

**Uzun I. S.**, PhD Candidate, Institute of Artificial Intelligence and Robotics, National University "Odessa Polytechnic"  
ORCID: 0000-0001-6619-4862

**Lobachev M. V.**, PhD, Professor, Institute of Artificial Intelligence and Robotics, National University "Odessa Polytechnic"  
ORCID: 0000-0002-4859-304X

## RELIABILITY-DRIVEN STREAMING MODELLING OF MULTIMODAL TIME SERIES FOR ROBUST DECISION SUPPORT UNDER DRIFT AND MODALITY DEGRADATION

*Streaming decision support systems that process multimodal time series must remain robust under simultaneous concept drift and temporary modality degradation. Existing approaches usually treat multimodal fusion, drift adaptation, anomaly detection, and probability calibration as separate problems, which makes it difficult to distinguish whether a loss of predictive quality is caused by a real change in the process or by a temporary failure of one modality. This paper presents a unified online pipeline in which online reliability estimation is used as a single control interface for reliability-adaptive dynamic fusion, drift-initiated budgeted micro-adaptation, and reliability-constrained anomaly detection under a false alarm rate budget. Reliability is modeled as a causal probability of the current non-degraded modality state, post-hoc calibrated, and reused in all downstream control rules. Evaluation follows the prequential protocol on controlled streams with deterministic injections of modality degradation, concept drift, and anomalies, and on real data from UCI Appliances Energy Prediction and UCI Air Quality. The calibrated reliability model retains high degradation-separation ability ( $ROC - AUC = 0.8624$ ) and improves calibration to  $ECE = 0.0845$  versus  $0.1840$  without calibration. RADF preserves clean-regime quality ( $MAE = 0.5557$ ;  $RMSE = 0.6952$ ) and improves degraded segments, for example  $MAE = 0.6402$  versus  $0.6820$  for early fusion under alternating missingness. Budgeted micro-adaptation improves post-drift forecasting relative to no adaptation ( $MAE = 0.6613$  versus  $0.7046$ ; average recovery 157 versus 800 steps) while updating only a three-parameter head within fixed budgets. RC-AD increases  $Recall@FAR$  at all tested budgets, including  $0.335$  versus  $0.103$  at a false alarm rate budget of  $0.05$  on controlled streams. In the integrated stress scenario, the system achieves  $ROC - AUC = 0.983$ ,  $ECE = 0.029$ ,  $Recall@FAR = 0.311$ , and  $PR - AUC = 0.287$ ; in robust application protocols with severe segment missingness, it reduces  $MAE$  by about 35 % in the energy domain and about 93 % in the BTC/USD domain relative to naive fusion. These results present one coherent streaming decision support result rather than isolated local methods.*

**Key words:** machine learning; data analysis; information systems; decision support systems; multimodal time series; modality degradation; concept drift; online reliability estimation; probability calibration; reliability-adaptive dynamic fusion; budgeted micro-adaptation; reliability-constrained anomaly detection; false alarm rate budget; prequential protocol.

**Узун І. С., Лобачев М. В. Надійнісно-кероване потокове моделювання мультимодальних часових рядів для робастної підтримки прийняття рішень за дрейфу та деградації модальностей**

*Потокові системи підтримки прийняття рішень, що працюють з мультимодальними часовими рядами, повинні зберігати робастність за спільної дії концептуального дрейфу та тимчасової деградації модальностей. Ізольований розгляд мультимодального злиття, адаптації до дрейфу, виявлення аномалій і калібрування ймовірнісних оцінок не дозволяє надійно відрізнити реальну зміну закономірностей потоку від короточасного погіршення якості окремого джерела даних. У статті подано єдиний онлайн-конвеєр для поточкових СППР, у якому онлайн-оцінювання надійності модальностей використовується як спільний керувальний інтерфейс для надійнісно-адаптивного динамічного злиття, дрейф-ініційованої бюджетованої мікроадаптації та надійнісно-обмеженого виявлення аномалій за бюджетом частки хибних тривог. Надійність інтерпретується як каузальна ймовірність поточного недеградованого стану модальності, калібрується ізотонічною регресією та повторно використовується в усіх правилах керування. Експериментальну перевірку виконано у передпослідовному протоколі на контрольованих потоках з детермінованими ін'єкціями деградації модальностей, дрейфу та аномалій, а також на реальних даних UCI Appliances Energy Prediction і UCI Air Quality. Калібрована модель надійності зберігає високу здатність відокремлювати деградовані стани ( $ROC - AUC = 0.8624$ ) та покращує каліброваність до  $ECE = 0.0845$  порівняно з  $0.1840$  без калібрування. Метод RADF зберігає якість у чистому режимі ( $MAE = 0.5557$ ,  $RMSE = 0.6952$ ) та зменшує похибку в деградованих сегментах, зокрема до  $MAE = 0.6402$  проти  $0.6820$  для раннього злиття у режимі *alternating\_missing*. Бюджетована мікроадаптація покращує пост-дрейфову якість відносно відсутності адаптації ( $MAE = 0.6613$  проти  $0.7046$ ; середній час відновлення 157 проти 800 кроків)*



---

за оновлення лише трьох параметрів малої голови у фіксованих бюджетах. Метод RC-AD підвищує  $Recall@FAR$  для всіх досліджених FAR-бюджетів, зокрема до 0.335 проти 0.103 для наївного мультимодального скорингу за бюджету 0.05. У контрольованому інтегрованому стрес-сценарії система досягає  $ROC - AUC = 0.983$ ,  $ECE = 0.029$ ,  $Recall@FAR = 0.311$  та  $PR - AUC = 0.287$ ; у робастних прикладних протоколах із сегментними пропусками знижує MAE приблизно на 35 % у домені енергоспоживання та приблизно на 93 % у домені BTC/USD порівняно з наївним злиттям. Отримані результати подають єдиний науковий результат, а не чотири локальні методи.

