selected best point, but ACI and PID move closer to the AdaPI solution and also appear in the non-dominated group. This indicates that windowed calibration can make conformal methods more competitive, even when the final weighted CING choice still favours AdaPI.

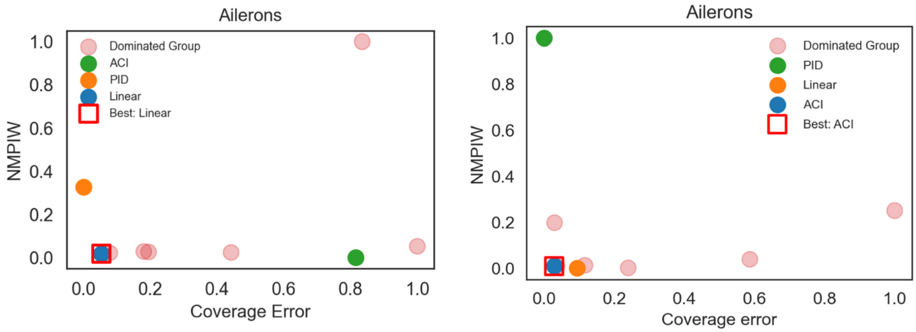
Figure 9 presents a complementary pattern on the Ailerons dataset. In the global setting, the linear variant of AdaPI is selected as the best point under the CING weighting, with PID

[8] The remaining plots are provided in "Appendix D".



(a) Global (b) 500-Windowed

Fig. 8 CING results of AdaPI vs. conformal methods on abalone



(a) Global (b) 500-Windowed

Fig. 9 CING results of AdaPI vs. conformal methods on ailerons

and ACI also appearing on the non-dominated frontier. Under the 500-instance windowed setting, the preferred solution shifts to ACI, while linear remains close to the lower-left region. This example shows that applying the same local calibration protocol can change the ranking materially and may favour conformal methods on some streams.

Overall, these examples do not suggest that conformal methods are inherently weak. Rather, they show that the relative ranking is sensitive to the calibration protocol and dataset. AdaPI remains competitive in both examples, but the windowed setting can substantially improve the position of conformal methods in the CING trade-off space.

7.6.2 CING sensitivity evaluation

This subsection examines how sensitive the CING evaluation is to different weighting schemes. Figure 10 shows a heatmap of the winning candidate under different weights across all datasets. The columns represent the weight assigned to coverage error (CE), and the rows represent datasets. The number shown to the right of each row reports how many distinct methods are selected for that dataset across the full range of weights; smaller values indicate that the CING choice is relatively insensitive to the weighting. Figure 11 shows the corresponding results for the 500-windowed setting.

Figure 10 shows that the global setting has some weight dependence, but the dependence is not uniformly strong. Across the 176 dataset weight combinations, PID is selected most

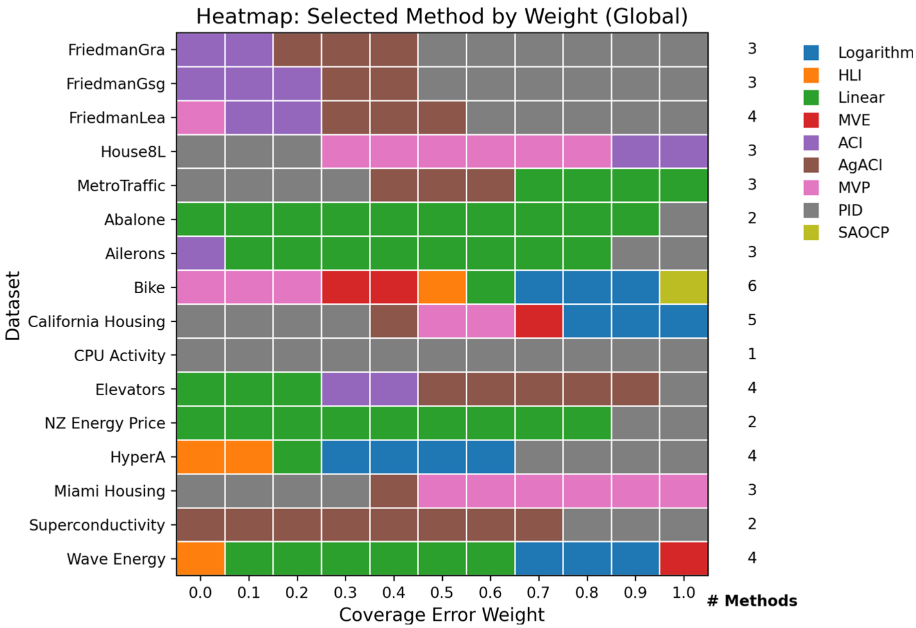


Fig. 10 Heatmap for CING sensitivity evaluation (Global)

often, with 56 selections, followed by linear with 42, AgACI with 26, MVP with 18, logarithm with 13, and ACI with 12. Thus, linear remains the most frequent AdaPI winner, but it is not the overall most frequent winner once conformal methods are included. The distinct winner counts also show that the CING choice is relatively stable for most datasets: 10 of the 16 datasets select no more than three distinct methods across all weights, and 14 select no more than four. CPU Activity is completely stable and is always assigned to PID, while Abalone, NZ Energy Price, and Superconductivity use only two selected methods. The main exceptions are Bike and California Housing, where the selected method changes more frequently as the CE weight increases.

Figure 11 shows a different pattern under the 500-windowed setting, but again the weighting sensitivity is concentrated in a limited number of datasets. Linear becomes the most frequent winner with 65 selections, followed by ACI with 48, MVP with 26, and PID with 19. This indicates that windowed calibration changes the trade-off structure: AdaPI remains highly competitive, but ACI and MVP become frequent winners, especially when CE receives moderate or high weight. At the dataset level, 11 of the 16 datasets select no more than three distinct methods across all weights, and 13 select no more than four. Several datasets therefore show broad stability under the windowed protocol, including FriedmanGra, FriedmanLea, House8L, MetroTraffic, Abalone, CPU Activity, and Wave Energy, each of which uses only two selected methods. Bike, HyperA, and NZ Energy Price remain the most weight-sensitive cases. Overall, the sensitivity analysis suggests that CING is reasonably robust to moderate changes in weighting for most datasets. When sensitivity appears, it is mainly dataset dependent rather than a general weakness of the CING weighting scheme and occurs on streams where several methods have close CE/NMPIW trade-offs.

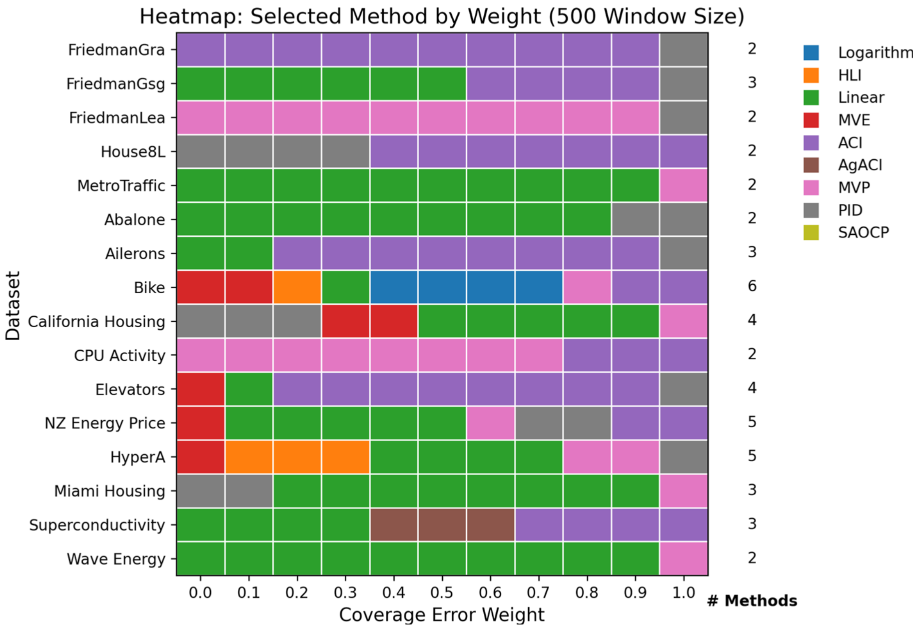


Fig. 11 Heatmap for CING sensitivity evaluation (500-windowed)

7.6.3 Full results

Tables 4 and 5 report the full comparison under the global and 500-windowed settings, respectively. As before, coverage is preferred when it is closer to 95%, and NMPIW is preferred when it is smaller.

Under the global setting (Table 4), the comparison is more balanced than in the earlier CING plots alone. The AdaPI variants remain close to the target coverage on many datasets and usually keep NMPIW within a moderate range. The linear variant is often the strongest AdaPI option, especially where narrow intervals are achieved without moving far from 95% coverage. However, the conformal baselines are also competitive on several datasets. For example, ACI, MVP, and PID often achieve coverage very close to the target and in some cases also produce narrower intervals than the AdaPI variants.

The global results also show that no method is uniformly reliable across all datasets. Some conformal results are attractive on one metric but weak on the other: PID gives very narrow intervals on House8L and California Housing, but with clear under-coverage, while PID and MVP produce very large NMPIW on Elevators. AdaPI avoids the largest width failures in this table, but it is not always the closest method to the target coverage and it is not always the narrowest. Thus, the main conclusion from the full table is not that one family dominates the other, but that the coverage–width trade-off remains dataset dependent.

Overall, these results reinforce the need to evaluate coverage and width jointly. A method that is excellent on CE can still be unattractive if it produces very wide intervals, and a method with compact intervals can be unsuitable when coverage falls far below the nominal level. CING is useful here because it exposes these trade-offs rather than reducing the comparison to a single raw metric. To complement the descriptive comparison, we apply a Friedman test

Table 4 Coverage (%) and NMPIW (%) comparison for AdaPI and conformal prediction methods (global): confidence level $\mathcal{L} = 95\%$ and limit $\mathcal{M} = 0.1$

Method	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW
Logarithm	95.32	26.99	95.15	27.12	94.78	22.00	96.19	27.13
HLI	95.40	27.13	95.17	27.17	94.56	21.79	96.17	27.27
Linear	95.11	26.70	95.05	26.93	94.63	21.81	95.56	25.08
MVE	95.92	27.96	95.62	27.92	95.04	22.48	96.61	28.83
ACI	94.02	**24.62**	93.97	**24.82**	94.47	20.26	**94.97**	24.75
AgACI	94.43	24.96	94.36	25.17	94.74	20.63	94.12	22.73
MVP	93.97	25.00	93.73	25.07	94.20	**20.21**	94.00	16.39
PID	**95.00**	26.33	**95.00**	26.79	**95.00**	21.86	73.57	**7.84**
SAOCP	95.59	28.10	95.60	28.44	95.62	23.97	95.80	36.18
Dataset	**FriedmanGra**		**FriedmanGsg**		**FriedmanLea**		**House8L**	
Method	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW
Logarithm	95.71	86.26	95.30	30.92	95.40	28.97	93.97	33.39
HLI	95.83	86.99	95.36	32.34	95.37	29.18	92.69	30.48
Linear	**95.26**	84.71	95.10	**30.69**	95.11	28.31	93.47	32.01
MVE	96.32	88.96	95.62	31.90	95.16	28.71	89.56	25.86
ACI	94.39	85.48	94.64	39.08	93.32	**26.94**	90.11	34.64
AgACI	93.51	78.13	94.82	38.01	94.09	28.87	90.51	34.69
MVP	93.20	79.83	93.74	37.78	96.72	100.48	78.41	**21.22**

Table 4 continued

Method	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW
PID	87.53	**71.13**	**95.07**	34.68	**95.00**	51.03	93.15	38.64
SAOCP	95.53	84.73	95.88	39.17	97.06	30.88	**95.71**	45.46
Dataset	**MetroTraffic**		**Abalone**		**Ailerons**		**Bike**	
Method	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW
Logarithm	**94.62**	44.56	97.60	23.75	97.85	5.61	97.27	3.88
HLI	93.89	42.42	97.23	23.17	97.51	5.37	97.22	3.83
Linear	94.48	44.15	96.36	20.48	96.64	**4.76**	96.72	**3.46**
MVE	93.20	39.92	98.44	27.00	98.17	6.34	97.45	4.20
ACI	93.39	46.45	94.73	17.62	94.46	84.34	90.83	23.30
AgACI	89.86	34.86	96.72	19.74	94.97	166.74	93.83	23.34
MVP	91.95	37.14	96.66	18.13	96.26	884.92	86.49	22.79
PID	65.00	**16.17**	**95.15**	**16.95**	**95.00**	601.05	**94.84**	23.71
SAOCP	96.68	54.48	96.19	25.90	99.60	139.35	97.04	26.48
Dataset	**California Housing**		**CPU Activity**		**Elevators**		**NZ Energy Price**	
Method	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW
Logarithm	94.68	17.11	97.37	27.42	96.34	44.50	94.74	6.17
HLI	94.18	**16.63**	97.02	26.83	96.35	44.50	94.28	**5.96**
Linear	94.62	17.02	97.19	24.93	95.90	41.85	94.63	5.98
MVE	94.17	16.68	96.91	27.59	96.54	45.82	**94.75**	6.47
ACI	94.24	18.97	92.71	22.88	91.56	40.36	94.08	23.22
AgACI	94.51	19.15	89.79	16.07	95.13	**39.02**	93.03	22.08
MVP	93.70	18.79	**93.43**	17.80	92.45	39.44	92.98	22.77
PID	**95.00**	19.98	63.56	**7.33**	**95.04**	39.69	61.56	10.93
SAOCP	95.67	21.68	96.60	24.42	96.54	41.62	96.56	35.92
Dataset	**HyperA**		**Miami Housing**		**Superconductivity**		**Wave Energy**	

Bold values indicate the best performance for each dataset

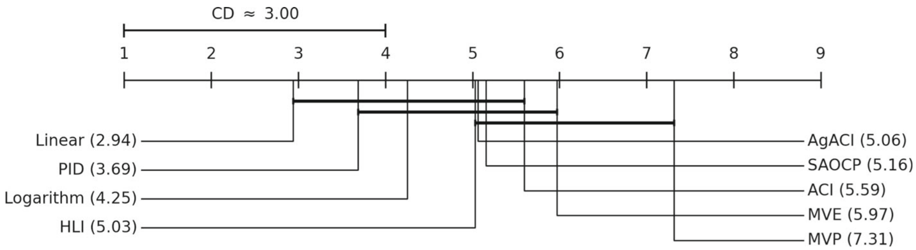


Fig. 12 Critical difference diagram for coverage error (global)

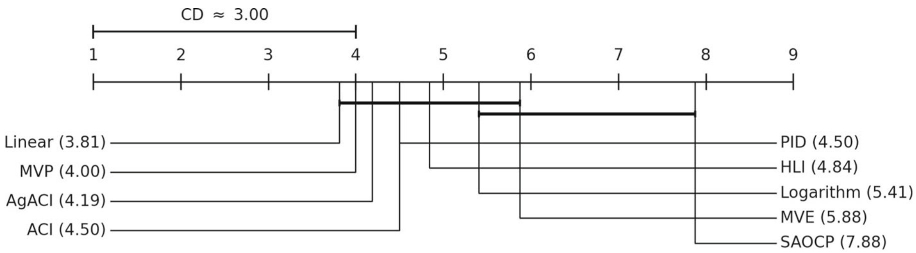


Fig. 13 Critical difference diagram for NMPIW (global)

followed by Nemenyi post hoc comparison and report the resulting average-rank relationships using critical difference diagrams.

Figure 12 shows that linear obtains the best average rank for coverage error in the global setting, followed by PID and logarithm. The gap between several middle-ranked methods is within the critical difference, so the figure should be read as evidence of a broad competitive group rather than a strict total ordering. MVP has the weakest average CE rank in this setting, reflecting the dataset-specific calibration failures shown in Table 4.

For NMPIW, Fig. 13 again places linear at the best average rank, but several conformal methods are close: MVP, AgACI, PID, and ACI all rank ahead of HLI and MVE on average. This differs from the older interpretation that conformal methods are generally separated towards the worst ranks. Instead, the global CD diagrams show that linear is the most consistently balanced single method, while some conformal baselines are competitive on width but less stable on coverage.

Taken together, the global CD diagrams support a more nuanced conclusion: AdaPI, especially the linear variant, provides a strong empirical compromise, but the conformal baselines should not be characterized as uniformly weak. Their performance depends strongly on the dataset and on whether the metric emphasizes coverage error or interval width.

Under the 500-windowed setting (Table 5), all methods, including AdaPI, use only the most recent 500 instances for their calibration signal. This makes the comparison more locally adaptive and changes the relative behaviour of the conformal methods. ACI and MVP in particular often move closer to the target 95% coverage than they do in the global setting, and in several datasets they become the strongest options on coverage error. This supports the view that the calibration protocol matters substantially for these baselines.

