Multisensor Data Fusion for Real-Time Water Quality Monitoring and Aquaculture Management in Floating Marine Systems

Seamus Kerney
Department of Computer Science, University of Central Florida, Orlando, FL, USA.
seamusk@ucf.edu

Abstract

Floating marine aquaculture exposes water-quality sensing to waves, biofouling, variable salinity, and intermittent power and communication, so single-sensor alarms and static fusion rules can produce late or biased decisions. This manuscript develops a reliability-aware hierarchical Bayesian fusion framework for real-time water-quality monitoring and management in floating marine systems. The framework represents temperature, salinity, dissolved oxygen, pH, un-ionized ammonia, and turbidity as a coupled state, combines measurement-level redundancy with model-based state estimation, and allows covariance and credibility weights to evolve with residual statistics, calibration age, and cleaning actions. A residual-based anomaly detector separates sensor faults from genuine hydrochemical events, while a decision layer converts posterior states into feed, aeration, and harvesting actions under explicit loss functions. The central contribution is not a new sensor but an integrated information-processing design that treats reliability as a dynamic state rather than a fixed tuning parameter. Analytical comparison shows why exponential reliability penalties and robust residual tests outperform static averaging when fouling is spatially clustered and when pH-temperature interactions alter ammonia toxicity. The framework provides testable design rules for edge implementation, uncertainty reporting, and maintenance scheduling. It also clarifies where learned models are appropriate as residual and event classifiers rather than as replacements for conservative water-quality state models. The resulting architecture supports defensible aquaculture decisions under communication and energy constraints without requiring a dense sensor grid.

Keywords

multisensor fusion; aquaculture management; water quality; Bayesian state estimation; sensor reliability; real-time monitoring

1. Introduction

Floating marine systems, including offshore cages, floating raceways, and hybrid aquaculture platforms, place water-quality instrumentation in an environment that is physically hostile to measurement. Wireless sensor networks provide the basic communication and sampling substrate for such systems [1], but network connectivity alone does not resolve measurement ambiguity. Real-time remote monitoring has long been used in coastal and marine applications to detect episodic water-quality change [2], yet the interpretation of concurrent temperature, salinity, dissolved oxygen, pH, ammonia, and turbidity signals remains difficult when sensors degrade at different rates and when events propagate through a moving water body.

Multisensor data fusion offers a principled response because it combines redundant, complementary, and temporally displaced observations into a state estimate with explicit uncertainty [3]. Recursive estimation supplies the formal machinery for updating that estimate as new observations arrive [4], and nonlinear filtering addresses the non-Gaussian and nonlinear relationships common in environmental sensing [5]. Evidence-theoretic methods are useful when individual sensors report categorical quality states rather than calibrated numeric values [6]. Temporal learning models such as long short-term memory networks can capture lagged dependence in water-quality series [7], random forests can model nonlinear tabular relationships among sensor and management variables [8], and Gaussian processes can represent smooth but uncertain environmental fields [9]. These methods are valuable, but they do not by themselves answer the engineering question that governs floating marine deployment: how should fusion logic change when a sensor's reliability is itself a time-varying, partly observable state?

The knowledge gap addressed here concerns the coupling between sensor reliability and aquaculture decision risk. Conventional fusion schemes often assume stationary measurement noise, fixed credibility weights, or centralized computation. Floating systems violate these assumptions through biofouling, wave-induced fouling and occlusion, calibration drift, salinity-dependent electrode behavior, biofouling-related flow obstruction, and intermittent links. Internet-of-Things deployments in constrained living environments face related challenges of heterogeneous devices, energy limits, and service continuity [10]. For aquaculture, the consequence is not merely a noisy estimate but a management error: aeration may be delayed, feeding may be excessive, or a harvest decision may be taken under unrecognized uncertainty.

This paper develops a reliability-aware hierarchical Bayesian fusion framework for real-time water-quality monitoring and aquaculture management in floating marine systems. The contribution is an integrated design that represents water-quality state, sensor bias, and sensor credibility in a single inference architecture; uses residual and maintenance information to update measurement covariance; distinguishes sensor faults from genuine hydrochemical events; and maps posterior state uncertainty into management actions through a loss function. The emphasis is on methodological coherence and analytical consequences rather than on reporting a field deployment. The framework is designed for edge implementation and is intentionally conservative where public health, animal welfare, and stock loss are at stake.