Ключові слова: машинне навчання; аналіз даних; інформаційні системи; системи підтримки прийняття рішень; мультимодальні часові ряди; деградація модальностей; концептуальний дрейф; онлайн-оцінювання надійності; калібрування ймовірностей; надійнісно-адаптивне динамічне злиття; бюджетована мікроадаптація; надійнісно-обмежене виявлення аномалій; бюджет частки хибних тривог; передпослідовний протокол.

**Formulation of the problem.** Streaming decision support systems increasingly rely on multimodal time series collected from several heterogeneous data sources. In practical deployments, the quality of a forecast or an alarm may deteriorate because the underlying process changes over time, because one modality becomes temporarily unreliable, or because both effects occur simultaneously. This distinction is operationally critical: concept drift requires model adaptation, whereas modality degradation requires conservative reweighting, threshold correction, or operator-side data-quality actions. In streaming decision support systems, these actions must be causal, interpretable, and compatible with low-latency processing [1–7].

The dissertation on which this article is based treats this problem as a single research loop rather than as several disconnected local tasks. The central idea is that a calibrated online reliability scale  $r_t^{(m)} \in [0,1]$  can be reused as a common control signal for forecasting, drift-triggered adaptation, and anomaly detection. This article therefore presents the integrated pipeline that combines online reliability estimation and probability calibration, reliability-adaptive dynamic fusion, budgeted micro-adaptation under relevance drift, and reliability-constrained anomaly detection under a false alarm rate budget.

**Analysis of recent research and publications.** Related studies provide strong foundations for individual components of the problem, but not yet for their coordinated streaming use. Multimodal machine learning and multimodal data fusion are well established as separate research areas, with surveys covering representation, fusion strategies, and missing-view settings [1–3]. These studies show why multiple modalities can improve decision quality, yet they usually do not make online modality reliability a shared control interface across several downstream procedures.

Concept drift adaptation is likewise a mature topic, with surveys emphasizing that non-stationarity changes the relationship between inputs and targets and therefore requires explicit adaptation logic [4, 5]. However, drift adaptation methods often react to error changes without explicitly separating drift from temporary modality degradation. This is problematic in streaming decision support systems because degradation-related error spikes may trigger unnecessary updates and consume adaptation budgets.

Probability calibration and anomaly detection also have mature literatures [9–13]. Calibration studies explain why raw model scores may be poorly aligned with empirical frequencies [9–11], while anomaly detection surveys and ROC-oriented evaluation frameworks explain how ranking quality and threshold choice should be interpreted [12, 13]. Nevertheless, the literature still lacks a unified streaming narrative in which calibrated modality reliability is reused to control multimodal fusion, to guard budgeted drift response, and to preserve a false alarm rate budget under modality degradation.

**The purpose of the article.** The purpose of the article is to present and experimentally validate a unified online pipeline for robust decision support under concept drift and modality degradation, and to show that online reliability estimation is the common control interface within one streaming narrative. The research gap addressed in the article is the absence of one operationally stable interface that links multimodal forecasting, budgeted drift response, and false-alarm-constrained anomaly detection in the same streaming decision support system.

The proposed pipeline contains four tightly connected functional parts: online reliability estimation with probability calibration, reliability-adaptive dynamic fusion, drift-initiated budgeted micro-adaptation, and reliability-constrained anomaly detection under a false alarm rate budget. Their interaction is described below and visualized in Fig. 1.

#### Presenting main material.

**Unified formal model of the streaming multimodal process.** Let a streaming multimodal process be defined by the sequence

$$\left\{ \left( \mathbf{x}_t^{(1)}, \dots, \mathbf{x}_t^{(M)}, y_t \right) \right\}_{t \geq 1},$$

where  $\mathbf{x}_t^{(m)}$  is the observation from modality  $m$  at time  $t$ ,  $y_t$  is the target for one-step-ahead decision support, and only the causal information set  $\mathcal{I}_t$  is available at decision time. For each modality, a latent degradation state  $d_t^{(m)} \in \{0,1\}$  may temporarily corrupt the stream through missingness, noise, or scale distortion. At the same time, the predictive relevance of modalities may change because of concept drift. The integrated pipeline starts from the

causal interpretation of reliability. For each modality  $m$ , reliability is defined as the probability that the current state is non-degraded:

$$r_t^{(m)} = \mathbb{P}(d_t^{(m)} = 0 | \mathcal{I}_t).$$

This definition is important because it turns modality quality into a quantity that can be reused across tasks. The scale is not a generic confidence score of the predictive model; it is an estimate of the current health of the data source itself. This distinction is central when concept drift and modality degradation produce similar error symptoms but require different actions.

In the forecasting branch, reliability is transformed into fusion weights through normalized relevance:

$$w_t^{(m)} = \frac{\pi^{(m)} r_t^{(m)}}{\sum_{k=1}^M \pi^{(k)} r_t^{(k)}},$$

where  $\pi^{(m)} > 0$  is the clean-regime prior relevance of modality  $m$ . The weights are then smoothed with an exponential moving average to prevent unstable switching. In the adaptation branch, a smoothed estimate of optimal fusion relevance is monitored for persistent sign changes that indicate modality-relevance drift. In the anomaly branch, the system emits a score  $s_t$  and an alarm  $a_t = \mathbb{I}[s_t > \tau]$  under the explicit constraint

$$FAR(\tau) = \frac{\sum_{t: y_t^{anom}=0} \mathbb{I}[s_t > \tau]}{\sum_{t: y_t^{anom}=0} 1} \leq \beta,$$

where  $\beta$  is the false alarm rate budget. This common notation directly links methods, metrics, and operational decisions.

The overall control loop of the proposed system is summarized in Fig. 1.