The windowed table also shows that this improvement is not uniform across all conformal methods and datasets. PID remains unstable in several cases: it under-covers strongly on House8L, California Housing, Miami Housing, and NZ Energy Price, and it still produces a very large NMPIW value on Elevators. AdaPI variants are generally more moderate in

Table 5 Coverage (%) and NMPIW (%) comparison for AdaPI and conformal prediction methods (500-windowed): confidence level $\mathcal{L} = 95\%$ and limit $\mathcal{M} = 0.1$

Method	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW
Logarithm	95.88	27.63	95.81	27.91	95.42	23.24	96.22	27.23
HLI	95.71	27.34	95.60	27.54	95.13	22.64	96.16	27.18
Linear	95.21	26.85	95.14	**27.16**	94.94	22.68	95.62	25.69
MVE	95.92	27.96	95.62	27.92	95.04	22.48	96.61	28.83
ACI	94.99	**25.58**	94.99	27.55	94.99	22.67	**94.99**	22.19
AgACI	94.98	25.66	94.98	27.88	94.96	23.29	94.62	30.81
MVP	94.98	27.86	94.98	27.91	94.97	**21.54**	94.83	22.13
PID	**95.00**	26.44	**95.00**	28.03	**95.00**	24.02	73.57	**7.96**
SAOCP	95.37	27.46	95.31	28.83	95.49	23.68	95.80	31.80
Dataset	**FriedmanGra**		**FriedmanGsg**		**FriedmanLea**		**House8L**	
Method	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW
Logarithm	95.78	86.57	95.92	32.58	95.81	30.00	94.03	34.70
HLI	95.67	86.03	95.36	32.27	95.33	28.16	92.96	31.65
Linear	95.26	**84.87**	95.18	**31.78**	95.13	**27.98**	93.52	32.74
MVE	96.32	88.96	95.62	31.90	95.16	28.71	89.56	**25.86**
ACI	95.02	92.12	94.92	50.78	94.96	28.48	**94.97**	59.75
AgACI	94.78	91.72	94.64	51.33	94.95	28.51	94.59	52.69
MVP	**95.01**	94.61	95.98	51.01	95.04	37.99	95.36	47.95
PID	87.53	89.71	**95.07**	51.77	**95.00**	78.32	93.15	64.48

Table 5 continued

Method	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW
SAOCP	95.14	88.92	95.60	58.64	96.38	40.61	95.60	60.15
Dataset	**MetroTraffic**		**Abalone**		**Ailerons**		**Bike**	
Method	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW
Logarithm	95.40	44.55	97.47	23.55	96.00	26.21	97.67	5.17
HLI	94.64	42.43	97.14	22.98	95.39	24.61	97.13	4.71
Linear	94.70	42.22	96.33	20.68	95.28	24.45	96.57	4.43
MVE	93.20	39.92	98.44	27.00	98.17	**6.34**	97.45	**4.20**
ACI	**95.03**	81.51	**94.92**	14.84	94.95	43.46	**95.05**	40.24
AgACI	94.52	73.88	94.87	15.76	94.88	43.76	94.64	29.77
MVP	**94.97**	50.49	95.22	**13.09**	94.87	61.92	95.22	32.02
PID	65.00	**23.85**	95.15	14.69	**95.00**	1371.91	94.84	33.54
SAOCP	96.47	75.11	96.09	16.47	96.49	49.80	96.65	37.07
Dataset	**California Housing**		**CPU Activity**		**Elevators**		**NZ Energy Price**	
Method	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW
Logarithm	95.19	17.57	97.70	24.29	96.68	40.28	95.57	6.51
HLI	94.80	17.06	97.06	21.44	96.32	38.35	95.14	6.24
Linear	94.81	17.20	96.67	20.41	96.10	**37.15**	95.04	**6.21**
MVE	94.17	**16.68**	96.91	27.59	96.54	45.82	94.75	6.47
ACI	95.00	31.69	94.84	35.48	**94.99**	48.71	**95.01**	54.28
AgACI	94.94	34.48	94.02	34.47	94.82	43.50	94.65	60.34
MVP	94.99	28.83	**95.04**	46.11	94.80	46.86	94.99	53.46
PID	**95.00**	31.23	63.56	**13.02**	95.04	47.81	61.56	24.93
SAOCP	95.49	31.05	96.50	40.53	96.36	59.85	95.76	56.54
Dataset	**HyperA**		**Miami Housing**		**Superconductivity**		**Wave Energy**	

Bold values indicate the best performance for each dataset

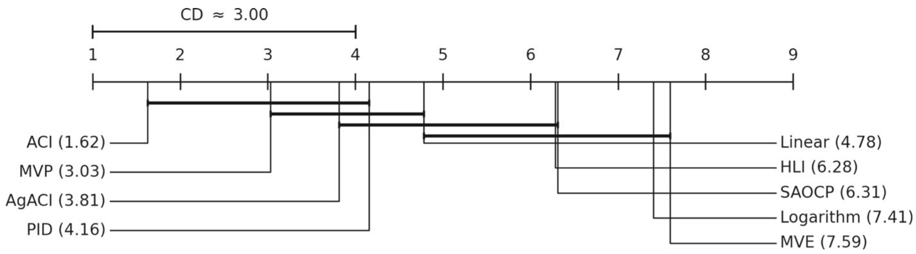


Fig. 14 Critical difference diagram for coverage error (500-windowed)

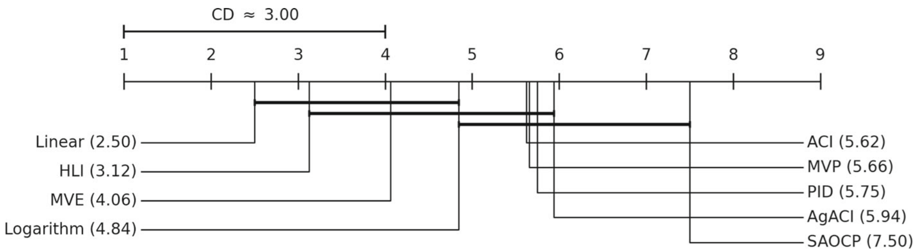


Fig. 15 Critical difference diagram for NMPIW (500-windowed)

width, and the linear variant frequently gives compact intervals, but the AdaPI variants are not always closest to the nominal coverage under the windowed protocol.

Overall, the 500-windowed setting narrows the gap between AdaPI and conformal approaches and, for coverage error, often reverses the ordering seen under global calibration. The best method therefore depends on the operational preference: conformal methods can be preferable when the priority is coverage alignment, while AdaPI, especially linear, remains attractive when interval efficiency and avoidance of very large widths are important.

Figure 14 confirms the change in coverage behaviour. ACI obtains the best average rank by a clear margin, followed by MVP, AgACI, PID, and linear. This indicates that the 500-windowed calibration protocol substantially improves the relative CE performance of the conformal methods. The AdaPI variants, especially HLI and logarithm, move to weaker average CE ranks in this setting, because their windowed factor updates are less directly targeted to the same coverage objective as the conformal calibration rules.

For NMPIW, Fig. 15 shows the complementary pattern. Linear has the best average rank, followed by HLI, MVE, and logarithm. The conformal methods are mostly in the middle or lower part of the ranking, with SAOCP the weakest on average. Thus, the 500-windowed protocol improves conformal calibration accuracy but does not remove the width advantage of the AdaPI variants, particularly linear.

Overall, the windowed setting highlights a genuine trade-off rather than a one-sided result. Conformal methods, especially ACI and MVP, become strong choices when coverage error is the primary objective. AdaPI remains competitive when interval width is also central, with linear providing the most consistent width efficiency across the evaluated datasets.

7.7 AdaPI with ensemble quantile

AdaPI, as a post-calibration framework, is not restricted to a specific base prediction interval (PI) estimator and can be applied to alternative PI constructions. In this section, we consider a simple incremental PI approach, namely ensemble quantile (EQ), and evaluate its performance when combined with AdaPI.

Specifically, ARF-Reg is used as the underlying ensemble model, and the $(1-\alpha)$ empirical quantiles of the base tree predictions are used to form the lower and upper bounds of the PIs, where $1-\alpha$ corresponds to the target confidence level (95%). This experiment aims to assess the generality of AdaPI when applied to PI estimators that are not explicitly designed for calibrated uncertainty estimation.

As shown in Table 6, the raw EQ method exhibits clear under-coverage across all datasets, indicating poor calibration. After applying AdaPI, all variants (logarithm, HLI, and linear) substantially improve coverage, bringing it closer to the nominal confidence level. This demonstrates that AdaPI can effectively correct miscalibrated intervals even when the base PI estimator is relatively weak.

Although MVE-based methods still dominate in terms of both coverage and interval efficiency (NMPIW), the consistent improvement over EQ confirms that AdaPI is model agnostic and can be integrated with diverse PI generation mechanisms.

8 Limitations and threats to validity

Several limitations should be considered when interpreting the results. First, AdaPI does not provide the same kind of formal sequential validity guarantees offered by some conformal methods. Its value is therefore primarily practical and empirical rather than guarantee driven. Second, because AdaPI acts as a post-calibration layer on top of a point predictor, its effectiveness depends strongly on the quality and behaviour of the underlying regressor. Poor point predictions can limit what any adaptive rescaling rule can recover.

Third, most AdaPI experiments in this study use a global coverage signal for adaptation. While this makes the method simple and fully incremental, it may also cause slower response or oscillatory behaviour when the stream changes rapidly and the current coverage is close to the target value. The conformal comparison experiments partly address this issue by also evaluating a 500-instance windowed AdaPI factor generator, showing that the calibration window can materially change the relative ranking of AdaPI and conformal methods. However, this single window size is not intended to be a complete study of local calibration; a broader investigation of window lengths, adaptive window selection, and other local coverage estimators remains open. Fourth, some of the evaluation conclusions depend on the weighting used in CING. We chose a coverage-heavy weighting because coverage is the primary calibration objective in many PI applications, but different application domains may prefer different tradeoffs between calibration and interval width.

Finally, although the experimental suite includes both real-world and synthetic streams, it cannot represent the full diversity of streaming regression problems. Some datasets exhibit idiosyncratic behaviour, and comparisons with online conformal methods were necessarily restricted to settings where those methods could be run with a common base learner and compatible feature types. These constraints should be kept in mind when generalizing the conclusions beyond the studied benchmarks.

Table 6 Coverage (%) and NMPIW (%) comparison for ensemble quantile and MVE-based prediction interval methods at 95% confidence level

Method	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW
EQ	89.33	41.71	79.85	**62.77**	91.18	53.49	78.39	**22.94**
EQ-Logarithm	94.49	63.50	93.81	90.37	94.53	62.25	93.78	35.67
EQ-HLI	93.45	48.65	92.24	81.79	93.68	57.80	92.29	29.56
EQ-Linear	94.23	49.74	93.12	84.09	94.41	59.34	93.08	30.42
MVE	95.62	31.90	96.32	88.96	93.20	**39.92**	96.54	45.82
MVE-Logarithm	95.30	30.92	95.71	86.27	**94.62**	44.56	96.34	44.50
MVE-HLI	95.36	32.34	95.82	86.97	93.89	42.42	96.35	44.50
MVE-Linear	**95.10**	**30.69**	**95.26**	84.73	94.48	44.15	**95.90**	41.85
Dataset	**Abalone**		**MetroTraffic**		**California Housing**		**Superconductivity**	

Bold values indicate the best performance for each dataset

9 Conclusions

This paper revisited AdaPI as a general strategy for online post-calibration of streaming prediction intervals. Building on MVE, we studied three adaptive scaling rules and evaluated them on a broad set of real-world and synthetic regression streams. The results show that simple coverage-driven rescaling can often move MVE closer to the desired confidence level while maintaining competitive interval width. Across many 95% confidence level experiments, especially under the default 0.8/0.2 CING weighting, the linear variant provides the strongest empirical balance among the three adaptive rules. However, this advantage is conditional on the target confidence level and evaluation weighting.

The experiments also highlight several broader points. First, PI quality in streams depends strongly on the base regressor, so calibration and point prediction should not be studied in isolation. Second, adaptation is particularly useful under drift, where interval width must respond quickly to changing predictive difficulty. Third, comparisons with online conformal methods suggest that AdaPI is a practical alternative when the goal is lightweight, model agnostic, fully incremental calibration rather than formal sequential guarantees.

9.1 Future work

Several directions remain open for future work:

- developing additional fully incremental PI constructions beyond the MVE and ensemble quantile instantiations studied here;
- investigating alternative scaling rules, including more aggressive or data-dependent adaptation strategies;
- systematically studying local calibration choices, including window size, adaptive window selection, and alternatives to fixed-window coverage estimates;
- studying tighter integration between adaptive intervals and drift detection or anomaly detection modules; and
- extending the general post-calibration idea to other predictive settings beyond regression.

Author Contributions Y.S. implemented the algorithms, conducted the experiments, and drafted the manuscript. All authors contributed to refining the ideas, designing the experiments, and reviewing the manuscript.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions

Data Availability No datasets were generated or analysed during the current study.

Declarations

Conflict of interest The authors declare no conflict of interest.

Appendix A: full raw results

We present the full raw results in appendix (Tables 7, 8, 9, 10, 11, 12, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28 and 13).

Table 7 Coverage (%) and NMPIW (%) results: confidence level $\mathcal{L}$ = 80% and limit $\mathcal{M}$ = 0.05 (Part I)

	ARF-Reg		SOKNL		KNN	
	coverage	NMPIW	coverage	NMPIW	coverage	NMPIW
Logarithm	85.83	20.99	85.79	19.17	85.15	20.95
HLI	85.63	20.84	85.63	18.98	85.19	21.05
Linear	83.2	19.01	**83.0**	**17.45**	83.28	19.57
MVE	87.98	22.87	88.13	20.86	86.3	22.11
Abalone						
Logarithm	84.2	24.0	91.69	14.11	93.38	13.9
HLI	84.14	23.78	89.94	12.67	91.2	12.06
Linear	**81.53**	22.76	86.74	11.12	88.31	**10.24**
MVE	89.5	26.51	94.95	17.65	96.92	19.84
CPU Activity						
Logarithm	87.84	7.17	99.66	0.68	99.45	0.91
HLI	87.24	6.65	99.51	0.29	98.16	0.52
Linear	**83.28**	6.27	99.12	**0.28**	96.82	0.5
MVE	89.24	8.66	99.99	4.43	99.99	4.46
Naval Propulsion Plant						
Logarithm	86.04	18.09	84.72	17.82	83.31	14.44
HLI	85.85	17.98	84.65	17.82	83.46	14.56
Linear	83.24	16.38	82.18	16.38	**81.75**	**13.79**
MVE	87.7	19.37	86.44	18.77	84.24	14.91
Ailerons						
Logarithm	92.77	16.26	92.71	15.41	92.85	13.78
HLI	92.24	15.29	92.31	14.59	92.21	12.97
Linear	**90.36**	13.23	90.57	12.7	90.37	**11.19**
MVE	94.33	19.3	94.09	18.04	94.39	16.34
Miami Housing						
Logarithm	91.5	2.68	92.93	2.27	90.86	2.42
HLI	91.07	2.55	92.12	2.07	90.44	2.32
Linear	89.13	2.18	90.2	**1.74**	**88.3**	2.02
MVE	93.11	3.08	94.75	2.75	92.36	2.75

Table 7 continued

	ARF-Reg		SOKNL		KNN	
	coverage	NMPIW	coverage	NMPIW	coverage	NMPIW
NZ Energy Price						
Logarithm	92.11	3.42	92.39	3.4	91.05	3.98
HLI	90.61	3.14	90.96	3.11	89.78	3.71
Linear	87.52	2.68	88.08	**2.66**	**86.85**	3.19
MVE	94.62	18.12	94.82	4.14	93.38	4.69
Elevators						
Logarithm	80.7	18.73	79.67	17.48	**79.79**	28.97
HLI	80.62	18.75	78.73	**16.87**	76.6	26.38
Linear	80.25	18.51	79.04	17.13	78.22	27.74
MVE	80.83	18.88	78.92	16.91	75.49	25.56
Bike						
Logarithm	83.64	28.22	83.44	24.74	82.65	27.5
HLI	83.75	28.4	83.68	24.84	82.79	27.65
Linear	82.16	26.47	82.03	**23.33**	**81.26**	26.11
MVE	84.82	29.76	84.87	26.1	83.88	28.92
California Housing						
Logarithm	87.68	29.48	87.61	29.08	89.32	24.35
HLI	87.42	29.3	87.34	28.79	89.22	23.97
Linear	86.15	26.9	**86.08**	26.5	87.59	**21.49**
MVE	88.19	30.55	87.9	29.96	90.28	25.95
Superconductivity						

Table 8 Coverage (%) and NMPIW (%) results: confidence level $\mathcal{L}$ = 80% and limit $\mathcal{M}$ = 0.05 (Part II)

	ARF-Reg		SOKNL		KNN	
	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW
Logarithm	80.92	50.38	**79.87**	51.42	79.79	46.52
HLI	81.4	50.67	78.65	50.34	78.37	**45.36**
Linear	80.34	50.01	79.56	51.09	79.29	46.16
MVE	81.79	50.94	78.34	50.02	78.08	45.13
Health Insurance						
Logarithm	90.86	16.17	90.59	15.68	89.54	17.95
HLI	89.91	14.9	89.55	14.61	88.83	16.96
Linear	87.55	12.47	86.95	**12.43**	**86.37**	14.48
MVE	92.94	19.61	92.93	18.85	91.5	20.93
House8L						
Logarithm	81.02	59.92	80.76	58.0	80.06	70.95
HLI	81.35	60.33	80.87	58.07	79.34	69.9
Linear	80.49	59.3	80.35	**57.55**	**79.89**	70.72
MVE	81.66	60.66	80.95	58.17	79.21	69.74