**Online reliability estimation as a unified control interface.** The dissertation models online reliability on lightweight degradation proxies  $\phi_t^{(m)}$  so that the signal remains causal and inexpensive to compute. The learnable form is

$$r_t^{(m)} = \sigma(z_t^{(m)}), \quad z_t^{(m)} = \mathbf{w}^\top \phi_t^{(m)} + b,$$

followed by post-hoc isotonic calibration

$$\hat{r}_t^{(m)} = g(z_t^{(m)}), \quad g: \mathbb{R} \rightarrow [0, 1].$$

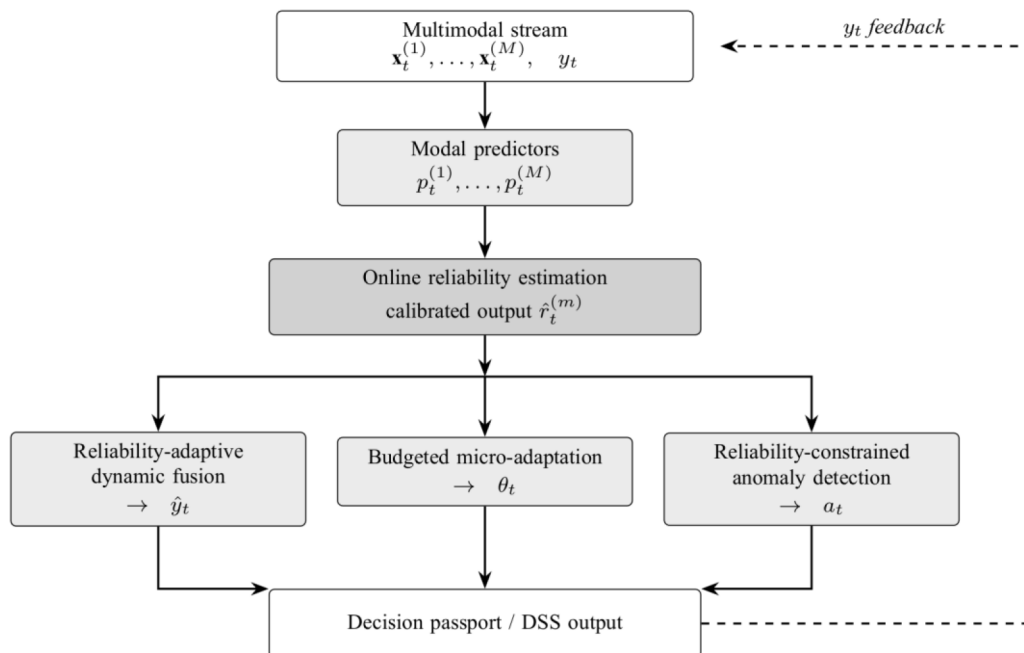


Fig. 1. Unified reliability-driven online pipeline for multimodal time-series decision support under concept drift and modality degradation

Source: compiled by the author based on the dissertation formal model and repository pipeline implementation

The control logic relies on the calibrated value  $\hat{r}_t^{(m)}$ , because calibrated probabilities are required for stable thresholding and conservative operator rules. This is why the article treats reliability estimation not as an auxiliary diagnostic but as the single control interface of the whole streaming system.

The empirical results support this interpretation. On controlled degradation detection, the calibrated model achieved  $ROC - AUC = 0.8624$  and balanced accuracy 0.8415, while the uncalibrated online model achieved slightly higher discrimination ( $ROC - AUC = 0.8755$ ) but substantially worse calibration ( $ECE = 0.1840$  versus 0.0845). This distinction matters because the pipeline uses reliability values in hard control rules, not only in ranking. The ablation analysis also shows that removing missing rate from the proxy set decreases AUC to 0.7473, confirming that degradation signals are not redundant. In other words, the practical value of the reliability module is not only detection quality, but the availability of a stable probability-like control scale that downstream modules can trust.

The role of the calibrated reliability scale as a shared control interface is illustrated in Fig. 2.

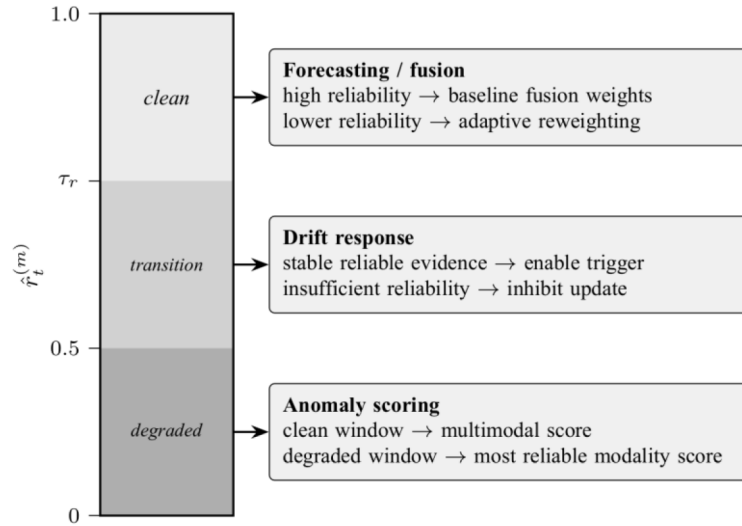


Fig. 2. Online reliability estimation as the unified control interface for fusion, micro-adaptation, and anomaly detection  
Source: compiled by the author based on the dissertation control rules and reliability scale semantics

**Reliability-adaptive dynamic fusion.** RADF converts online reliability into dynamic fusion weights and adds a clean-gate rule to guarantee no quality regression in windows without degradation. In the dissertation, the reliability-driven component is defined as

$$r_t^{(m)} = r_{t,err}^{(m)} \cdot r_{t,miss}^{(m)},$$

and the two-modality target weight is

$$w_t^* = \frac{w_0 r_t^{(1)}}{w_0 r_t^{(1)} + (1 - w_0) r_t^{(2)} + \epsilon}.$$

If both modalities satisfy clean-gate conditions on missingness and normalized error, RADF outputs the strong early-fusion baseline; otherwise it uses reliability-adaptive late fusion. This design makes RADF safe in clean segments and robust in degraded segments.

The results show exactly this pattern. In the clean condition, RADF reproduces the early-fusion baseline with  $MAE = 0.5557$  and  $RMSE = 0.6952$ . Under alternating missingness, RADF reduces  $MAE$  to 0.6402, compared with 0.6820 for early fusion and 0.6654 for late fusion. Under alternating mixed degradation, RADF reaches  $MAE = 0.6298$ , compared with 0.6584 and 0.6608 respectively. The robustness study further shows that at missing level 0.9, RADF keeps  $MAE = 0.6425$ , while early fusion reaches 0.7050. On real UCI Appliances data, RADF reaches  $MAE = 63.6943$  in degraded segments versus 381.9107 for early fusion, while keeping competitive clean-segment quality ( $MAE = 59.9149$ ). These results confirm that the proposed fusion module is not a generic dynamic fusion heuristic: it is a reliability-controlled fusion rule for streaming forecasting under modality degradation.

**Drift-initiated budgeted micro-adaptation.** The adaptation module is deliberately small so that concept drift can be addressed without converting the pipeline into continuous retraining. The predictive head is

$$\hat{y}_t = \alpha_t p_t^{(1)} + \beta_t p_t^{(2)} + c_t,$$

where only the three parameters  $\theta_t = (\alpha_t, \beta_t, c_t)$  are updated. A drift event is defined through the optimal fusion weight on a sliding window:

$$\hat{w}_t \in \arg \min_{w \in [0,1]} \sum_{i \in \mathcal{V}_t} (y_i - (w p_i^{(1)} + (1-w) p_i^{(2)}))^2,$$

followed by exponential smoothing  $\bar{w}_t$ . Drift is triggered when  $\bar{w}_t$  crosses the neutral threshold 0.5 to the side opposite to the clean-regime base relevance and stays there for  $P$  steps, after which a cooldown period  $C$  is enforced. Once triggered, the head is updated on the last  $n \leq B_s$  samples under ridge regularization. This makes drift response event-triggered, parameter-efficient, and explicitly budgeted.

The numerical evidence is consistent with this budgeted interpretation. The proposed micro-adaptation achieves  $MAE = 0.6613$  and  $RMSE = 0.8401$ , compared with 0.7046 and 0.8952 without adaptation. Its average recovery time is 157 steps, versus 241.67 for periodic adaptation and 800 for no adaptation. At the same time, it updates only 9 parameters in total and uses 240 samples for updates, whereas periodic adaptation updates 72 parameters and uses 4800 samples. This point is important: periodic adaptation attains lower absolute error ( $MAE = 0.6434$ ) but spends a much larger update budget. The proposed adaptation module is therefore best interpreted as a controlled adaptation strategy, not as an unrestricted accuracy maximizer. The trigger diagnostics also support this reading: recall is 1.000 with average delay 43.9 in the abrupt-strong scenario, and 0.867 with average delay 257.8 in the weak-gradual-noisy scenario.

**Reliability-constrained anomaly detection under a false alarm rate budget.** RC-AD treats anomaly detection as a budgeted streaming control problem. The base residual score is

$$s_t = \frac{\varepsilon_t - \mu_\varepsilon}{\sigma_\varepsilon + \epsilon},$$

and the reliability-constrained scoring rule is

$$s_t^{RC} = \begin{cases} s_t^{mm}, & \text{if the window is clean,} \\ s_t^{(m^*)}, & \text{otherwise,} \end{cases}$$

where  $m^* = \arg \max_m r_t^{(m)}$  is the currently most reliable modality. The threshold is selected from the calibration segment as the empirical  $(1 - \beta)$ -quantile, so that the operational rule satisfies the false alarm rate budget. In the dissertation, the target metric is defined as

$$Recall@FAR(\beta) = \max_{\tau: FAR(\tau) \leq \beta} TPR(\tau).$$

This definition directly aligns the evaluation with the operator-facing false alarm rate requirement of the decision support system.

The controlled-stream results show that RC-AD improves Recall@FAR at all studied budgets. At  $\beta = 0.01$ , RC-AD reaches 0.121, compared with 0.034 for naive multimodal scoring. At  $\beta = 0.05$ , it reaches 0.335, compared with 0.103 for naive multimodal scoring and 0.186 for equal-weight fusion. At  $\beta = 0.1$ , it reaches 0.426, compared with 0.182 and 0.273 respectively. On UCI Air Quality, RC-AD remains the strongest multimodal method under the strict budget  $\beta = 0.01$ , with  $Recall@FAR = 0.097$  and  $PR - AUC = 0.290$ , compared with 0.070/0.277 for equal-weight fusion and 0.064/0.216 for naive multimodal scoring. At the same time, the real-data results also show a strong single-modality baseline on modality 2 ( $Recall@FAR = 0.105$ ), which honestly indicates that RC-AD should be understood as the most reliable multimodal budgeted strategy, not as a universal winner against every domain-specific unimodal configuration.

**Prequential evaluation protocol on controlled streams and real data.** All module-level evaluations follow a common prequential protocol [6, 7]. Each method is trained on a clean prefix, then used in test-then-update mode: a prediction or anomaly score is produced first, and only then are model statistics, buffers, or calibrated components updated from newly available information. This protocol ensures that all reported metrics remain causal and prevents leakage of future information into current control decisions. Aggregation is performed over fixed seed sets, and the repository preserves reference CSV artifacts and reproducibility instructions for the main tables [19].

The evaluation design uses both controlled streams and real data. Controlled streams are necessary because they provide known injections of modality degradation, concept drift, and anomalies, which makes causal interpretation possible. Real datasets are required because the same control logic must remain meaningful outside the synthetic setup. UCI Appliances Energy Prediction is used for forecasting with controlled missingness in test segments [15, 16], while UCI Air Quality is used for anomaly-detection demonstration under modality missingness [17, 18]. Integrated system evaluation adds a synthetic stress scenario and robust protocols in the energy and BTC/USD domains.

Table 1 summarizes the datasets, scenarios, and metrics used in the unified prequential protocol.