Table 8 continued

	ARF-Reg		SOKNL		KNN	
	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW
MetroTraffic						
Logarithm	88.52	15.17	90.33	11.92	92.58	11.71
HLI	86.26	14.64	90.3	11.69	92.13	10.87
Linear	86.81	14.03	89.25	10.07	90.48	**9.04**
MVE	**86.13**	14.71	90.82	12.92	93.29	13.68
Diamonds						
Logarithm	89.04	11.03	89.79	6.52	89.8	6.39
HLI	88.56	10.6	89.09	6.14	89.23	6.04
Linear	**86.46**	8.96	86.86	5.16	87.18	**5.04**
MVE	90.49	12.45	91.46	7.6	91.24	7.36
Video Transcoding						
Logarithm	79.84	14.99	**79.97**	3.86	80.04	3.67
HLI	78.19	14.68	79.96	3.8	80.04	3.61
Linear	78.94	14.76	79.13	3.72	78.98	**3.5**
MVE	79.12	14.93	82.57	4.23	82.41	4.02
Wave Energy						
Logarithm	80.9	24.1	81.59	17.71	80.22	21.76
HLI	81.25	24.31	81.93	17.86	80.37	21.84
Linear	80.28	23.78	80.63	**17.32**	**80.07**	21.68
MVE	81.77	24.61	82.89	18.28	80.51	21.9
FriedmanGra						
Logarithm	80.73	24.09	81.49	17.7	80.2	21.76
HLI	81.04	24.28	81.84	17.84	80.29	21.81
Linear	80.2	23.81	80.6	**17.32**	**80.09**	21.71
MVE	81.5	24.57	82.73	18.25	80.41	21.87
FriedmanGsg						
Logarithm	**79.94**	19.36	81.93	**14.1**	79.54	18.67
HLI	80.21	19.48	82.23	14.21	78.63	18.19
Linear	79.52	19.17	80.52	13.62	78.92	18.33
MVE	80.73	19.72	83.54	14.7	78.72	18.22
FriedmanLea						
Logarithm	81.72	18.32	82.19	10.5	80.58	11.23
HLI	81.92	18.43	82.47	10.59	80.85	11.3
Linear	80.97	17.96	81.04	**10.17**	**80.16**	11.11
MVE	82.25	18.58	83.5	10.91	81.08	11.37
$Hyper_a$						

Table 9 Coverage (%) and NMPIW (%) results: confidence level $\mathcal{L}$ = 90% and limit $\mathcal{M}$ = 0.05 (Part I)

	ARF-Reg		SOKNL		KNN	
	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW
Logarithm	91.92	27.66	92.2	25.04	91.5	27.8
HLI	91.96	27.97	92.16	25.3	91.42	27.98
Linear	**90.84**	26.29	91.02	**23.69**	91.0	27.12
MVE	92.75	29.36	93.33	26.77	92.0	28.37
Abalone						
Logarithm	94.14	29.23	95.58	19.01	96.55	18.68
HLI	93.51	28.48	94.79	17.7	95.41	16.35
Linear	**91.5**	27.18	93.15	15.34	93.79	**14.0**
MVE	98.21	34.02	97.28	22.66	98.21	25.46
CPU Activity						
Logarithm	94.0	9.26	99.86	0.83	99.78	1.1
HLI	93.39	8.67	99.81	0.34	99.34	0.56
Linear	**91.34**	8.25	99.69	**0.33**	98.74	0.54
MVE	94.93	11.11	99.99	5.69	99.99	5.73
Naval Propulsion Plant						
Logarithm	92.08	23.97	91.9	23.34	90.95	18.91
HLI	92.08	24.13	91.91	23.51	90.99	19.04
Linear	91.02	22.64	90.78	22.02	**90.36**	**18.53**
MVE	92.62	24.86	92.44	24.1	91.17	19.14
Ailerons						
Logarithm	95.64	22.09	95.58	20.87	95.68	18.74
HLI	95.43	21.45	95.49	20.19	95.51	18.15
Linear	94.8	19.03	94.96	18.01	**94.77**	**16.21**
MVE	96.21	24.77	96.05	23.16	96.22	20.97
Miami Housing						
Logarithm	95.13	3.58	95.79	3.06	94.7	3.23
HLI	95.03	3.48	95.53	2.87	94.59	3.18
Linear	94.09	3.05	94.5	2.46	93.61	2.84
MVE	95.65	3.96	96.56	3.53	95.38	3.53
NZ Energy Price						
Logarithm	95.96	4.48	96.04	4.48	95.3	5.28
HLI	95.17	4.19	95.24	4.16	94.59	5.02
Linear	93.5	3.65	93.62	**3.62**	**92.9**	4.43
MVE	97.09	5.3	97.27	5.32	96.35	6.02
Elevators						
Logarithm	**89.29**	28.58	89.19	25.34	89.04	41.54
HLI	87.39	25.63	87.04	23.06	86.47	36.67
Linear	88.52	27.26	88.27	24.34	87.82	39.03
MVE	86.33	24.24	85.64	**21.7**	83.91	32.81

Table 9 continued

	ARF-Reg		SOKNL		KNN	
	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW
Bike						
Logarithm	90.13	39.03	90.33	33.43	**90.01**	37.73
HLI	89.61	37.94	90.08	33.08	89.44	36.8
Linear	89.96	38.49	90.16	**32.96**	89.79	37.23
MVE	89.78	38.2	90.28	33.5	89.61	37.12
California Housing						
Logarithm	93.52	37.81	93.43	37.42	93.96	32.53
HLI	93.43	37.63	93.31	37.14	93.75	32.06
Linear	92.62	34.78	**92.59**	34.25	93.1	**29.63**
MVE	93.87	39.21	93.62	38.46	94.04	33.31

Table 10 Coverage (%) and NMPIW (%) results: confidence level $\mathcal{L}$ = 90% and limit $\mathcal{M}$ = 0.05 (Part II)

	ARF-Reg		SOKNL		KNN	
	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW
Logarithm	92.72	60.69	91.14	62.52	90.05	57.69
HLI	92.67	60.71	91.44	62.99	90.18	57.88
Linear	91.02	58.85	90.39	61.86	**89.97**	**57.56**
MVE	95.71	65.38	92.45	64.2	90.28	57.93
Health Insurance						
Logarithm	94.05	22.16	94.19	21.26	93.23	24.6
HLI	93.75	21.43	93.84	20.38	93.08	24.32
Linear	92.5	18.82	92.43	**18.12**	**91.86**	21.81
MVE	95.02	25.17	95.37	24.2	94.0	26.87
House8L						
Logarithm	91.26	75.4	90.93	72.86	90.57	88.22
HLI	91.51	75.98	91.19	73.44	90.84	88.81
Linear	90.51	73.72	90.36	**71.59**	**90.19**	87.42
MVE	92.2	77.85	91.79	74.66	91.2	89.51
MetroTraffic						
Logarithm	93.7	21.35	93.67	18.8	93.97	17.43
HLI	92.11	20.01	92.71	17.05	93.96	16.92
Linear	92.84	19.98	93.08	17.02	93.65	**15.79**
MVE	**90.86**	18.87	92.41	16.58	94.07	17.56
Diamonds						
Logarithm	92.8	15.11	93.13	9.03	93.06	8.88
HLI	92.81	15.14	93.03	8.94	93.02	8.86
Linear	**91.8**	13.87	91.92	**7.95**	92.0	**7.95**
MVE	93.3	15.98	93.88	9.75	93.59	9.45

Table 10 continued

	ARF-Reg		SOKNL		KNN	
	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW
Video Transcoding						
Logarithm	91.12	17.98	**89.81**	5.03	89.65	4.85
HLI	91.31	18.01	89.44	4.89	88.91	4.68
Linear	90.4	17.63	89.48	4.85	89.16	**4.67**
MVE	93.65	19.17	90.84	5.43	90.13	5.16
Wave Energy						
Logarithm	90.45	30.85	90.85	22.62	90.03	27.91
HLI	90.65	31.06	91.04	22.77	90.14	28.0
Linear	90.11	30.55	90.31	**22.22**	**89.98**	27.88
MVE	91.15	31.59	91.86	23.46	90.26	28.11
FriedmanGra						
Logarithm	90.39	30.84	90.57	22.67	90.07	27.94
HLI	90.58	31.03	90.75	22.8	90.09	27.96
Linear	90.1	30.58	90.2	**22.36**	**90.02**	27.89
MVE	91.03	31.53	91.49	23.43	90.21	28.07
FriedmanGsg						
Logarithm	89.88	25.03	90.42	18.06	89.32	24.45
HLI	89.72	24.91	90.57	18.15	88.14	23.45
Linear	89.63	24.84	**89.87**	**17.71**	88.83	24.01
MVE	90.18	25.32	91.56	18.86	88.09	23.38
FriedmanLea						
Logarithm	90.43	23.46	90.13	13.67	89.79	14.76
HLI	90.57	23.58	90.22	13.71	89.44	14.56
Linear	90.12	23.18	**89.98**	**13.57**	89.66	14.69
MVE	90.89	23.84	90.68	14.0	89.49	14.59
Hyper$_a$						

Table 11 Coverage (%) and NMPIW (%) results: confidence level $\mathcal{L}$ = 95% and limit $\mathcal{M}$ = 0.05 (Part I)

	ARF-Reg		SOKNL		KNN	
	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW
Logarithm	94.88	35.09	95.26	30.86	95.26	34.32
HLI	94.72	36.13	95.36	32.29	**94.98**	35.18
Linear	94.82	35.18	95.1	**30.68**	95.16	34.37
MVE	94.68	34.98	95.62	31.9	94.96	33.81
Abalone						
Logarithm	97.74	33.71	97.56	23.63	97.96	24.32
HLI	97.19	33.32	97.19	23.33	97.46	22.9
Linear	**96.07**	31.23	96.33	20.61	96.59	**19.12**
MVE	99.71	40.54	98.44	27.0	98.68	30.34
CPU Activity						
Logarithm	96.47	11.34	99.93	0.96	99.9	1.28
HLI	96.32	10.93	99.93	**0.38**	99.72	0.6
Linear	**95.42**	10.48	99.87	**0.38**	99.5	0.58
MVE	97.1	13.24	99.99	6.78	99.99	6.82
Naval Propulsion Plant						
Logarithm	95.19	30.32	95.38	28.95	94.92	23.4
HLI	95.05	30.42	95.37	29.19	94.71	23.3
Linear	**95.04**	29.97	95.08	28.29	94.88	23.35
MVE	94.8	29.62	95.16	28.71	94.41	**22.8**
Ailerons						
Logarithm	97.28	29.54	97.37	27.33	97.38	24.51
HLI	96.96	28.41	97.02	26.81	97.06	24.0
Linear	97.04	26.91	97.18	24.92	97.07	**22.37**
MVE	97.03	29.52	**96.91**	27.59	97.17	24.99
Miami Housing						
Logarithm	96.86	4.53	97.27	3.87	96.79	4.0
HLI	96.88	4.53	97.2	3.82	96.84	4.02
Linear	96.56	4.15	96.7	**3.44**	**96.49**	3.71
MVE	96.95	4.72	97.45	4.2	96.99	4.2
NZ Energy Price						
Logarithm	97.68	5.61	97.82	5.58	97.37	6.54
HLI	97.39	5.39	97.48	5.33	97.11	6.35
Linear	96.55	4.79	96.57	**4.73**	**96.25**	5.76
MVE	97.98	6.31	98.17	6.34	97.66	7.17
Elevators						
Logarithm	93.85	39.56	**93.97**	33.39	93.76	53.74
HLI	92.58	35.61	92.69	30.48	92.33	49.38
Linear	93.35	37.65	93.53	32.15	93.12	51.75
MVE	89.53	28.88	89.56	**25.86**	88.16	39.1

Table 11 continued

	ARF-Reg		SOKNL		KNN	
	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW
Bike						
Logarithm	94.55	51.58	**94.61**	44.54	94.53	49.53
HLI	93.71	49.28	93.9	42.42	93.65	47.29
Linear	94.37	51.46	94.5	44.27	94.37	49.32
MVE	92.65	45.52	93.2	**39.92**	92.75	44.23
California Housing						
Logarithm	96.31	46.24	96.33	44.43	96.54	41.04
HLI	96.34	45.35	96.34	44.46	96.31	39.84
Linear	95.89	42.68	**95.88**	41.71	96.12	**37.64**
MVE	96.55	46.72	96.54	45.82	96.12	39.69
Superconductivity						

Table 12 Coverage (%) and NMPIW (%) results: confidence level $\mathcal{L}$ = 95% and limit $\mathcal{M}$ = 0.05 (Part II)

	ARF-Reg		SOKNL		KNN	
	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW
Logarithm	97.1	68.82	96.46	70.45	95.41	67.74
HLI	96.83	68.25	96.37	70.48	95.53	68.56
Linear	95.82	**65.89**	95.51	68.86	**95.15**	67.29
MVE	98.85	77.9	98.63	76.5	95.75	69.03
Health Insurance						
Logarithm	95.94	28.52	96.18	27.06	95.37	31.43
HLI	95.97	28.83	96.16	27.21	95.43	31.93
Linear	95.43	26.69	95.54	**25.02**	**95.15**	30.89
MVE	96.32	29.99	96.61	28.83	95.46	32.01
House8L						
Logarithm	95.62	90.01	95.7	86.17	95.77	102.25
HLI	95.75	90.74	95.82	86.89	95.88	102.78
Linear	**95.22**	88.51	95.24	**84.67**	95.23	100.54
MVE	96.22	92.77	96.32	88.96	96.88	106.66
MetroTraffic						
Logarithm	96.23	29.22	96.08	27.47	96.12	26.89
HLI	95.35	27.21	95.16	25.11	95.55	23.57
Linear	95.7	28.28	95.43	26.35	95.8	25.04
MVE	93.06	22.49	93.43	**19.75**	**94.89**	20.92
Diamonds						
Logarithm	95.14	19.3	95.33	11.53	95.17	11.59
HLI	95.1	19.18	95.37	11.59	95.09	11.42
Linear	95.09	19.18	95.12	11.3	95.12	11.54
MVE	94.96	19.04	95.36	11.62	**94.98**	**11.26**

Table 12 continued

	ARF-Reg		SOKNL		KNN	
	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW
Video Transcoding						
Logarithm	96.33	20.62	94.73	6.16	94.6	5.92
HLI	96.31	20.55	94.28	5.98	93.94	**5.7**
Linear	95.53	19.96	94.65	6.01	94.47	5.78
MVE	98.28	22.84	**94.75**	6.47	94.27	6.14
Wave Energy						
Logarithm	95.25	36.74	95.3	26.98	95.01	33.38
HLI	95.38	36.98	95.4	27.13	95.01	33.42
Linear	95.09	36.4	95.11	**26.7**	**95.0**	33.36
MVE	95.73	37.64	95.92	27.96	95.09	33.5
FriedmanGra						
Logarithm	95.16	36.76	95.15	27.12	95.02	33.48
HLI	95.27	36.97	95.16	27.17	94.93	33.38
Linear	95.05	36.5	95.05	**26.94**	**95.0**	33.47
MVE	95.59	37.57	95.62	27.92	**95.0**	33.44
FriedmanGsg						
Logarithm	94.85	29.99	94.78	22.01	94.2	29.72
HLI	94.59	29.76	94.55	**21.79**	93.35	28.46
Linear	94.79	29.88	94.64	21.84	94.02	29.43
MVE	94.86	30.17	**95.04**	22.48	92.93	27.86
FriedmanLea						
Logarithm	94.9	28.18	94.68	17.11	94.64	18.27
HLI	94.86	28.13	94.18	**16.63**	94.04	17.7
Linear	94.84	28.09	94.63	17.03	94.58	18.21
MVE	**95.03**	28.41	94.17	16.68	93.7	17.39
Hyper_a						

Table 13 Coverage (%) and NMPIW (%) results: confidence level $\mathcal{L}$ = 99% and limit $\mathcal{M}$ = 0.05 (Part I)

	ARF-Reg		SOKNL		KNN	
	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW
Logarithm	98.49	52.77	98.55	45.64	98.55	51.39
HLI	98.49	68.58	98.55	61.34	98.51	67.03
Linear	98.65	63.65	**98.71**	55.68	98.69	59.91
MVE	97.57	45.97	98.03	**41.93**	97.61	44.43
Abalone						
Logarithm	99.68	40.9	99.32	**33.19**	99.08	40.82
HLI	99.52	48.89	99.28	41.5	99.1	51.24
Linear	99.32	42.76	99.1	36.35	**99.06**	44.97
MVE	99.93	53.28	99.57	35.48	**99.06**	39.88
CPU Activity						
Logarithm	98.74	17.14	99.97	1.45	99.97	2.12
HLI	98.72	18.61	99.98	0.48	99.95	0.74
Linear	**98.84**	18.45	99.98	**0.46**	99.92	0.69
MVE	98.8	17.41	99.99	8.91	99.99	8.97
Naval Propulsion Plant						
Logarithm	98.46	49.83	98.51	47.17	98.61	34.36
HLI	98.46	52.22	98.51	49.35	98.6	36.46
Linear	98.63	53.24	98.68	50.48	**98.81**	36.53
MVE	97.0	38.93	97.38	37.73	97.58	**29.97**
Ailerons						
Logarithm	99.17	49.33	99.2	43.64	99.17	40.22
HLI	99.17	50.1	99.2	45.46	99.18	42.33
Linear	99.2	49.69	99.24	44.79	99.23	41.7
MVE	98.04	38.79	98.13	36.26	**98.16**	**32.85**
Miami Housing						
Logarithm	98.77	7.4	98.78	6.47	98.87	6.22
HLI	98.77	7.53	98.77	6.61	98.87	6.31
Linear	98.87	7.95	98.88	6.95	**98.94**	6.66
MVE	98.12	6.2	98.39	**5.52**	98.41	**5.52**
NZ Energy Price						
Logarithm	99.12	8.81	99.17	8.69	99.11	8.95
HLI	99.11	8.86	99.17	8.69	99.08	10.4
Linear	99.11	8.95	99.14	8.77	99.09	10.6
MVE	98.88	**8.3**	**98.96**	8.33	98.65	9.42
Elevators						
Logarithm	98.29	57.64	98.31	49.07	98.27	75.03
HLI	98.29	58.16	98.34	49.44	98.31	75.61
Linear	98.42	59.09	**98.46**	50.43	98.42	77.27
MVE	93.49	37.96	94.33	**33.99**	93.14	51.38