**Results.** The results should be read as evidence for one coordinated streaming pipeline. The reliability module provides a calibrated control scale; the fusion module turns this scale into conservative forecast reweighting; the adaptation module uses relevance shifts derived from the same pipeline to trigger bounded updates; and the anomaly-detection module reuses reliability to protect the false alarm rate budget. The empirical question is therefore not

Table 1

Evaluation settings, datasets, and primary metrics

Evaluation part	Data and scenario	Primary metrics
Controlled degradation detection	Controlled multimodal stream, 10 seeds, known degradation labels, lightweight proxy signals, and post-hoc calibration	ROC-AUC, balanced accuracy, ECE
Controlled forecasting	Clean, missingness, and mixed-degradation regimes with segment-wise modality degradation	MAE, RMSE
Controlled drift adaptation	Relevance-drift benchmark with abrupt-strong and weak-gradual-noisy scenarios under fixed update budgets	MAE, RMSE, recovery time
Controlled anomaly detection	FAR-budget benchmark for $\beta \in \{0.01, 0.05, 0.1\}$ under degradation-aware scoring	Recall@FAR, PR-AUC
Real data protocols	UCI Appliances Energy Prediction and UCI Air Quality with chronological splits and controlled missingness	MAE, Recall@FAR, PR-AUC
Integrated system effect	Synthetic stress scenario and robust energy/BTC-USD protocols	ROC-AUC, ECE, Recall@FAR, PR-AUC, MAE, RMSE

Source: compiled by the author based on the dissertation evaluation protocol and repository reference artifacts.

Table 2

Key confirmed results for online reliability estimation, RADF, budgeted micro-adaptation, and RC-AD

Module	Key evidence	Operational meaning
Online reliability estimation	Calibrated model: $ROC - AUC = 0.8624$ , balanced accuracy = 0.8415, $ECE = 0.0845$ ; uncalibrated online model: $ROC - AUC = 0.8755$ , $ECE = 0.1840$	Stable control scale for downstream rules
Reliability-adaptive dynamic fusion	Clean regime: $MAE = 0.5557$ , $RMSE = 0.6952$ ; alternating missingness: $MAE = 0.6402$ versus 0.6820; degraded UCI segments: $MAE = 63.6943$ versus 381.9107	Safe in clean windows, robust in degraded windows
Budgeted micro-adaptation	$MAE = 0.6613$ versus 0.7046 without adaptation; recovery time 157 versus 800; 9 updated parameters and 240 update samples	Restores post-drift quality under bounded cost
Reliability-constrained anomaly detection	$Recall@FAR = 0.121/0.335/0.426$ for $\beta = 0.01/0.05/0.1$ ; naive multimodal baseline = 0.034/0.103/0.182; UCI Air Quality: $Recall@FAR = 0.097$ , $PR - AUC = 0.290$	Preserves FAR control under degradation

Source: compiled by the author based on repository reference CSV artifacts for the four module groups.

whether each component can be published alone, but whether the same control interface consistently links methods, metrics, and system behavior.

Table 2 reports the key confirmed quantitative results for online reliability estimation, RADF, budgeted micro-adaptation, and RC-AD.

Table 2 makes two points explicit. First, the reliability module is the enabling condition for the rest of the pipeline: the calibrated model is not the strongest AUC-only variant, but it is the most usable control scale because it reduces calibration error. Second, the remaining modules have task-specific success criteria that match their operational role: clean-regime safety and degraded-regime robustness for RADF, bounded recovery cost for micro-adaptation, and sensitivity under a fixed false alarm rate budget for RC-AD.

Figure 3 provides a compact visual summary of the same module-level evidence, while Table 2 preserves the exact values.

The integrated system results confirm that the unified pipeline is operationally meaningful. In the representative synthetic stress scenario, the same pipeline reached  $ROC - AUC = 0.983$  and  $ECE = 0.029$  for degradation separation, while anomaly detection under the fixed false alarm rate budget  $\beta = 0.05$  achieved  $Recall@FAR = 0.311$  and  $PR - AUC = 0.287$ . The energy and BTC/USD protocols show that the system effect is not only diagnostic: the integrated system materially changes forecasting robustness when modality degradation is severe. The implementation and reference result artifacts are available in the open repository [19].

Table 3 summarizes the integrated system effect and operational characteristics, and Fig. 4 visualizes the same effect across the synthetic and real-domain protocols.

**Discussion.** The article is intentionally organized as a single narrative: reliability scale  $\rightarrow$  fusion  $\rightarrow$  micro-adaptation  $\rightarrow$  anomaly detection  $\rightarrow$  integrated streaming decision support. This ordering reflects the control logic of the dissertation. Reliability estimation comes first because it translates modality state into a calibrated operational scale. RADF then uses that scale to keep forecasting robust without harming clean-regime quality. Micro-adaptation uses the same pipeline to decide when relevance drift has become persistent enough to justify a bounded update. RC-AD uses reliability again to prevent degradation from spending the false alarm rate budget on pseudo-anomalies. The system effect appears only when these components are used together.

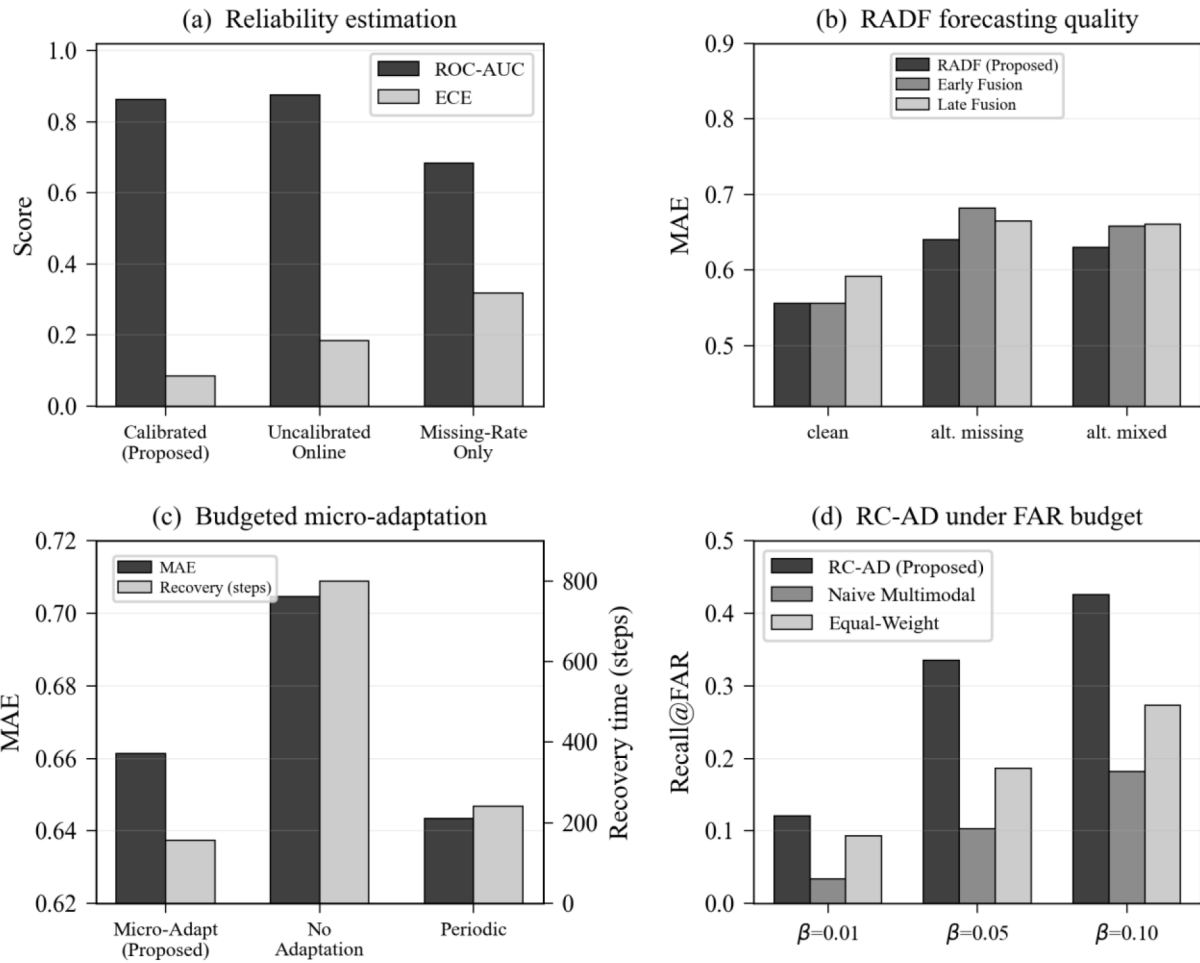


Fig. 3. Summary of module-level evidence for the reliability-driven streaming pipeline

Source: compiled by the author based on repository reference CSV artifacts for the four module groups.

Table 3

**Integrated system effect and operational characteristics**

Protocol or domain	Integrated system metrics	Comparison evidence
Synthetic stress scenario	$MAE = 0.7415$ , $RMSE = 1.0523$ , $Recall@FAR = 0.3114$ , $PR - AUC = 0.2868$	10 drift triggers, 10 adaptations, base relevance 0.0000 $\rightarrow$ 0.6337, all modules enabled
Energy-domain robust protocol ( $p = 0.90$ missingness)	Integrated system: $MAE = 0.2608$ , $RMSE = 0.3352$	Naive fusion: $MAE = 0.4015$ , $RMSE = 0.5035$ ; $MAE$ reduction about 35 %
BTC/USD robust protocol ( $p = 0.90$ missingness)	Integrated system: $MAE = 0.0239$ , $RMSE = 0.0316$	Naive fusion: $MAE = 0.3453$ , $RMSE = 0.7269$ ; $MAE$ reduction about 93 %

Source: compiled by the author based on dissertation system evaluation and repository metrics.

This interpretation also explains why the metrics are heterogeneous but compatible.  $ROC - AUC$  and  $ECE$  matter because thresholds and gates depend on the reliability scale.  $MAE$  and  $RMSE$  matter because fusion and adaptation are judged by forecasting error under streaming constraints.  $Recall@FAR$  matters because anomaly detection is constrained by operator capacity and alarm cost, not only by ranking quality. The prequential protocol links these metrics causally: all control decisions are evaluated in the same online order in which they would be executed in a deployed decision support system.

The results should not be overgeneralized. For budgeted micro-adaptation, periodic adaptation gives lower absolute error, but it does so by spending far more update resources. For reliability-constrained anomaly detection, a single-modality detector on UCI Air Quality can outperform RC-AD in recall, which indicates domain asymmetry rather than failure of the unified pipeline. These observations do not weaken the integrated claim. Instead, they clarify the precise scope of the contribution: the dissertation proposes a coordinated streaming control pipeline for