Table 13 continued

	ARF-Reg		SOKNL		KNN	
	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW
Bike						
Logarithm	98.42	78.05	98.43	69.09	98.46	75.96
HLI	98.45	84.28	98.45	75.71	98.49	82.17
Linear	98.61	83.81	98.61	75.27	**98.68**	82.81
MVE	96.06	59.82	96.29	**52.46**	96.17	58.13
California Housing						
Logarithm	98.82	62.04	98.82	59.83	98.94	56.19
HLI	98.81	62.15	98.83	60.76	98.96	56.99
Linear	98.87	61.36	98.86	60.18	**98.99**	55.58
MVE	98.79	61.4	98.86	60.22	98.36	**52.17**
Superconductivity						

Table 14 Coverage (%) and NMPIW (%) results: confidence level $\mathcal{L}$ = 99% and limit $\mathcal{M}$ = 0.05 (Part II)

	ARF-Reg		SOKNL		KNN	
	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW
Logarithm	99.49	89.87	99.54	86.1	99.17	87.72
HLI	99.42	91.65	99.45	87.95	99.2	91.63
Linear	99.21	85.4	99.22	**82.77**	**99.07**	89.01
MVE	99.79	102.38	99.82	100.54	99.43	90.72
Health Insurance						
Logarithm	98.46	49.63	98.5	46.28	98.42	54.77
HLI	98.47	51.54	98.52	48.36	98.44	56.91
Linear	98.62	53.96	**98.67**	50.35	98.6	58.67
MVE	97.72	39.41	97.91	**37.89**	97.31	42.07
House8L						
Logarithm	99.31	113.16	99.3	109.23	99.35	126.08
HLI	99.27	114.77	99.27	111.11	99.31	127.96
Linear	99.09	111.63	**99.07**	**108.25**	99.12	124.21
MVE	99.76	121.92	99.75	116.91	99.88	140.17
MetroTraffic						
Logarithm	**98.98**	43.63	98.84	43.52	98.51	56.32
HLI	98.86	45.68	98.79	44.77	98.62	56.0
Linear	98.82	48.64	98.69	47.69	98.48	57.06
MVE	95.51	29.56	94.96	**25.96**	95.96	27.49
Diamonds						
Logarithm	98.37	35.24	98.29	20.9	98.33	21.05
HLI	98.41	35.47	98.33	21.06	98.36	21.28
Linear	**98.56**	37.06	98.49	22.09	98.5	22.18
MVE	96.91	25.03	97.07	15.27	96.84	**14.8**

Table 14 continued

	ARF-Reg		SOKNL		KNN	
	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW
Video Transcoding						
Logarithm	99.47	25.5	98.65	8.87	98.62	8.24
HLI	99.39	25.58	98.65	9.0	98.63	8.34
Linear	99.19	24.72	**98.83**	9.19	98.81	8.46
MVE	99.93	30.02	98.05	8.51	98.15	**8.07**
Wave Energy						
Logarithm	99.04	48.05	98.91	**36.25**	98.95	44.19
HLI	99.06	48.59	98.89	36.53	98.93	44.42
Linear	99.03	48.04	**98.98**	36.64	**98.98**	44.47
MVE	99.2	49.47	98.92	36.74	98.93	44.03
FriedmanGra						
Logarithm	99.02	48.21	98.88	**36.44**	98.93	44.33
HLI	**99.01**	48.71	98.86	36.74	98.9	44.53
Linear	99.02	48.29	98.97	36.93	98.98	44.86
MVE	99.13	49.38	98.83	36.69	98.85	43.95
FriedmanGsg						
Logarithm	98.8	40.11	98.56	32.07	98.41	40.86
HLI	98.78	40.35	98.58	32.56	98.43	41.25
Linear	**98.94**	41.18	98.74	33.89	98.63	42.37
MVE	98.61	39.64	97.92	**29.54**	97.31	36.62
FriedmanLea						
Logarithm	98.84	38.07	98.61	24.64	98.63	25.56
HLI	98.8	38.0	98.61	24.73	98.63	25.65
Linear	**98.93**	38.48	98.81	25.47	98.82	26.39
MVE	98.65	37.34	98.05	**8.51**	97.67	22.85
Hyper$_a$						

Table 15 Coverage (%) and NMPIW (%) results: confidence level $\mathcal{L}$ = 80% and limit $\mathcal{M}$ = 0.1 (Part I)

	ARF-Reg		SOKNL		KNN	
	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW
Logarithm	85.94	21.05	85.92	19.22	85.21	20.99
HLI	85.73	20.89	85.73	19.03	85.23	21.07
Linear	83.32	19.11	**83.14**	**17.46**	83.34	19.63
MVE	87.98	22.87	88.13	20.86	86.3	22.11
Abalone						
Logarithm	84.34	24.05	91.85	14.22	93.59	14.13
HLI	84.22	23.82	90.16	12.8	91.42	12.18
Linear	**81.58**	22.77	87.01	11.18	88.61	**10.38**
MVE	89.5	26.51	94.95	17.65	96.92	19.84

Table 15 continued

	ARF-Reg		SOKNL		KNN	
	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW
CPU Activity						
Logarithm	87.87	7.22	99.82	0.78	99.53	0.95
HLI	87.34	6.66	99.89	**0.45**	98.64	0.6
Linear	**83.46**	6.29	99.84	**0.45**	97.74	0.58
MVE	89.24	8.66	99.99	4.43	99.99	4.46
Naval Propulsion Plant						
Logarithm	86.09	18.14	84.82	17.86	83.37	14.46
HLI	85.94	18.02	84.73	17.85	83.49	14.58
Linear	83.36	16.46	82.28	16.46	**81.83**	**13.81**
MVE	87.7	19.37	86.44	18.77	84.24	14.91
Ailerons						
Logarithm	92.86	16.37	92.74	15.53	92.93	13.87
HLI	92.33	15.41	92.37	14.7	92.32	13.08
Linear	**90.55**	13.37	90.7	12.85	90.59	**11.32**
MVE	94.33	19.3	94.09	18.04	94.39	16.34
Miami Housing						
Logarithm	91.55	2.7	93.05	2.29	90.95	2.43
HLI	91.13	2.57	92.23	2.09	90.51	2.33
Linear	89.25	2.21	90.38	**1.77**	**88.44**	2.04
MVE	93.11	3.08	94.75	2.75	92.36	2.75
NZ Energy Price						
Logarithm	92.24	3.45	92.48	3.43	91.13	4.01
HLI	90.78	3.17	91.09	3.15	89.91	3.74
Linear	87.86	2.72	88.37	**2.7**	**87.12**	3.23
MVE	94.62	4.13	94.82	4.14	93.38	4.69
Elevators						
Logarithm	80.71	18.75	79.66	17.47	**79.79**	28.97
HLI	80.63	18.76	78.74	**16.87**	76.6	26.38
Linear	80.26	18.5	79.03	17.12	78.17	27.67
MVE	80.83	18.88	78.92	16.91	75.49	25.56
Bike						
Logarithm	83.68	28.27	83.5	24.79	82.69	27.55
HLI	83.78	28.44	83.72	24.88	82.81	27.71
Linear	82.22	26.53	82.06	**23.39**	**81.3**	26.16
MVE	84.82	29.76	84.87	26.1	83.88	28.92
California Housing						
Logarithm	87.73	29.54	87.65	29.14	89.36	24.43
HLI	87.45	29.36	87.39	28.85	89.28	24.05
Linear	86.28	27.06	**86.22**	26.67	87.72	**21.67**
MVE	88.19	30.55	87.9	29.96	90.28	25.95
Superconductivity						

Table 16 Coverage (%) and NMPIW (%) results: confidence level $\mathcal{L}$ = 80% and limit $\mathcal{M}$ = 0.1 (Part II)

	ARF-Reg		SOKNL		KNN	
	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW
Logarithm	80.94	50.4	**79.87**	51.42	79.79	46.52
HLI	81.42	50.68	78.65	50.34	78.37	45.36
Linear	80.36	50.03	79.53	51.07	79.27	46.14
MVE	81.79	50.94	78.34	50.02	78.08	**45.13**
Health Insurance						
Logarithm	90.94	16.31	90.73	15.8	89.62	18.07
HLI	90.03	15.05	89.66	14.74	88.92	17.08
Linear	87.76	12.64	87.21	**12.58**	**86.57**	14.66
MVE	92.94	19.61	92.93	18.85	91.5	20.93
House8L						
Logarithm	81.05	59.94	80.77	58.01	**80.06**	70.95
HLI	81.36	60.34	80.87	58.07	79.34	69.9
Linear	80.51	59.32	80.36	**57.59**	79.89	70.69
MVE	81.66	60.66	80.95	58.17	79.21	69.74
MetroTraffic						
Logarithm	88.53	15.2	90.35	11.97	92.62	11.8
HLI	**86.26**	14.65	90.32	11.75	92.2	10.98
Linear	86.79	14.08	89.32	10.18	90.62	**9.21**
MVE	86.13	14.71	90.82	12.92	93.29	13.68
Diamonds						
Logarithm	89.11	11.09	89.86	6.57	89.88	6.43
HLI	88.64	10.67	89.18	6.19	89.32	6.09
Linear	**86.63**	9.07	87.06	5.23	87.36	**5.11**
MVE	90.49	12.45	91.46	7.6	91.24	7.36
Video Transcoding						
Logarithm	79.83	14.99	80.04	3.87	80.09	3.68
HLI	78.2	14.68	**80.0**	3.81	80.1	3.61
Linear	78.9	14.75	79.12	3.72	78.98	**3.5**
MVE	79.12	14.93	82.57	4.23	82.41	4.02
Wave Energy						
Logarithm	80.93	24.12	81.63	17.73	80.23	21.76
HLI	81.27	24.32	81.96	17.87	80.37	21.84
Linear	80.3	23.79	80.67	**17.32**	**80.08**	21.69
MVE	81.77	24.61	82.89	18.28	80.51	21.9
FriedmanGra						
Logarithm	80.76	24.1	81.53	17.72	80.2	21.77
HLI	81.06	24.29	81.86	17.85	80.29	21.81
Linear	80.21	23.82	80.62	**17.33**	**80.09**	21.71
MVE	81.5	24.57	82.73	18.25	80.41	21.87

Table 16 continued

	ARF-Reg		SOKNL		KNN	
	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW
FriedmanGsg						
Logarithm	**79.94**	19.36	81.98	14.12	79.54	18.67
HLI	80.23	19.49	82.27	14.22	78.63	18.19
Linear	79.51	19.17	80.56	**13.63**	78.9	18.32
MVE	80.73	19.72	83.54	14.7	78.72	18.22
FriedmanLea						
Logarithm	81.74	18.33	82.23	10.51	80.6	11.23
HLI	81.94	18.43	82.5	10.6	80.86	11.3
Linear	81.01	17.98	81.08	**10.19**	**80.17**	11.11
MVE	82.25	18.58	83.5	10.91	81.08	11.37
Hyper_a						

Table 17 Coverage (%) and NMPIW (%) results: confidence level $\mathcal{L}$ = 90% and limit $\mathcal{M}$ = 0.1 (Part I)

	ARF-Reg		SOKNL		KNN	
	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW
Logarithm	91.94	27.74	92.24	25.1	91.54	27.84
HLI	91.94	27.97	92.22	25.37	91.42	28.02
Linear	**90.88**	26.39	91.08	**23.76**	91.02	27.12
MVE	92.75	29.36	93.33	26.77	92.0	28.37
Abalone						
Logarithm	94.29	29.33	95.63	19.15	96.63	18.9
HLI	93.59	28.44	94.87	17.85	95.53	16.47
Linear	**91.59**	27.18	93.3	15.47	93.95	**14.16**
MVE	98.21	34.02	97.28	22.66	98.21	25.46
CPU Activity						
Logarithm	94.08	9.33	99.99	1.0	99.82	1.14
HLI	93.47	8.66	99.99	**0.57**	99.59	0.67
Linear	**91.39**	8.26	99.99	**0.57**	99.21	0.66
MVE	94.93	11.11	99.99	5.69	99.99	5.73
Naval Propulsion Plant						
Logarithm	92.1	24.02	91.93	23.39	90.97	18.93
HLI	92.07	24.15	91.91	23.52	90.99	19.05
Linear	91.05	22.69	90.81	22.09	**90.33**	**18.53**
MVE	92.62	24.86	92.44	24.1	91.17	19.14
Ailerons						
Logarithm	95.65	22.2	95.59	20.98	95.69	18.83
HLI	95.47	21.57	95.51	20.3	95.54	18.26
Linear	**94.83**	19.24	95.01	18.19	94.84	**16.34**
MVE	96.21	24.77	96.05	23.16	96.22	20.97

Table 17 continued

	ARF-Reg		SOKNL		KNN	
	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW
Miami Housing						
Logarithm	95.16	3.6	95.84	3.07	94.73	3.24
HLI	95.05	3.5	95.57	2.9	94.61	3.19
Linear	94.13	3.08	94.58	**2.49**	**93.67**	2.86
MVE	95.65	3.96	96.56	3.53	95.38	3.53
NZ Energy Price						
Logarithm	96.01	4.52	96.1	4.51	95.36	5.31
HLI	95.25	4.22	95.32	4.19	94.68	5.05
Linear	93.65	3.7	93.78	**3.67**	**93.02**	4.47
MVE	97.09	5.3	97.27	5.32	96.35	6.02
Elevators						
Logarithm	**89.29**	28.57	89.19	25.35	89.05	41.54
HLI	87.39	25.63	87.04	23.05	86.47	36.67
Linear	88.47	27.2	88.2	24.3	87.74	38.87
MVE	86.33	24.24	85.64	**21.7**	83.91	32.81
Bike						
Logarithm	90.13	39.0	90.33	33.46	**90.01**	37.73
HLI	89.62	37.94	90.09	33.1	89.45	36.82
Linear	89.96	38.41	90.16	**32.98**	89.77	37.22
MVE	89.78	38.2	90.28	33.5	89.61	37.12
California Housing						
Logarithm	93.54	37.88	93.44	37.49	93.97	32.58
HLI	93.46	37.69	93.34	37.19	93.76	32.11
Linear	92.7	34.95	**92.66**	34.44	93.16	**29.84**
MVE	93.87	39.21	93.62	38.46	94.04	33.31

Table 18 Coverage (%) and NMPIW (%) results: confidence level $\mathcal{L} = 90\%$ and limit $\mathcal{M} = 0.1$ (Part II)

	ARF-Reg		SOKNL		KNN	
	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW
Logarithm	92.79	60.8	91.17	62.57	**90.05**	57.69
HLI	92.71	60.77	91.45	63.01	90.18	57.89
Linear	91.06	58.89	90.39	61.88	89.97	**57.58**
MVE	95.71	65.38	92.45	64.2	90.28	57.93
Health Insurance						
Logarithm	94.13	22.26	94.26	21.36	93.25	24.7
HLI	93.79	21.55	93.89	20.49	93.12	24.41
Linear	92.57	18.97	92.53	**18.25**	**91.92**	21.9
MVE	95.02	25.17	95.37	24.2	94.0	26.87
House8L						
Logarithm	91.29	75.48	90.95	72.91	90.59	88.24
HLI	91.52	76.04	91.21	73.47	90.85	88.84
Linear	90.53	73.8	90.38	**71.6**	**90.19**	87.44
MVE	92.2	77.85	91.79	74.66	91.2	89.51
MetroTraffic						
Logarithm	93.7	21.37	93.67	18.82	93.97	17.46
HLI	92.1	20.01	92.71	17.06	93.97	16.95
Linear	92.8	20.02	93.07	17.04	93.67	15.87
MVE	**90.86**	18.87	92.41	**16.58**	94.07	17.56
Diamonds						
Logarithm	92.84	15.15	93.17	9.06	93.09	8.91
HLI	92.83	15.17	93.07	8.97	93.05	8.88
Linear	**91.84**	13.94	91.98	8.0	92.06	**7.99**
MVE	93.3	15.98	93.88	9.75	93.59	9.45
Video Transcoding						
Logarithm	91.18	18.0	**89.81**	5.04	89.65	4.86
HLI	91.35	18.03	89.46	4.89	88.93	4.68
Linear	90.42	17.63	89.47	4.84	89.13	**4.66**
MVE	93.65	19.17	90.84	5.43	90.13	5.16
Wave Energy						
Logarithm	90.47	30.86	90.87	22.63	90.03	27.91
HLI	90.67	31.06	91.06	22.77	90.14	28.0
Linear	90.12	30.55	90.33	**22.22**	**89.98**	27.88
MVE	91.15	31.59	91.86	23.46	90.26	28.11
FriedmanGra						
Logarithm	90.4	30.85	90.59	22.68	90.07	27.94
HLI	90.59	31.04	90.77	22.8	90.09	27.96
Linear	90.1	30.58	90.21	**22.36**	**90.01**	27.89
MVE	91.03	31.53	91.49	23.43	90.21	28.07

Table 18 continued

	ARF-Reg		SOKNL		KNN	
	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW
FriedmanGsg						
Logarithm	**89.88**	25.03	90.45	18.07	89.32	24.45
HLI	89.73	24.92	90.59	18.16	88.14	23.45
Linear	89.62	24.83	89.87	**17.71**	88.81	23.98
MVE	90.18	25.32	91.56	18.86	88.09	23.38
FriedmanLea						
Logarithm	90.44	23.47	90.14	13.67	89.79	14.76
HLI	90.58	23.59	90.23	13.72	89.45	14.56
Linear	90.13	23.19	**89.98**	**13.57**	89.66	14.69
MVE	90.89	23.84	90.68	14.0	89.49	14.59
Hyper$_a$						

Table 19 Coverage (%) and NMPIW (%) results: confidence level $\mathcal{L}$ = 99% and limit $\mathcal{M}$ = 0.1 (Part I)