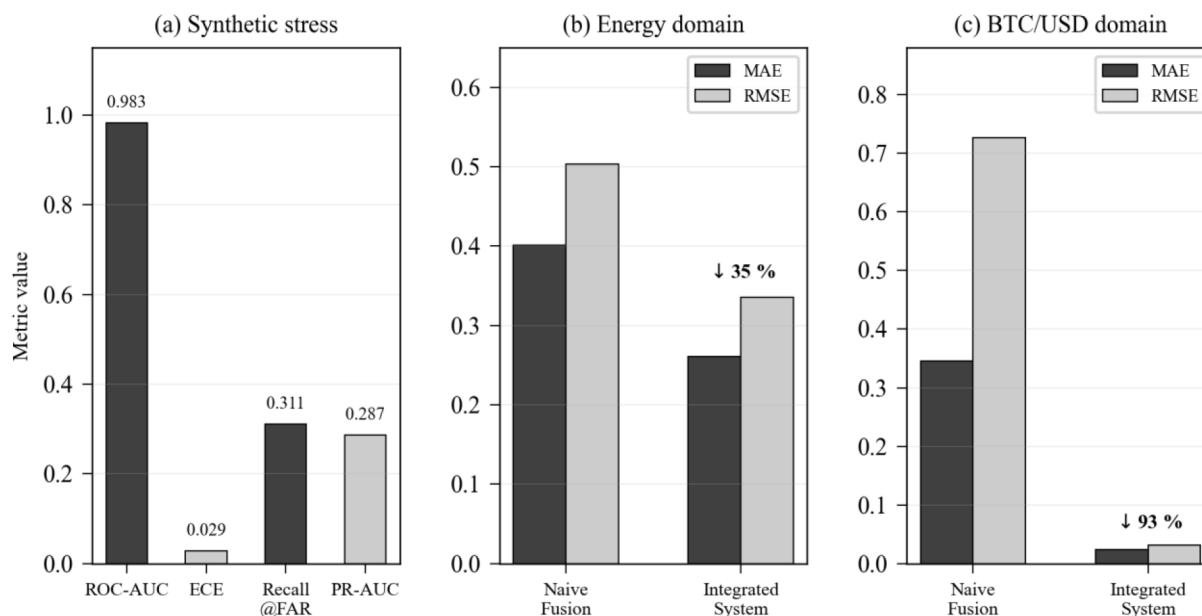


Fig. 4. Integrated system effect of the reliability-driven streaming pipeline under controlled stress and robust application protocols ( $p = 0.90$  missingness,  $\beta = 0.05$ )

Source: compiled by the author based on repository system metrics and dissertation domain-level comparisons

decision support under joint concept drift and modality degradation, not a claim that every module dominates every possible baseline under every evaluation criterion.

**Limitations and threats to validity.** Several limitations are already explicit in the dissertation and remain valid for this article. First, degradation signals are proxies rather than direct measurements of source failure, so separation quality may vary across domains. Second, error-based reliability components assume that target feedback is eventually available in the stream; under delayed labels, the control interface must rely more heavily on missingness and other causal proxies. Third, the proposed micro-adaptation is best justified for drift that changes relative modality relevance; deeper structural drift may require heavier adaptation strategies. Fourth, RC-AD is optimized for a false alarm rate budget, so different operational criteria would require different thresholding rules. Fifth, latency evidence is based on repository microbenchmarks and dissertation tables rather than on a hardware-invariant deployment benchmark; across project artifacts, exact point estimates vary, but all retained sources place reliability estimation in the low-hundreds-of-microseconds range and RADF inference in the microsecond range. Finally, external validity is bounded by the controlled streams and the real domains used in the dissertation.

**Conclusions and prospects for further research.** The manuscript presents the dissertation as one reliability-driven streaming decision support result. The integrated contribution is not merely the coexistence of four separate methods, but the fact that online reliability estimation serves as a unified control interface for reliability-adaptive dynamic fusion, drift-initiated budgeted micro-adaptation, and reliability-constrained anomaly detection under a false alarm rate budget. The prequential evaluation on controlled streams and real data shows that this pipeline is formally coherent, empirically supported, and practically relevant for streaming decision support systems operating under joint concept drift and modality degradation. The open-source implementation and reference result artifacts further support reproducibility of the reported claims [19]. Further research should address delayed-label settings, wider real-domain validation, and heavier adaptation strategies for deeper structural drift.

#### Bibliography:

1. Baltrušaitis T., Ahuja C., Morency L.-P. Multimodal Machine Learning: A Survey and Taxonomy // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2019. Vol. 41, no. 2. P. 423–443. DOI: 10.1109/TPAMI.2018.2798607
2. Lahat D., Adali T., Jutten C. Multimodal Data Fusion: An Overview of Methods, Challenges, and Prospects // Proceedings of the IEEE. 2015. Vol. 103, no. 9. P. 1449–1477. DOI: 10.1109/JPROC.2015.2460697
3. Hall D. L., Llinas J. An Introduction to Multisensor Data Fusion // Proceedings of the IEEE. 1997. Vol. 85, no. 1. P. 6–23. DOI: 10.1109/5.554205
4. Gama J., Žliobaitė I., Bifet A., Pechenizkiy M., Bouchachia A. A Survey on Concept Drift Adaptation // ACM Computing Surveys. 2014. Vol. 46, no. 4. Art. 44. DOI: 10.1145/2523813
5. Lu J., Liu A., Dong F., Gu F., Gama J., Zhang G. Learning under Concept Drift: A Review // IEEE Transactions on Knowledge and Data Engineering. 2019. Vol. 31, no. 12. P. 2346–2363. DOI: 10.1109/TKDE.2018.2876857