	ARF-Reg		SOKNL		KNN	
	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW
Logarithm	98.49	52.77	98.55	45.65	98.55	51.4
HLI	98.49	68.58	98.55	61.36	98.51	67.04
Linear	98.65	63.04	**98.73**	55.35	98.71	60.0
MVE	97.57	45.97	98.03	**41.93**	97.61	44.43
Abalone						
Logarithm	99.69	41.56	99.33	33.32	99.08	40.88
HLI	99.54	45.38	99.28	35.74	99.08	45.51
Linear	99.34	41.11	99.08	**33.34**	**99.05**	41.89
MVE	99.93	53.28	99.57	35.48	99.06	39.88
CPU Activity						
Logarithm	98.74	17.13	99.99	1.57	99.97	1.74
HLI	98.72	17.69	99.99	**0.89**	99.97	0.93
Linear	**98.84**	17.73	99.99	**0.89**	99.97	0.92
MVE	98.8	17.41	99.99	8.91	99.99	8.97
Naval Propulsion Plant						
Logarithm	98.46	49.83	98.51	47.17	98.61	34.36
HLI	98.47	52.22	98.52	49.35	98.6	36.46
Linear	98.62	53.25	98.69	50.39	**98.8**	36.39
MVE	97.0	38.93	97.38	37.73	97.58	**29.97**
Ailerons						
Logarithm	**99.17**	49.36	99.2	43.68	**99.17**	40.26
HLI	99.18	50.18	99.2	45.25	99.19	42.33
Linear	99.2	49.71	99.25	44.5	99.22	41.68
MVE	98.04	38.79	98.13	36.26	98.16	**32.85**

Table 19 continued

	ARF-Reg		SOKNL		KNN	
	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW
Miami Housing						
Logarithm	98.77	7.4	98.78	6.47	98.87	6.22
HLI	98.77	7.53	98.77	6.61	98.87	6.31
Linear	98.87	7.92	98.87	6.92	**98.94**	6.61
MVE	98.12	6.2	98.39	**5.52**	98.41	**5.52**
NZ Energy Price						
Logarithm	99.12	8.81	99.18	41.34	99.07	10.39
HLI	99.12	8.86	99.17	8.69	99.08	10.4
Linear	99.122	8.95	99.15	8.78	99.07	10.64
MVE	98.88	**8.3**	**98.96**	8.33	98.65	9.42
Elevators						
Logarithm	98.29	57.64	98.31	49.07	98.27	75.03
HLI	98.29	58.16	98.34	49.44	98.31	75.61
Linear	98.38	58.81	**98.43**	50.19	98.39	76.86
MVE	93.49	37.96	94.33	**33.99**	93.14	51.38
Bike						
Logarithm	98.42	78.05	98.43	69.1	98.46	75.96
HLI	98.45	84.28	98.45	75.71	98.5	82.17
Linear	98.59	83.39	98.6	74.91	**98.66**	82.22
MVE	96.06	59.82	96.29	**52.46**	96.17	58.13
California Housing						
Logarithm	98.82	62.08	98.82	59.82	98.94	56.2
HLI	98.8	61.92	98.82	60.5	98.96	56.73
Linear	98.87	61.24	98.84	60.11	**98.98**	55.42
MVE	98.79	61.4	98.86	60.22	98.36	**52.17**
Superconductivity						

Table 20 Coverage (%) and NMPIW (%) results: confidence level $\mathcal{L}$ = 99% and limit $\mathcal{M}$ = 0.1 (Part II)

	ARF-Reg		SOKNL		KNN	
	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW
Logarithm	99.51	90.11	99.55	86.42	99.17	87.66
HLI	99.43	91.62	99.45	87.93	99.20	91.68
Linear	99.21	85.32	99.23	**82.79**	**99.07**	89.04
MVE	99.79	102.38	99.82	100.54	99.43	90.72
Health Insurance						
Logarithm	98.46	49.63	98.50	46.28	98.42	54.78
HLI	98.47	51.54	98.52	48.36	98.44	56.91
Linear	98.60	53.49	**98.66**	50.26	98.60	58.32
MVE	97.72	39.41	97.91	**37.89**	97.31	42.07
House8L						
Logarithm	99.32	113.31	99.30	109.22	99.36	126.40
HLI	99.28	115.06	99.28	111.12	99.32	128.12
Linear	**99.09**	111.73	**99.09**	**108.25**	99.12	124.06
MVE	99.76	121.92	99.75	116.91	99.88	140.17
MetroTraffic						
Logarithm	**98.98**	43.63	98.84	43.52	98.51	56.32
HLI	98.86	45.68	98.79	44.77	98.62	56.00
Linear	98.80	48.51	98.66	47.56	98.44	56.56
MVE	95.51	29.56	94.96	**25.96**	95.96	27.49
Diamonds						
Logarithm	98.37	35.24	98.29	20.90	98.33	21.05
HLI	98.41	35.47	98.33	21.06	98.36	21.28
Linear	**98.55**	36.85	98.47	21.99	98.49	22.06
MVE	96.91	25.03	97.07	15.27	96.84	**14.80**
Video Transcoding						
Logarithm	99.49	25.61	98.65	8.88	98.62	8.25
HLI	99.40	25.67	98.65	8.79	98.62	8.16
Linear	99.20	24.75	**98.82**	9.03	98.80	8.34
MVE	99.93	30.02	98.05	8.51	98.15	**8.07**
Wave Energy						
Logarithm	99.05	48.07	98.91	**36.24**	98.95	44.19
HLI	99.06	48.56	98.89	36.47	98.93	44.38
Linear	99.03	47.96	98.97	36.61	**98.98**	44.42
MVE	99.20	49.47	98.92	36.74	98.93	44.03
FriedmanGra						
Logarithm	99.02	48.25	98.88	**36.42**	98.93	44.33
HLI	**99.01**	48.66	98.86	36.69	98.90	44.49
Linear	99.02	48.23	98.96	36.89	98.98	44.81
MVE	99.13	49.38	98.83	36.69	98.85	43.95

Table 20 continued

	ARF-Reg		SOKNL		KNN	
	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW
FriedmanGsg						
Logarithm	98.80	40.10	98.56	32.05	98.41	40.86
HLI	98.78	40.32	98.58	32.51	98.43	41.21
Linear	**98.94**	41.09	98.72	33.70	98.62	42.27
MVE	98.61	39.64	97.92	**29.54**	97.31	36.62
FriedmanLea						
Logarithm	98.84	38.07	98.61	24.64	98.63	25.56
HLI	98.80	38.01	98.62	24.73	98.63	25.65
Linear	**98.92**	38.47	98.80	25.44	98.81	26.34
MVE	98.65	37.34	98.05	**8.51**	97.67	22.85
$Hyper_a$						

Table 21 Coverage (%) and NMPIW (%) results: confidence level $\mathcal{L}$ = 80% and limit $\mathcal{M}$ = 0.5 (Part I)

	ARF-Reg		SOKNL		KNN	
	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW
Logarithm	86.82	21.66	86.64	19.77	85.73	21.4
HLI	86.48	21.49	86.38	19.59	85.75	21.38
Linear	84.61	19.97	84.45	**18.19**	**84.29**	20.25
MVE	87.98	22.87	88.13	20.86	86.3	22.11
Abalone						
Logarithm	85.75	24.72	93.24	15.47	95.14	16.03
HLI	85.23	24.25	91.93	14.03	93.64	13.89
Linear	**82.6**	23.22	89.92	**12.68**	92.03	12.71
MVE	89.5	26.51	94.95	17.65	96.92	19.84
CPU Activity						
Logarithm	88.14	7.7	99.99	2.4	99.99	2.42
HLI	87.87	7.19	99.99	**2.22**	99.99	2.23
Linear	**85.87**	6.58	99.99	**2.22**	99.99	2.23
MVE	89.24	8.66	99.99	4.43	99.99	4.46
Naval Propulsion Plant						
Logarithm	86.71	18.62	85.45	18.25	83.73	14.65
HLI	86.57	18.48	85.31	18.17	83.64	14.69
Linear	84.84	17.26	83.46	17.16	**82.52**	**14.09**
MVE	87.7	19.37	86.44	18.77	84.24	14.91
Ailerons						
Logarithm	93.45	17.48	93.28	16.49	93.53	14.79
HLI	93.01	16.68	92.96	15.8	93.12	14.07
Linear	**91.98**	15.17	92.0	14.4	92.1	**12.77**
MVE	94.33	19.3	94.09	18.04	94.39	16.34

Table 21 continued

	ARF-Reg		SOKNL		KNN	
	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW
Miami Housing						
Logarithm	92.18	2.85	93.83	2.46	91.49	2.55
HLI	91.82	2.74	93.18	2.31	91.21	2.47
Linear	90.47	2.49	91.93	**2.08**	**89.7**	2.26
MVE	93.11	3.08	94.75	2.75	92.36	2.75
NZ Energy Price						
Logarithm	93.36	3.7	93.53	3.7	92.21	4.27
HLI	92.36	3.48	92.57	3.47	91.3	4.06
Linear	90.63	3.15	90.97	**3.14**	**89.53**	3.69
MVE	94.62	4.13	94.82	4.14	93.38	4.69
Elevators						
Logarithm	80.78	18.82	79.67	17.47	**79.79**	28.97
HLI	80.68	18.8	78.81	**16.88**	76.6	26.39
Linear	80.39	18.57	78.88	17.04	77.52	27.15
MVE	80.83	18.88	78.92	16.91	75.49	25.56
Bike						
Logarithm	84.1	28.81	84.05	25.23	83.06	28.03
HLI	84.12	28.77	84.09	25.16	83.11	28.02
Linear	82.84	27.3	82.78	**23.93**	**81.91**	26.69
MVE	84.82	29.76	84.87	26.1	83.88	28.92
California Housing						
Logarithm	87.97	30.03	87.86	29.6	89.76	25.03
HLI	87.76	29.86	**87.6**	29.32	89.71	24.8
Linear	87.12	28.49	87.07	28.0	88.75	**23.24**
MVE	88.19	30.55	87.9	29.96	90.28	25.95
Superconductivity						

Table 22 Coverage (%) and NMPIW (%) results: confidence level $\mathcal{L}$ = 80% and limit $\mathcal{M}$ = 0.5 (Part II)

	ARF-Reg		SOKNL		KNN	
	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW
Logarithm	81.19	50.56	**79.87**	51.43	79.79	46.52
HLI	81.53	50.77	78.65	50.34	78.38	45.36
Linear	80.57	50.15	79.35	50.89	79.02	45.96
MVE	81.79	50.94	78.34	50.02	78.08	**45.13**
Health Insurance						
Logarithm	91.81	17.57	91.65	16.94	90.42	19.13
HLI	91.05	16.49	90.9	15.97	89.83	18.33
Linear	89.78	14.75	89.3	**14.38**	**88.41**	16.49
MVE	92.94	19.61	92.93	18.85	91.5	20.93
House8L						
Logarithm	81.23	60.17	80.88	58.12	**80.06**	70.95
HLI	81.46	60.43	80.89	58.1	79.34	69.89
Linear	80.73	59.56	80.5	**57.72**	79.77	70.51
MVE	81.66	60.66	80.95	58.17	79.21	69.74
MetroTraffic						
Logarithm	88.55	15.41	90.55	12.37	92.98	12.58
HLI	86.32	14.74	90.53	12.23	92.73	12.0
Linear	86.48	14.47	89.95	11.21	91.9	**10.69**
MVE	**86.13**	14.71	90.82	12.92	93.29	13.68
Diamonds						
Logarithm	89.68	11.65	90.54	6.97	90.44	6.81
HLI	89.33	11.32	90.02	6.67	90.05	6.55
Linear	**88.06**	10.18	88.76	5.97	88.87	**5.85**
MVE	90.49	12.45	91.46	7.6	91.24	7.36
Video Transcoding						
Logarithm	**79.84**	15.01	80.74	3.96	80.79	3.76
HLI	78.4	14.73	80.56	3.89	80.61	**3.7**
Linear	78.47	14.69	79.3	3.77	79.3	3.56
MVE	79.12	14.93	82.57	4.23	82.41	4.02
Wave Energy						
Logarithm	81.19	24.28	82.02	17.9	80.3	21.8
HLI	81.43	24.41	82.22	17.98	80.41	21.86
Linear	80.51	23.9	81.0	**17.47**	**80.12**	21.71
MVE	81.77	24.61	82.89	18.28	80.51	21.9
FriedmanGra						
Logarithm	80.99	24.25	81.91	17.89	80.23	21.79
HLI	81.19	24.37	82.1	17.97	80.33	21.83
Linear	80.38	23.91	80.94	**17.47**	**80.11**	21.72
MVE	81.5	24.57	82.73	18.25	80.41	21.87

Table 22 continued

	ARF-Reg		SOKNL		KNN	
	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW
FriedmanGsg						
Logarithm	**80.15**	19.47	82.5	14.31	79.54	18.67
HLI	80.38	19.56	82.64	14.36	78.67	18.2
Linear	79.61	19.2	81.12	**13.82**	78.74	18.24
MVE	80.73	19.72	83.54	14.7	78.72	18.22
FriedmanLea						
Logarithm	81.87	18.4	82.65	10.64	80.75	11.27
HLI	82.05	18.48	82.78	10.68	80.93	11.32
Linear	81.33	18.14	81.55	**10.32**	**80.35**	11.16
MVE	82.25	18.58	83.5	10.91	81.08	11.37
$Hyper_a$						

Table 23 Coverage (%) and NMPIW (%) results: confidence level $\mathcal{L}$ = 90% and limit $\mathcal{M}$ = 0.5 (Part I)

	ARF-Reg		SOKNL		KNN	
	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW
Logarithm	92.26	28.26	92.67	25.6	91.7	28.06
HLI	92.22	28.05	92.57	25.32	91.54	27.78
Linear	**91.38**	26.54	91.54	**24.11**	91.18	27.1
MVE	92.75	29.36	93.33	26.77	92.0	28.37
Abalone						
Logarithm	95.53	30.56	96.35	20.41	97.34	21.35
HLI	94.56	29.46	95.68	19.03	96.59	18.74
Linear	**92.65**	27.92	94.64	17.26	95.7	**16.9**
MVE	98.21	34.02	97.28	22.66	98.21	25.46
CPU Activity						
Logarithm	94.25	9.87	99.99	3.09	99.99	3.11
HLI	94.08	9.19	99.99	**2.84**	99.99	2.86
Linear	**92.37**	8.62	99.99	**2.84**	99.99	2.86
MVE	94.93	11.11	99.99	5.69	99.99	5.73
Naval Propulsion Plant						
Logarithm	92.35	24.35	92.18	23.73	91.13	19.07
HLI	92.28	24.36	92.14	23.68	91.02	19.04
Linear	91.53	23.24	91.29	22.69	**90.55**	**18.68**
MVE	92.62	24.86	92.44	24.1	91.17	19.14
Ailerons						
Logarithm	95.89	23.15	95.73	21.84	95.89	19.65
HLI	95.74	22.65	95.71	21.31	95.79	19.17
Linear	**95.36**	20.87	**95.36**	19.86	95.43	**17.78**
MVE	96.21	24.77	96.05	23.16	96.22	20.97

Table 23 continued

	ARF-Reg		SOKNL		KNN	
	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW
Miami Housing						
Logarithm	95.35	3.75	96.12	3.25	94.96	3.35
HLI	95.28	3.67	95.92	3.11	94.86	3.31
Linear	94.63	3.37	95.31	**2.81**	**94.17**	3.06
MVE	95.65	3.96	96.56	3.53	95.38	3.53
NZ Energy Price						
Logarithm	96.56	4.82	96.71	4.82	95.94	17.52
HLI	96.06	4.56	96.14	4.55	94.86	**3.31**
Linear	95.1	4.17	95.16	4.16	**94.36**	4.93
MVE	97.09	5.3	97.27	5.32	97.66	7.17
Elevators						
Logarithm	**89.29**	28.55	89.2	25.36	89.05	41.55
HLI	87.39	25.63	87.04	23.05	86.47	36.67
Linear	87.98	26.39	87.62	23.7	86.98	37.44
MVE	86.33	24.24	85.64	**21.7**	83.91	32.81
Bike						
Logarithm	90.15	39.03	90.36	33.53	**90.0**	37.77
HLI	89.65	37.87	90.13	33.08	89.52	36.76
Linear	89.83	38.1	90.11	**33.01**	89.65	36.93
MVE	89.78	38.2	90.28	33.5	89.61	37.12
California Housing						
Logarithm	93.64	38.46	93.59	38.01	94.08	33.1
HLI	93.65	38.3	93.46	37.72	93.91	32.62
Linear	93.16	36.57	**93.15**	36.02	93.61	**31.3**
MVE	93.87	39.21	93.62	38.46	94.04	33.31

Table 24 Coverage (%) and NMPIW (%) results: confidence level $\mathcal{L}$ = 90% and limit $\mathcal{M}$ = 0.5 (Part II)

	ARF-Reg		SOKNL		KNN	
	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW
Logarithm	93.64	61.87	91.56	62.98	90.12	57.74
HLI	93.34	61.54	91.71	63.27	90.2	57.91
Linear	91.7	59.53	90.65	62.12	**90.0**	**57.57**
MVE	95.71	65.38	92.45	64.2	90.28	57.93
Health Insurance						
Logarithm	94.53	23.35	94.69	22.41	93.56	25.5
HLI	94.25	22.6	94.39	21.66	93.48	25.15
Linear	93.39	20.68	93.46	**19.71**	**92.59**	23.24
MVE	95.02	25.17	95.37	24.2	94.0	26.87
House8L						
Logarithm	91.6	76.19	91.21	73.44	90.78	88.62
HLI	91.7	76.53	91.37	73.77	90.97	89.02
Linear	90.81	74.41	90.59	**72.08**	**90.33**	87.68
MVE	92.2	77.85	91.79	74.66	91.2	89.51
MetroTraffic						
Logarithm	93.7	21.55	93.69	18.98	94.02	17.65
HLI	92.1	20.04	92.71	17.12	94.02	17.21
Linear	92.33	20.06	92.91	17.08	93.81	**16.52**
MVE	**90.86**	18.87	92.41	16.58	94.07	17.56
Diamonds						
Logarithm	93.06	15.5	93.49	9.33	93.31	9.14
HLI	93.03	15.48	93.38	9.24	93.25	9.09
Linear	**92.31**	14.54	92.58	8.53	92.59	**8.45**
MVE	93.3	15.98	93.88	9.75	93.59	9.45
Video Transcoding						
Logarithm	91.74	18.25	**89.89**	5.1	89.67	4.92
HLI	91.7	18.18	89.75	4.99	89.14	4.76
Linear	90.6	17.72	89.35	4.88	88.95	**4.68**
MVE	93.65	19.17	90.84	5.43	90.13	5.16

Table 24 continued

	ARF-Reg		SOKNL		KNN	
	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW
Wave Energy						
Logarithm	90.66	31.07	91.15	22.87	90.07	27.96
HLI	90.78	31.2	91.25	22.94	90.16	28.03
Linear	90.22	30.64	90.49	**22.34**	**89.98**	27.88
MVE	91.15	31.59	91.86	23.46	90.26	28.11
FriedmanGra						
Logarithm	90.58	31.04	90.84	22.88	90.08	27.96
HLI	90.7	31.17	90.93	22.95	90.11	27.99
Linear	90.18	30.66	90.31	**22.44**	**90.0**	27.89
MVE	91.03	31.53	91.49	23.43	90.21	28.07
FriedmanGsg						
Logarithm	89.92	25.07	90.78	18.3	89.32	24.46
HLI	89.84	25.02	90.82	18.33	88.15	23.46
Linear	89.53	24.77	**90.02**	**17.79**	88.5	23.75
MVE	90.18	25.32	91.56	18.86	88.09	23.38
FriedmanLea						
Logarithm	90.56	23.57	90.28	13.76	89.79	14.76
HLI	90.66	23.66	90.33	13.78	89.46	14.57
Linear	90.25	23.31	**90.01**	**13.6**	89.58	14.64
MVE	90.89	23.84	90.68	14.0	89.49	14.59
Hyper$_a$						

Table 25 Coverage (%) and NMPIW (%) results: confidence level $\mathcal{L}$ = 95% and limit $\mathcal{M}$ = 0.5 (Part I)

	ARF-Reg		SOKNL		KNN	
	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW
Logarithm	94.92	35.31	95.38	31.3	95.26	34.32
HLI	94.74	35.07	95.44	31.43	**94.98**	34.04
Linear	94.72	34.59	95.08	**30.21**	95.16	33.76
MVE	94.68	34.98	95.62	31.9	94.96	33.81
Abalone						
Logarithm	98.58	35.14	97.88	24.9	98.3	26.77
HLI	97.88	34.03	97.63	23.84	98.0	24.28
Linear	**96.89**	32.42	97.0	**22.05**	97.49	22.08
MVE	99.71	40.54	98.44	27.0	98.68	30.34
CPU Activity						
Logarithm	96.79	11.93	99.99	3.68	99.99	3.7
HLI	96.72	11.19	99.99	**3.39**	99.99	3.41
Linear	**95.74**	10.74	99.99	**3.39**	99.99	3.41
MVE	97.1	13.24	99.99	6.78	99.99	6.82
Naval Propulsion Plant						
Logarithm	95.24	30.44	95.46	29.17	94.91	23.44
HLI	**95.03**	30.21	95.37	29.01	94.71	23.07
Linear	95.08	30.2	95.21	28.57	94.86	23.18
MVE	94.8	29.62	95.16	28.71	94.41	**22.8**
Ailerons						
Logarithm	97.29	30.16	97.37	28.08	97.41	25.06
HLI	97.01	29.03	97.05	27.31	97.12	24.53
Linear	97.07	28.06	97.12	26.27	97.11	**23.69**
MVE	97.03	29.52	**96.91**	27.59	97.17	24.99
Miami Housing						
Logarithm	96.95	4.62	97.4	4.01	96.95	4.09
HLI	96.95	4.61	97.34	3.95	96.94	4.08
Linear	96.7	4.37	96.99	**3.68**	**96.63**	3.88
MVE	96.95	4.72	97.45	4.2	96.99	4.2
NZ Energy Price						
Logarithm	97.89	5.92	98.05	5.9	97.65	6.83
HLI	97.73	5.74	97.88	5.68	97.45	6.67
Linear	97.27	5.26	97.36	**5.2**	**96.93**	6.21
MVE	97.98	6.31	98.17	6.34	97.66	7.17

Table 25 continued

	ARF-Reg		SOKNL		KNN	
	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW
Elevators						
Logarithm	93.85	39.56	93.97	33.39	93.76	53.72
HLI	92.58	35.61	92.69	30.47	92.32	49.37
Linear	92.57	35.58	**92.82**	30.67	92.23	48.88
MVE	89.53	28.88	89.56	**25.86**	88.16	39.1
Bike						
Logarithm	94.55	51.71	**94.62**	44.58	94.53	49.58
HLI	93.71	48.69	93.89	41.83	93.65	46.72
Linear	94.0	49.62	94.14	42.94	94.03	47.68
MVE	92.65	45.52	93.2	**39.92**	92.75	44.23
California Housing						
Logarithm	96.42	46.85	96.41	45.04	96.59	41.52
HLI	96.44	45.88	96.41	45.03	96.36	40.05
Linear	**96.11**	44.2	96.17	43.3	96.3	**39.12**
MVE	96.55	46.72	96.54	45.82	96.12	39.69

Table 26 Coverage (%) and NMPIW (%) results: confidence level $\mathcal{L}$ = 95% and limit $\mathcal{M}$ = 0.5 (Part II)

	ARF-Reg		SOKNL		KNN	
	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW
Logarithm	97.73	71.11	97.05	71.61	95.54	68.18
HLI	97.28	69.83	96.77	71.33	95.61	68.83
Linear	96.39	**66.97**	95.88	69.35	**95.23**	67.45
MVE	98.85	77.9	98.63	76.5	95.75	69.03
Health Insurance						
Logarithm	96.09	29.1	96.35	27.82	95.41	31.68
HLI	96.09	29.15	96.34	27.75	95.45	31.86
Linear	95.66	27.56	95.84	**25.98**	**95.23**	31.08
MVE	96.32	29.99	96.61	28.83	95.46	32.01
House8L						
Logarithm	95.82	90.83	95.9	87.1	96.09	103.43
HLI	95.9	91.11	95.97	87.33	96.09	103.49
Linear	**95.38**	89.1	95.41	**85.17**	95.41	101.06
MVE	96.22	92.77	96.32	88.96	96.88	106.66
MetroTraffic						
Logarithm	96.23	29.31	96.08	27.54	96.12	26.97
HLI	95.35	27.21	95.16	25.11	95.55	23.58
Linear	95.18	27.15	94.93	24.72	95.59	23.74
MVE	93.06	22.49	93.43	**19.75**	**94.89**	20.92

Table 26 continued

	ARF-Reg		SOKNL		KNN	
	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW
Diamonds						
Logarithm	95.16	19.32	95.37	11.61	95.18	11.6
HLI	95.1	19.18	95.38	11.62	95.09	11.43
Linear	95.09	19.22	95.21	11.39	95.12	11.52
MVE	94.96	19.04	95.36	11.62	**94.98**	11.26
Video Transcoding						
Logarithm	96.86	21.09	**94.75**	6.23	94.61	5.98
HLI	96.64	20.83	94.36	6.03	94.0	5.75
Linear	95.81	20.19	94.46	6.0	94.21	**5.73**
MVE	98.28	22.84	**94.75**	6.47	94.27	6.14
Wave Energy						
Logarithm	95.37	36.98	95.46	27.2	**95.01**	33.38
HLI	95.46	37.13	95.5	27.28	95.02	33.42
Linear	95.13	36.48	95.17	**26.77**	**94.99**	33.34
MVE	95.73	37.64	95.92	27.96	95.09	33.5
FriedmanGra						
Logarithm	95.26	36.96	95.23	27.26	95.02	33.47
HLI	95.34	37.1	95.25	27.31	94.94	33.38
Linear	95.08	36.57	95.05	**26.96**	94.97	33.4
MVE	95.59	37.57	95.62	27.92	**95.0**	33.44
FriedmanGsg						
Logarithm	94.87	30.08	94.83	22.09	94.2	29.71
HLI	94.65	29.84	94.65	21.91	93.35	28.45
Linear	94.65	29.72	94.52	**21.71**	93.69	28.88
MVE	94.86	30.17	**95.04**	22.48	92.93	27.86
FriedmanLea						
Logarithm	94.91	28.22	94.69	17.14	94.64	18.27
HLI	94.9	28.19	94.2	**16.65**	94.04	17.69
Linear	94.83	28.09	94.46	16.87	94.39	18.02
MVE	**95.03**	28.41	94.17	16.68	93.7	17.39
$Hyper_a$						

Table 27 Coverage (%) and NMPIW (%) results: confidence level $\mathcal{L}$ = 99% and limit $\mathcal{M}$ = 0.5 (Part I)

	ARF-Reg		SOKNL		KNN	
	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW
Logarithm	98.49	52.78	98.57	45.72	98.51	51.35
HLI	98.49	58.89	98.55	51.78	98.51	57.33
Linear	98.43	54.84	**98.61**	47.98	98.53	53.26
MVE	97.57	45.97	98.03	**41.93**	97.61	44.43
Abalone						
Logarithm	99.83	44.56	99.4	34.26	99.11	41.28
HLI	99.68	44.75	99.35	36.56	99.1	45.95
Linear	99.55	40.89	99.19	**33.5**	**99.06**	42.81
MVE	99.93	53.28	99.57	35.48	**99.06**	39.88
CPU Activity						
Logarithm	98.74	17.08	99.99	4.83	99.99	4.86
HLI	98.73	16.6	99.99	**4.45**	99.99	4.48
Linear	98.76	16.5	99.99	**4.45**	99.99	4.48
MVE	**98.8**	17.41	99.99	8.91	99.99	8.97
Naval Propulsion Plant						
Logarithm	98.45	49.82	98.5	47.14	98.61	34.33
HLI	98.46	49.75	98.52	46.89	98.6	**33.97**
Linear	98.47	50.25	98.56	47.06	**98.68**	34.22
MVE	97.0	38.93	97.38	37.73	97.58	29.97
Ailerons						
Logarithm	99.15	48.34	99.2	43.88	99.17	40.5
HLI	99.17	49.12	99.2	44.11	99.17	41.19
Linear	**99.05**	48.68	99.15	43.39	99.07	40.95
MVE	98.04	38.79	98.13	36.26	98.16	**32.85**
Miami Housing						
Logarithm	98.77	7.4	98.78	6.47	98.88	6.2
HLI	98.77	7.47	98.77	6.56	98.87	6.26
Linear	98.8	7.44	98.81	6.51	**98.89**	6.4
MVE	98.12	6.2	98.39	**5.52**	98.41	**5.52**
NZ Energy Price						
Logarithm	99.12	8.84	99.18	8.74	99.07	10.41
HLI	99.12	8.86	99.17	8.71	99.08	10.39
Linear	99.12	8.79	99.17	8.62	99.06	10.45
MVE	98.88	**8.3**	**98.96**	8.33	98.65	9.42

Table 27 continued

	ARF-Reg		SOKNL		KNN	
	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW
Elevators						
Logarithm	98.29	57.64	98.31	49.07	98.27	75.03
HLI	98.29	58.04	**98.33**	49.25	98.31	75.49
Linear	97.96	56.76	98.1	48.34	98.04	74.09
MVE	93.49	37.96	94.33	**33.99**	93.14	51.38
Bike						
Logarithm	98.42	78.02	98.43	69.15	98.46	75.98
HLI	98.45	78.72	98.45	69.5	**98.49**	76.04
Linear	98.37	78.46	98.37	68.52	98.44	75.45
MVE	96.06	59.82	96.29	**52.46**	96.17	58.13
California Housing						
Logarithm	98.82	62.17	98.84	60.09	98.94	56.25
HLI	98.8	61.5	98.81	60.11	98.95	56.03
Linear	98.84	61.16	98.82	59.85	**98.97**	56.05
MVE	98.79	61.4	98.86	60.22	98.36	**52.17**
Superconductivity						

Table 28 Coverage (%) and NMPIW (%) results: confidence level $\mathcal{L}$ = 99% and limit $\mathcal{M}$ = 0.5 (Part II)

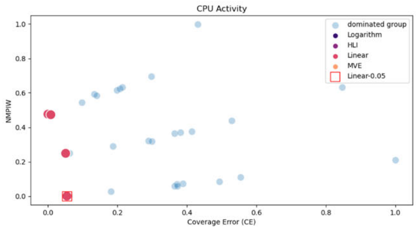
	ARF-Reg		SOKNL		KNN	
	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW
Logarithm	99.61	93.17	99.68	89.43	99.23	88.26
HLI	99.52	94.43	99.56	90.35	99.24	91.87
Linear	99.35	86.84	99.39	**83.1**	**99.1**	88.22
MVE	99.79	102.38	99.82	100.54	99.43	90.72
Health Insurance						
Logarithm	98.45	49.54	98.5	46.28	98.42	54.76
HLI	98.47	50.2	98.52	47.02	98.44	55.57
Linear	98.46	49.66	**98.54**	46.65	98.42	54.48
MVE	97.72	39.41	97.91	**37.89**	97.31	42.07
House8L						
Logarithm	99.45	114.99	99.42	110.85	99.49	128.57
HLI	99.37	114.33	99.35	110.15	99.4	127.61
Linear	99.16	111.6	**99.15**	**108.03**	99.2	124.09
MVE	99.76	121.92	99.75	116.91	99.88	140.17

Table 28 continued

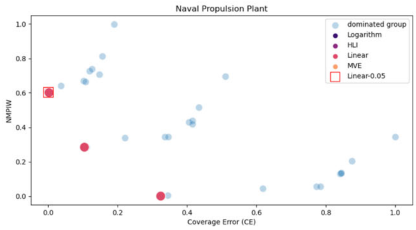
	ARF-Reg		SOKNL		KNN	
	Coverage	NMPIW	Coverage	NMPIW	Coverage	NMPIW
MetroTraffic						
Logarithm	**98.98**	43.63	98.84	43.52	98.51	56.32
HLI	98.86	45.67	98.79	44.77	98.62	56.0
Linear	98.47	46.51	98.27	44.83	98.07	50.83
MVE	95.51	29.56	94.96	**25.96**	95.96	27.49
Diamonds						
Logarithm	98.37	35.24	98.29	20.9	98.33	21.05
HLI	**98.41**	35.47	98.33	21.06	98.36	21.28
Linear	98.32	34.6	98.27	20.61	98.29	20.67
MVE	96.91	25.03	97.07	15.27	96.84	**14.8**
Video Transcoding						
Logarithm	99.61	26.28	98.65	8.92	98.63	8.25
HLI	99.51	25.82	98.65	8.8	98.63	8.14
Linear	99.31	24.92	**98.7**	8.84	98.68	8.16
MVE	99.93	30.02	98.05	8.51	98.15	**8.07**
Wave Energy						
Logarithm	99.07	48.26	98.92	**36.35**	98.95	44.18
HLI	99.08	48.69	98.89	36.52	98.93	44.38
Linear	**99.03**	47.96	98.94	36.46	**98.97**	44.33
MVE	99.2	49.47	98.92	36.74	98.93	44.03
FriedmanGra						
Logarithm	99.03	48.37	98.89	**36.58**	98.93	44.33
HLI	99.03	48.73	98.87	36.74	98.9	44.48
Linear	**99.01**	48.24	98.92	36.74	98.96	44.58
MVE	99.13	49.38	98.83	36.69	98.85	43.95
FriedmanGsg						
Logarithm	98.8	40.13	98.57	**32.04**	98.41	40.86
HLI	98.78	40.33	98.58	32.43	98.43	41.19
Linear	**98.86**	40.56	98.57	32.34	98.42	41.07
MVE	98.61	39.64	97.92	29.54	97.31	36.62
FriedmanLea						
Logarithm	98.84	38.09	98.61	24.65	98.63	25.56
HLI	98.8	37.95	98.62	24.65	98.63	25.58
Linear	**98.88**	38.29	98.67	24.91	98.7	25.86
MVE	98.65	37.34	98.05	**8.51**	97.67	22.85
Hyper_a						

Appendix B: full CING result

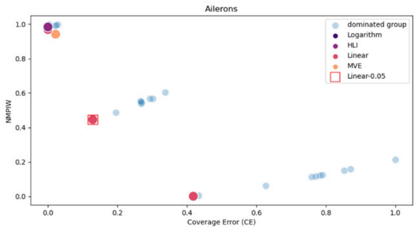
This appendix includes the supplementary figures for CING results.