- 
6. Dawid A. P. Present Position and Potential Developments: Some Personal Views: Statistical Theory: The Prequential Approach // *Journal of the Royal Statistical Society. Series A (General)*. 1984. Vol. 147, no. 2. P. 278–292. DOI: 10.2307/2981683
  7. Bifet A., Gavaldà R., Holmes G., Pfahringer B. *Machine Learning for Data Streams: With Practical Examples in MOA*. Cambridge: MIT Press, 2018. 336 p.
  8. Rubin D. B. Inference and Missing Data // *Biometrika*. 1976. Vol. 63, no. 3. P. 581–592. DOI: 10.1093/biomet/63.3.581
  9. Guo C., Pleiss G., Sun Y., Weinberger K. Q. On Calibration of Modern Neural Networks // *Proceedings of the 34th International Conference on Machine Learning*. 2017. P. 1321–1330. URL: <https://proceedings.mlr.press/v70/guo17a.html>
  10. Zadrozny B., Elkan C. Transforming Classifier Scores into Accurate Multiclass Probability Estimates // *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2002. P. 694–699. DOI: 10.1145/775047.775151
  11. Niculescu-Mizil A., Caruana R. Predicting Good Probabilities with Supervised Learning // *Proceedings of the 22nd International Conference on Machine Learning*. 2005. P. 625–632. DOI: 10.1145/1102351.1102430
  12. Chandola V., Banerjee A., Kumar V. Anomaly Detection: A Survey // *ACM Computing Surveys*. 2009. Vol. 41, no. 3. Art. 15. DOI: 10.1145/1541880.1541882
  13. Fawcett T. An Introduction to ROC Analysis // *Pattern Recognition Letters*. 2006. Vol. 27, no. 8. P. 861–874. DOI: 10.1016/j.patrec.2005.10.010
  14. Hyndman R. J., Koehler A. B. Another Look at Measures of Forecast Accuracy // *International Journal of Forecasting*. 2006. Vol. 22, no. 4. P. 679–688. DOI: 10.1016/j.ijforecast.2006.03.001
  15. UCI Machine Learning Repository. Appliances Energy Prediction. 2017. URL: <https://archive.ics.uci.edu/dataset/374/appliances+energy+prediction>
  16. Candanedo L. M., Feldheim V., Deramaix D. Data Driven Prediction Models of Energy Use of Appliances in a Low-Energy House // *Energy and Buildings*. 2017. Vol. 140. P. 81–97. DOI: 10.1016/j.enbuild.2017.01.083
  17. UCI Machine Learning Repository. Air Quality. 2008. DOI: 10.24432/C5060Z. URL: <https://archive.ics.uci.edu/dataset/387/air+quality>
  18. De Vito S., Massera E., Piga M., Martinotto L., Di Francia G. On Field Calibration of an Electronic Nose for Benzene Estimation in an Urban Pollution Monitoring Scenario // *Sensors and Actuators B: Chemical*. 2008. Vol. 129, no. 2. P. 750–757. DOI: 10.1016/j.snb.2007.09.060
  19. Uzun I. Multimodal DSS: Models and Methods for Multimodal Time-Series Analysis in Intelligent Decision Support Systems – Open-Source Repository. 2026. URL: <https://github.com/illiaUzun-pub/PhD-CS-UA>

#### References:

1. Baltrušaitis, T., Ahuja, C., & Morency, L.-P. (2019). Multimodal machine learning: A survey and taxonomy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(2), 423–443. <https://doi.org/10.1109/TPAMI.2018.2798607>
2. Lahat, D., Adali, T., & Jutten, C. (2015). Multimodal data fusion: An overview of methods, challenges, and prospects. *Proceedings of the IEEE*, 103(9), 1449–1477. <https://doi.org/10.1109/JPROC.2015.2460697>
3. Hall, D. L., & Llinas, J. (1997). An introduction to multisensor data fusion. *Proceedings of the IEEE*, 85(1), 6–23. <https://doi.org/10.1109/5.554205>
4. Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014). A survey on concept drift adaptation. *ACM Computing Surveys*, 46(4), Article 44. <https://doi.org/10.1145/2523813>
5. Lu, J., Liu, A., Dong, F., Gu, F., Gama, J., & Zhang, G. (2019). Learning under concept drift: A review. *IEEE Transactions on Knowledge and Data Engineering*, 31(12), 2346–2363. <https://doi.org/10.1109/TKDE.2018.2876857>
6. Dawid, A. P. (1984). Present position and potential developments: Some personal views: Statistical theory: The prequential approach. *Journal of the Royal Statistical Society. Series A (General)*, 147(2), 278–292. <https://doi.org/10.2307/2981683>
7. Bifet, A., Gavaldà, R., Holmes, G., & Pfahringer, B. (2018). *Machine learning for data streams: With practical examples in MOA*. MIT Press.
8. Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3), 581–592. <https://doi.org/10.1093/biomet/63.3.581>
9. Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017). On calibration of modern neural networks. In *Proceedings of the 34th International Conference on Machine Learning* (pp. 1321–1330). <https://proceedings.mlr.press/v70/guo17a.html>
10. Zadrozny, B., & Elkan, C. (2002). Transforming classifier scores into accurate multiclass probability estimates. In *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 694–699). <https://doi.org/10.1145/775047.775151>

- 
11. Niculescu-Mizil, A., & Caruana, R. (2005). Predicting good probabilities with supervised learning. In *Proceedings of the 22nd International Conference on Machine Learning* (pp. 625–632). <https://doi.org/10.1145/1102351.1102430>
  12. Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys*, 41(3), Article 15. <https://doi.org/10.1145/1541880.1541882>
  13. Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861–874. <https://doi.org/10.1016/j.patrec.2005.10.010>
  14. Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4), 679–688. <https://doi.org/10.1016/j.ijforecast.2006.03.001>
  15. UCI Machine Learning Repository. (2017). *Appliances Energy Prediction*. <https://archive.ics.uci.edu/dataset/374/appliances+energy+prediction>
  16. Candanedo, L. M., Feldheim, V., & Deramaix, D. (2017). Data driven prediction models of energy use of appliances in a low-energy house. *Energy and Buildings*, 140, 81–97. <https://doi.org/10.1016/j.enbuild.2017.01.083>
  17. UCI Machine Learning Repository. (2008). *Air Quality*. <https://doi.org/10.24432/C5060Z>
  18. De Vito, S., Massera, E., Piga, M., Martinotto, L., & Di Francia, G. (2008). On field calibration of an electronic nose for benzene estimation in an urban pollution monitoring scenario. *Sensors and Actuators B: Chemical*, 129(2), 750–757. <https://doi.org/10.1016/j.snb.2007.09.060>
  19. Uzun, I. (2026). *Multimodal DSS: Models and methods for multimodal time-series analysis in intelligent decision support systems – Open-source repository*. <https://github.com/illiaUzun-pub/PhD-CS-UA>

Дата першого надходження статті до видання: 22.03.2026

Дата прийняття статті до друку після рецензування: 20.04.2026

Дата публікації (оприлюднення) статті: 30.05.2026