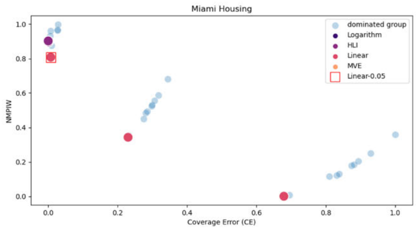
(a) CPU Activity



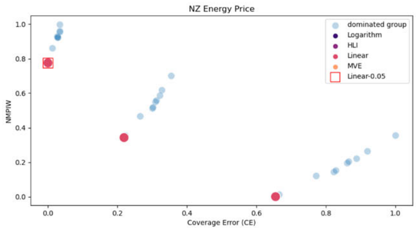
(b) Naval Propulsion Plant



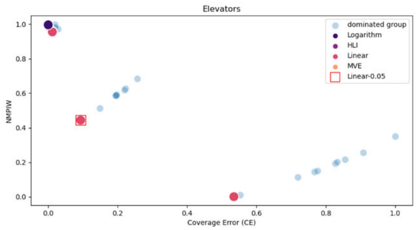
(c) Ailerons



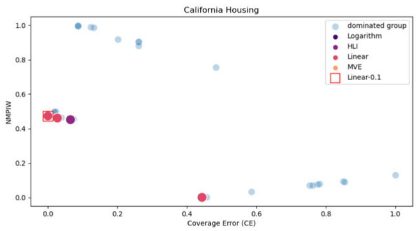
(d) Miami Housing



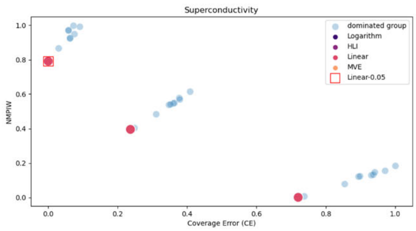
(e) NZ Energy Price



(f) Elevators

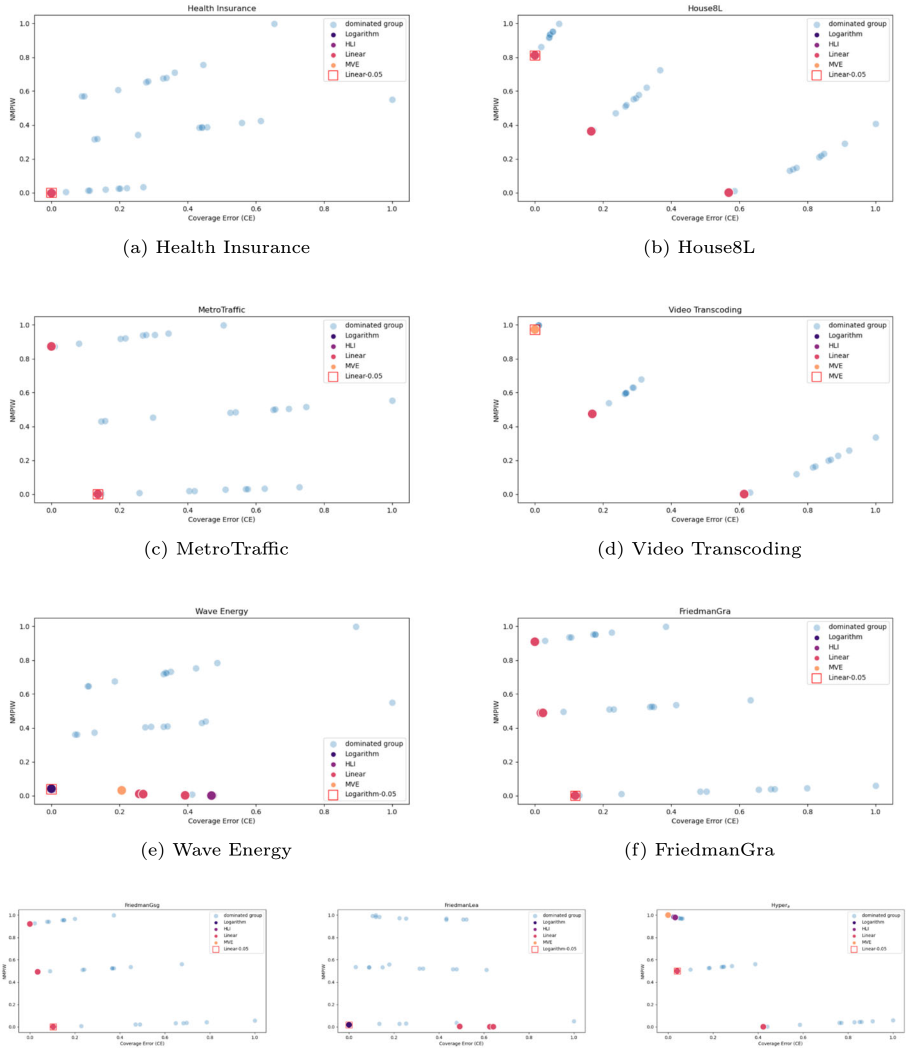


(g) California Housing



(h) Superconductivity

Fig. 16 Supplementary CING evaluation results for different datasets (Part I)



(a) Health Insurance (b) House8L

(c) MetroTraffic (d) Video Transcoding

(e) Wave Energy (f) FriedmanGra

(g) FriedmanGsg (h) FriedmanLea (i) Hyper$_a$

Fig. 17 Supplementary CING evaluation results for different datasets (Part II)

Appendix C: full coverage results comparison

We present scattergram for coverage results comparison for $\mathcal{L} = 80, 90,$ and 99 in this appendix (Figs. 18, 19 and 20).

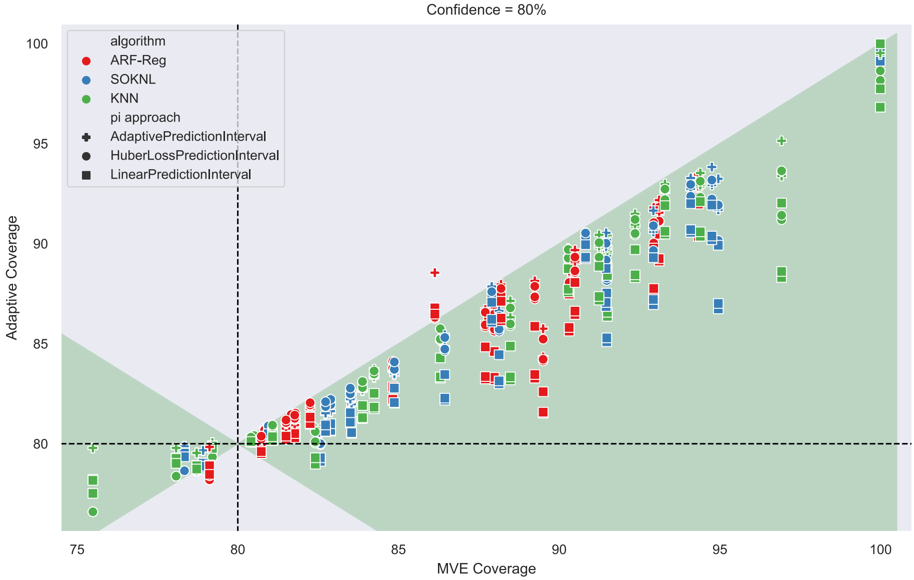


Fig. 18 Scattergram for $\mathcal{L} = 80\%$

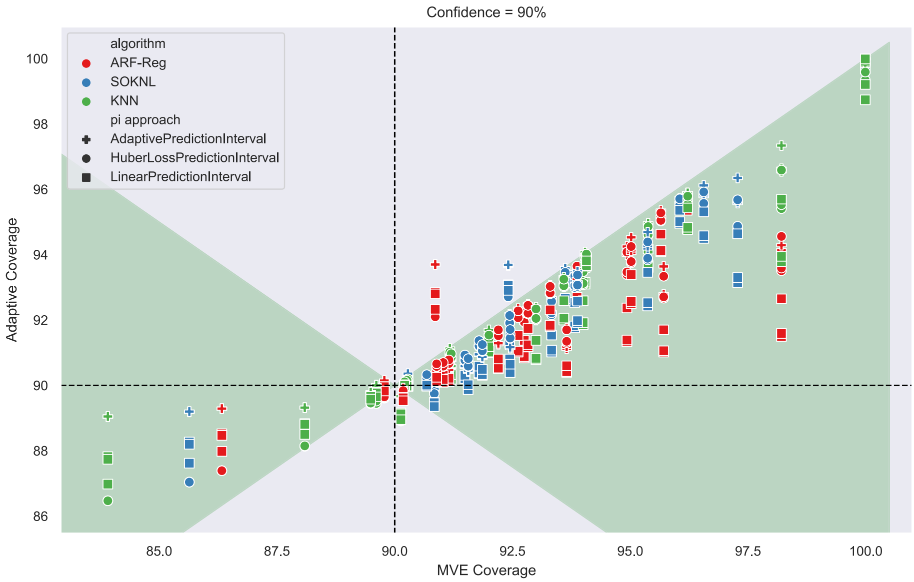


Fig. 19 Scattergram for $\mathcal{L} = 90\%$

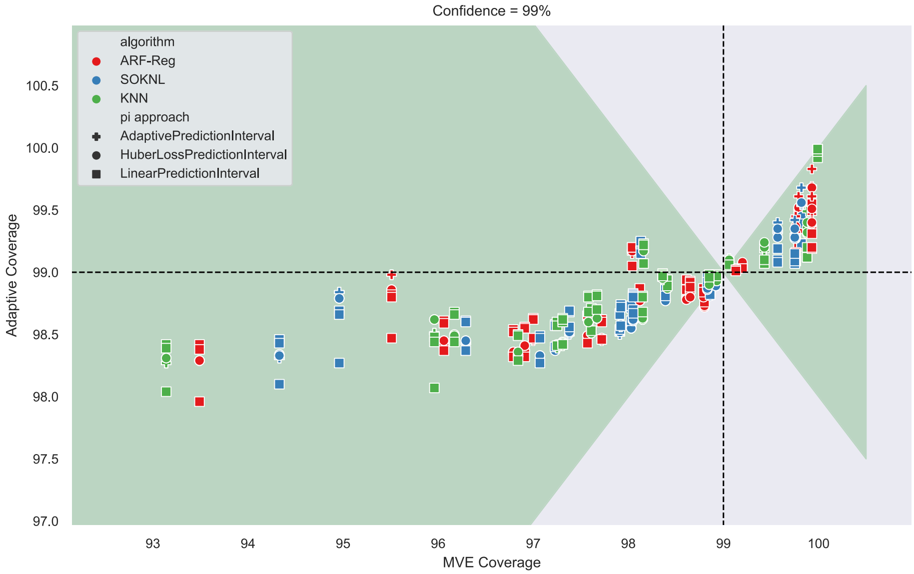
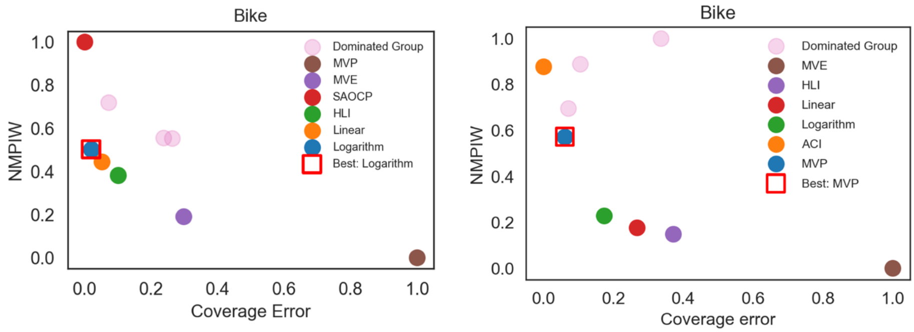


Fig. 20 Scattergram for $\mathcal{L} = 99\%$

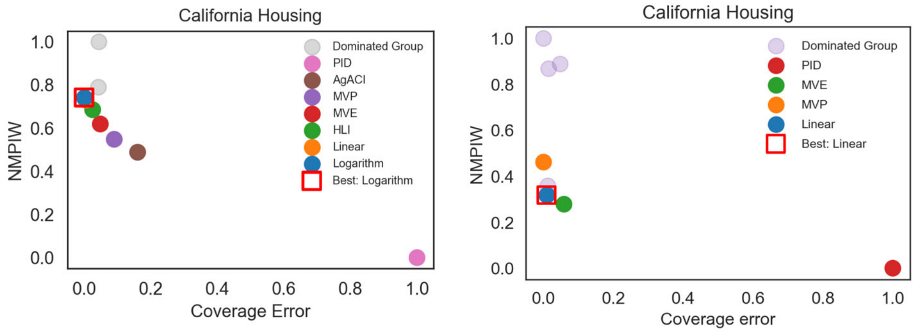
Appendix D: CING results for AdaPI and conformal prediction

This section provides the remaining CING evaluation plots for the comparison discussed in Sect. 7.6.1. Each dataset is shown under the Global and 500-windowed settings, matching the format used in Fig. 8 and Fig. 9. The files stored under `window=0` correspond to the Global setting (Figs. 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34).



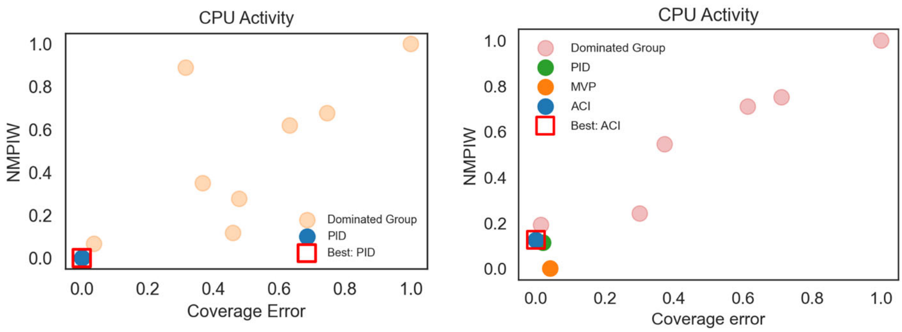
(a) Global (b) 500-Windowed

Fig. 21 CING results of AdaPI vs. conformal methods on bike



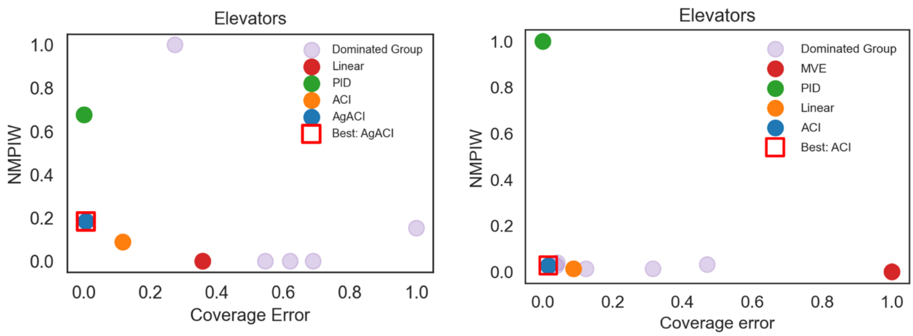
(a) Global (b) 500-Windowed

Fig. 22 CING results of AdaPI vs. conformal methods on California housing



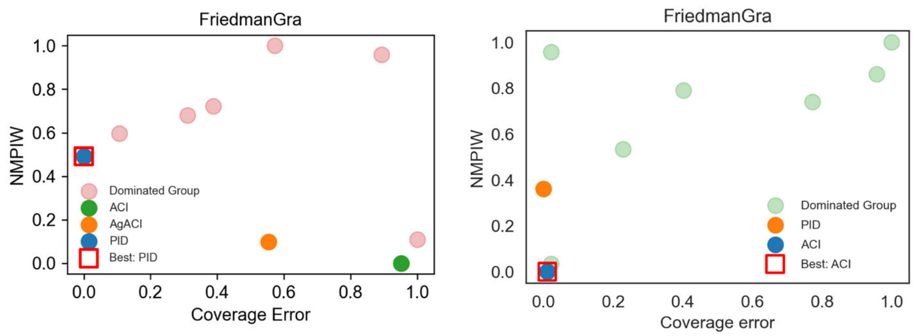
(a) Global (b) 500-Windowed

Fig. 23 CING results of AdaPI vs. conformal methods on CPU activity



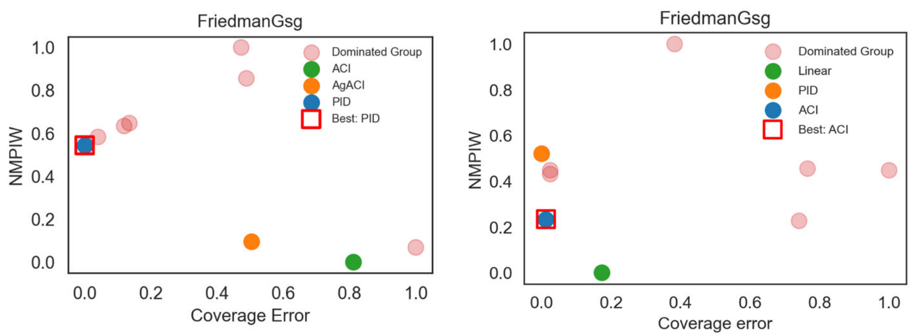
(a) Global (b) 500-Windowed

Fig. 24 CING results of AdaPI vs. conformal methods on elevators



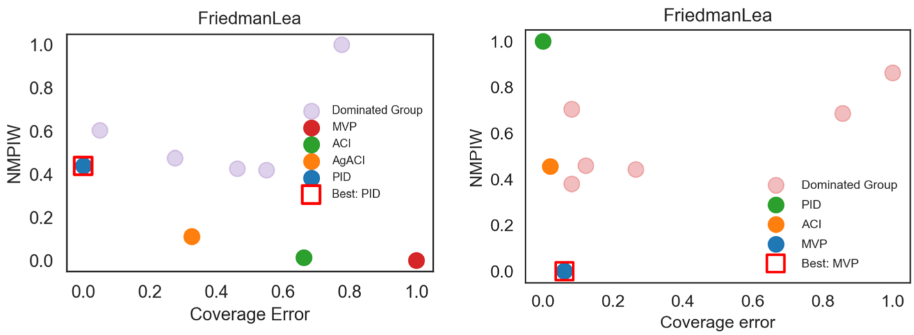
(a) Global (b) 500-Windowed

Fig. 25 CING results of AdaPI vs. conformal methods on FriedmanGra



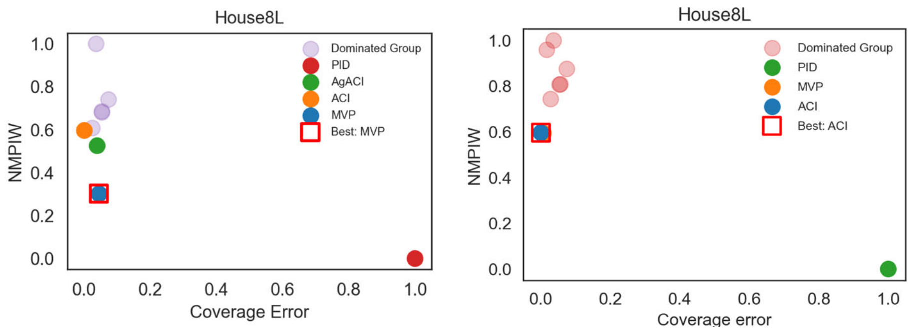
(a) Global (b) 500-Windowed

Fig. 26 CING results of AdaPI vs. conformal methods on FriedmanGsg



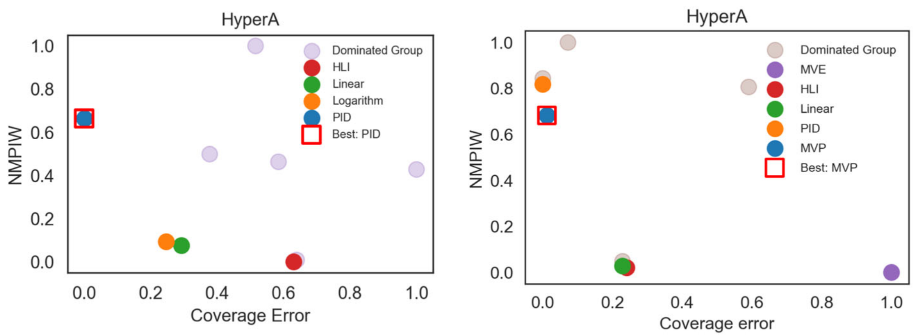
(a) Global (b) 500-Windowed

Fig. 27 CING results of AdaPI vs. conformal methods on FriedmanLea



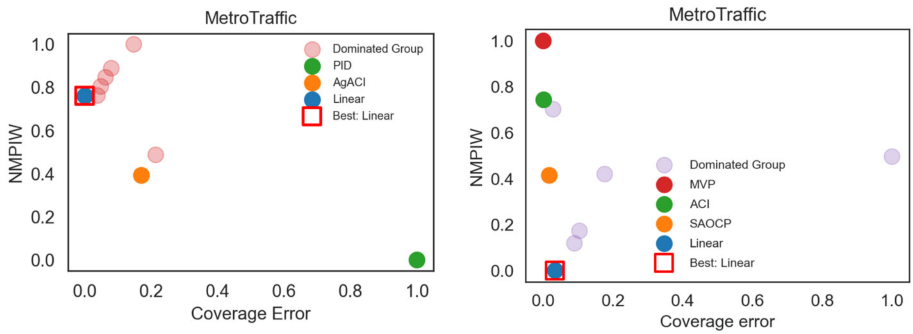
(a) Global (b) 500-Windowed

Fig. 28 CING results of AdaPI vs. conformal methods on House8L



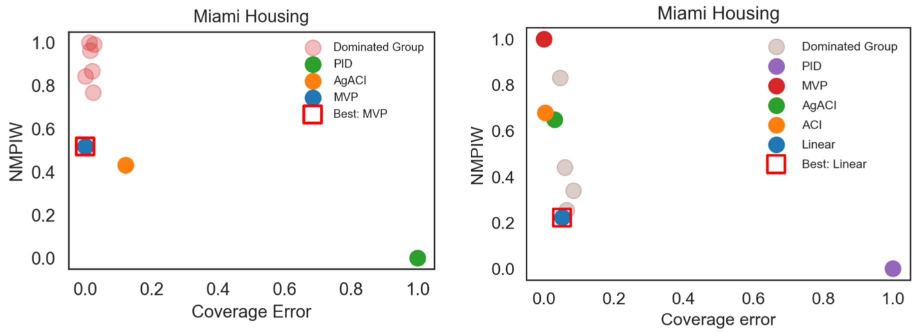
(a) Global (b) 500-Windowed

Fig. 29 CING results of AdaPI vs. conformal methods on HyperA



(a) Global (b) 500-Windowed

Fig. 30 CING results of AdaPI vs. conformal methods on MetroTraffic

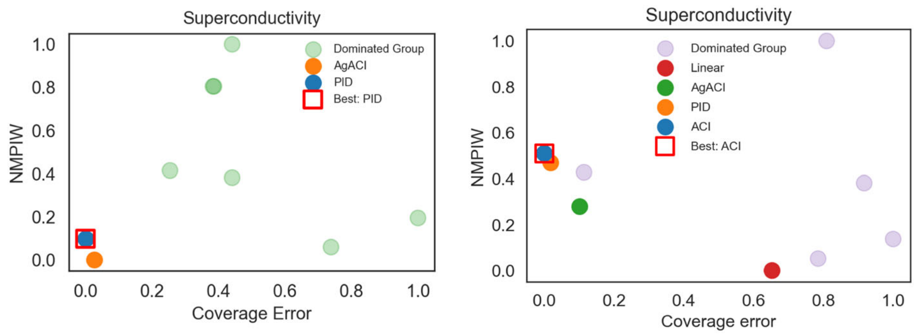


(a) Global (b) 500-Windowed

Fig. 31 CING results of AdaPI vs. conformal methods on Miami housing

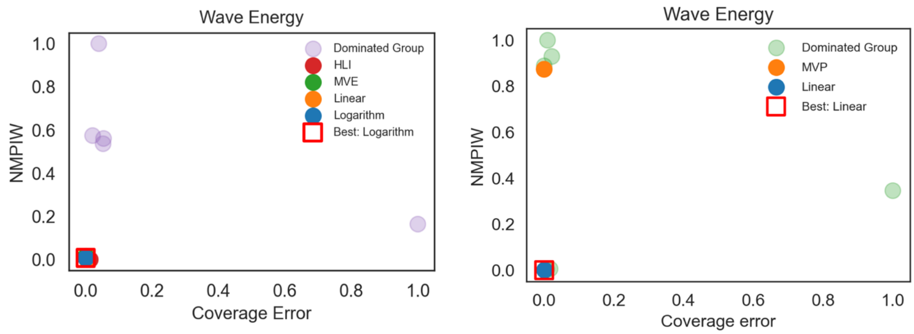
(a) Global (b) 500-Windowed

Fig. 32 CING results of AdaPI vs. conformal methods on NZ energy price



(a) Global (b) 500-Windowed

Fig. 33 CING results of AdaPI vs. conformal methods on superconductivity



(a) Global (b) 500-Windowed

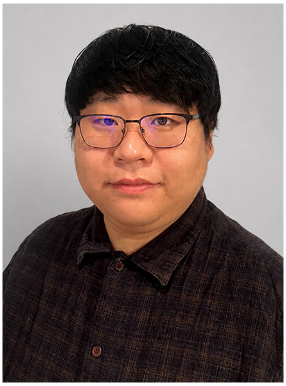
Fig. 34 CING results of AdaPI vs. conformal methods on wave energy

References

1. Hahn GJ (1969) Factors for calculating two-sided prediction intervals for samples from a normal distribution. J Am Stat Assoc 64(327):878–888
2. Shrestha DL, Solomatine DP (2006) Machine learning approaches for estimation of prediction interval for the model output. Neural Netw 19(2):225–235
3. Hahn GJ, Nelson W (1973) A survey of prediction intervals and their applications. J Qual Technol 5(4):178–188
4. Gomes HM, Montiel J, Mastelini SM, Pfahringer B, Bifet A (2020) On ensemble techniques for data stream regression. In: 2020 IJCNN. IEEE
5. Bifet A, Holmes G, Pfahringer B, Kranen P, Kremer H, Jansen T, Seidl T (2010) Moa: massive online analysis, a framework for stream classification and clustering. (PMLR)
6. Zhao J, Wang W, Sheng C, Zhao J, Wang W, Sheng C (2018) Industrial prediction intervals with data uncertainty. In: Data-driven prediction for industrial processes and their applications, pp 159–222
7. Hadjicharalambous M, Polycarpou MM, Panayiotou CG (2020) Neural network-based construction of online prediction intervals. Neural Comput Appl 32(11):6715–6733
8. Sun Y, Pfahringer B, Murilo Gomes H, Bifet A (2024) Adaptive prediction interval for data stream regression. In: Pacific-Asia conference on knowledge discovery and data mining. Springer, pp 130–141
9. SYa S, Melkumova L (2016) Normality assumption in statistical data analysis. In: CEUR Workshop Proceedings, vol 1638, pp 763–768
10. Sáez D, Ávila F, Olivares D, Cañizares C, Marín L (2014) Fuzzy prediction interval models for forecasting renewable resources and loads in microgrids. IEEE Trans Smart Grid 6(2):548–556
11. Shrivastava NA, Lohia K, Panigrahi BK (2016) A multiobjective framework for wind speed prediction interval forecasts. Renew Energy 87:903–910

12. Zhu D, Wang Q, Goddard J (2022) A new hedging hypothesis regarding prediction interval formation in stock price forecasting. J Forecast 41(4):697–717
13. Kavousi-Fard A (2016) Modeling uncertainty in tidal current forecast using prediction interval-based SVR. IEEE Trans Sustain Energy 8(2):708–715
14. Thombs LA, Schucany WR (1990) Bootstrap prediction intervals for autoregression. J Am Stat Assoc 85(410):486–492
15. Huang S-F, Hsu H-L (2020) Prediction intervals for time series and their applications to portfolio selection. REVSTAT-Stat J 18(1):131–151
16. Cortez P, Pereira PJ, Mendes R (2020) Multi-step time series prediction intervals using neuroevolution. Neural Comput Appl 32(13):8939–8953
17. Xu C, Xie Y (2021) Conformal prediction interval for dynamic time-series. In: International conference on machine learning. PMLR, pp 11559–11569
18. Liu Y, Zhao J, Wang W, Pedrycz W (2020) Prediction intervals for granular data streams based on evolving type-2 fuzzy granular neural network dynamic ensemble. IEEE Trans Fuzzy Syst 29(4):874–888
19. Barber RF, Candes EJ, Ramdas A, Tibshirani RJ (2021) Predictive inference with the jackknife. Ann Stat 49(1):486–507
20. Miller RG (1974) The jackknife-a review. Biometrika 61(1):1–15
21. Heskes T (1996) Practical confidence and prediction intervals. Adv Neural Inf Process Syst 9
22. Khosravi A, Nahavandi S, Creighton D, Atiya AF (2010) Lower upper bound estimation method for construction of neural network-based prediction intervals. IEEE Trans Neural Networks 22(3):337–346
23. Zhao X, Barber S, Taylor CC, Milan Z (2021) Interval forecasts based on regression trees for streaming data. Adv Data Anal Classif 15:5–36
24. Hothorn T, Hornik K, Van De Wiel MA, Zeileis A (2006) A lego system for conditional inference. Am Stat 60(3):257–263
25. Percival DB, Walden AT (2000) Wavelet methods for time series analysis, vol 4. Cambridge University Press, Cambridge, UK
26. Shafer G, Vovk V (2008) A tutorial on conformal prediction. J Mach Learn Res 9(3):371
27. Eklund M, Norinder U, Boyer S, Carlsson L (2015) The application of conformal prediction to the drug discovery process. Ann Math Artif Intell 74:117–132
28. Cortés-Ciriano I, Bender A (2019) Concepts and applications of conformal prediction in computational drug discovery. arXiv:1908.03569 arXiv preprint
29. Kath C, Ziel F (2021) Conformal prediction interval estimation and applications to day-ahead and intraday power markets. Int J Forecast 37(2):777–799
30. Angelopoulos A, Candes E, Tibshirani RJ (2023) Conformal PID control for time series prediction. Adv Neural Inf Process Syst 36:23047–23074
31. Bastani O, Gupta V, Jung C, Noarov G, Ramalingam R, Roth A (2022) Practical adversarial multivalid conformal prediction. Adv Neural Inf Process Syst 35:29362–29373
32. Bhatnagar A, Wang H, Xiong C, Bai Y (2023) Improved online conformal prediction via strongly adaptive online learning. In: International conference on machine learning. PMLR, pp 2337–2363
33. Gibbs I, Candes E (2021) Adaptive conformal inference under distribution shift. Adv Neural Inf Process Syst 34:1660–1672
34. Zaffran M, Féron O, Goude Y, Josse J, Dieuleveut A (2022) Adaptive conformal predictions for time series. In: International conference on machine learning. PMLR, pp 25834–25866
35. Nix DA, Weigend AS (1994) Estimating the mean and variance of the target probability distribution. ICNN'94. IEEE
36. Gunantara N (2018) A review of multi-objective optimization: methods and its applications. Cogent Eng 5(1):1502242
37. Eiben AE, Smith JE (2015) Introduction to evolutionary computing. Springer
38. Deb K, Pratap A, Agarwal S, Meyarivan T (2002) A fast and elitist multiobjective genetic algorithm: Nsga-II. IEEE Trans Evol Comput 6(2):182–197
39. Emmerich MT, Deutz AH (2018) A tutorial on multiobjective optimization: fundamentals and evolutionary methods. Nat Comput 17:585–609
40. Huber PJ (1992) Robust estimation of a location parameter. In: Breakthroughs in statistics: methodology and distribution. Springer, New York, pp 492–518
41. Nash W, Sellers T, Talbot S, Cawthorn A, Ford W (1995). Abalone UCI machine learning repository. https://doi.org/10.24432/C55C7W
42. Coraddu A, Oneto L, Ghio A, Savio S, Anguita D, Figari M (2014) Condition based maintenance of naval propulsion plants. UCI Mach Learn Repos. https://doi.org/10.24432/C5K31K
43. Torgo L (2000) Partial linear trees. ICML, pp 1007–1014

44. Fanaee-T H, Gama J (2014) Event labeling combining ensemble detectors and background knowledge. Progress Artif Intell 2:113–127
45. Pace RK, Barry R (1997) Sparse spatial autoregressions. Stat Prob Lett 33(3):291–297
46. Hamidieh K (2018) A data-driven statistical model for predicting the critical temperature of a superconductor. Comput Mater Sci 154:346–354
47. Wickham H (2016) Ggplot2: elegant graphics for data analysis. Springer, New York. https://ggplot2.tidyverse.org
48. Deneke T (2015) Online video characteristics and transcoding time dataset. UCI Mach Learn Repos. https://doi.org/10.24432/C58C9K
49. Ikonomovska E, Gama J, Džeroski S (2011) Learning model trees from evolving data streams. Data Min Knowl Disc 23:128–168
50. Hulten G, Spencer L, Domingos P (2001) Mining time-changing data streams. In: Proceedings of the seventh ACM SIGKDD international conference on knowledge discovery and data mining, pp 97–106
51. Gomes HM, Lee A, Gunasekara N, Sun Y, Cassales GW, Liu JJ, Heyden M, Cerqueira V, Bahri M, Koh YS, Pfahringer B, Bifet A (2025) CapyMOA: efficient machine learning for data streams in python. arXiv:2502.07432
52. Gomes HM, Barddal JP, Ferreira LEB, Bifet A (2018) Adaptive random forests for data stream regression. ESANN
53. Sun Y, Pfahringer B, Gomes HM, Bifet A (2022) SOKNL: a novel way of integrating k-nearest neighbours with adaptive random forest regression for data streams. Data Min Knowl Disc 36(5):2006–2032
54. Gomes HM, Bifet A, Read J, Barddal JP, Enembreck F, Pfharinger B, Holmes G, Abdessalem T (2017) Adaptive random forests for evolving data stream classification. Mach Learn 106:1469–1495
55. Bifet A, Holmes G, Pfahringer B (2010) Leveraging bagging for evolving data streams. In: machine learning and knowledge discovery in databases: European conference, ECML PKDD 2010, Barcelona, Spain, September 20-24, 2010, Proceedings, Part I 21. Springer, pp 135–150
56. Bifet A, Gavalda R (2007) Learning from time-changing data with adaptive windowing. In: Proceedings of the 2007 SIAM international conference on data mining. SIAM, pp 443–448



Yibin Sun is an AI researcher and engineer specializing in adaptive machine learning, data stream mining, continual learning, and trustworthy AI. He received his PhD in Computer Science from the University of Waikato, New Zealand in 2025. His research focuses on developing machine learning systems that can adapt to evolving environments, with applications spanning data streams, uncertainty estimation, and intelligent decision-making.

Bernhard Pfahringer received his PhD degree from the University of Technology in Vienna, Austria, in 1995. He is a Professor in the Department of Computer Science at the University of Waikato and a co-director of the Waikato AI Institute. His interests span several machine learning sub-fields, including deep learning, streaming data, randomization, and complex data in general.

Heitor Murilo Gomes (Member, IEEE) is currently a senior lecturer with the Victoria University of Wellington in New Zealand. He was a senior research fellow and co-director of the AI Institute with the University of Waikato. His main research interests include the application of machine learning for data streams in a variety of tasks, such as ensemble learning, drift detection and adaptation, and the intersection of online continual learning and streaming data. He has contributed to several open-source projects like MOA (Massive Online Analysis) and recently led the development of CapyMOA (www.capymoa.org), and a Python library for efficient data stream learning. He will also be the general co-chair of PAKDD 2027.

Albert Bifet is the Director of the Te Ipu o te Mahara AI Institute at the University of Waikato and Professor at TelecomParis, Institute Polytechnique de Paris. His research focuses on Artificial Intelligence, big data science, and machine learning for data streams. He is leading the TAIAO Environmental Data Science project and co-leading the open source projects MOA Massive On-line Analysis and Stream DM for Spark Streaming and SAMOA Scalable Advanced Massive Online Analysis. He is the co-author of a book on machine learning from data streams published at MITPress. He is one of the winners of the best paper award at the ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT) 2023, and he was the general co-chair of the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML-PKDD) 2024.