2. System Requirements and State-Space Problem Formulation

A floating marine aquaculture system must estimate a water-quality state that is only partially observable from any single point. The state vector at discrete time k is defined as x_k = [T, S, DO, pH, NH3, TU]^T, where T is temperature, S is salinity, DO is dissolved oxygen, NH3 is un-ionized ammonia concentration, and TU is turbidity. Un-ionized ammonia is not directly measured by most field sensors; it is inferred from total ammonia, pH, temperature, and salinity through equilibrium relationships [11]. This inference matters because the toxic fraction can change sharply with pH even when total ammonia remains nearly constant. Temperature and salinity also affect dissolved oxygen solubility and sensor response, while turbidity interacts with optical measurements and indicates particulate loading associated with feeding, waste, and algal activity.

The monitoring system is organized into three functional tiers. The node tier performs local validation, time stamping, and basic redundancy checks. The edge tier runs state estimation, reliability weighting, and anomaly detection near the floating platform. The

management tier converts posterior state distributions into alarms and operational actions. This hierarchy reduces dependence on continuous wide-area communication and allows time-critical aeration decisions to be made locally. Because floating systems may have limited maintenance access, the fusion logic must explicitly track calibration age and cleaning history rather than assuming that all sensors are equally trustworthy after deployment.

The management context imposes asymmetric costs. A false negative on dissolved oxygen depletion can cause rapid stock mortality, whereas a false positive may trigger unnecessary aeration and energy use. High un-ionized ammonia under warm, alkaline conditions is similarly urgent, while turbidity excursions may be tolerable over short intervals. The estimation problem therefore cannot be reduced to minimizing mean squared sensor error. It requires calibrated uncertainty, fault isolation, and decision rules that account for the relative cost of missed events and unnecessary interventions.

3. Reliability-Aware Hierarchical Fusion Framework

3.1 Sensor Measurement and Water-Quality State Models

Each sensor observation is represented as a noisy, biased function of the latent water-quality state. For sensor i at time k, the measurement model is

$$y_{i,k} = h_i(x_k, \theta_i) + b_{i,k} + \varepsilon_{i,k}$$

where y_{i,k} is the observed value, h_i is the sensor-specific response function, θ_i contains calibration and environmental compensation parameters, b_{i,k} is a slowly varying bias associated with drift or fouling, and ε_{i,k} is zero-mean measurement noise conditional on the sensor's current quality state. The function h_i may be linear for a calibrated temperature probe, mildly nonlinear for a pH electrode, or state-dependent for an optical turbidity sensor whose window is progressively occluded. The bias term is essential in floating systems because biofouling often produces persistent offset rather than random error. The model also permits observations to be transformed into common units before fusion, so that a pH-derived ammonia estimate and an ion-selective ammonia measurement can be compared on the same scale.

The latent state evolves according to a process model that combines local transport, reaction, and management forcing. A discrete-time approximation is

$$x_{k+1} = F\,x_k + B\,u_k + w_k, w_k \sim N(0, Q)$$

where F encodes local persistence, dilution, and linearized biogeochemical interactions over the sampling interval, B maps management inputs u_k such as feeding, aeration, or water exchange into state changes, and w_k represents process uncertainty with covariance Q. In a floating system, F should not be treated as a universal constant; it depends on flow, stratification, and farm loading. The analytical role of the equation is to make those dependencies explicit and to prevent a fusion algorithm from interpreting a real environmental transition as a sensor fault. When F is poorly known, its uncertainty is absorbed into Q, which increases reliance on measurements but also widens predictive intervals. This trade-off is preferable to a hidden assumption of stationary dynamics.

3.2 Bayesian Fusion and Dynamic Reliability Weighting

The fusion objective is the posterior distribution of the water-quality state given all observations up to time k. Recursive Bayesian updating provides the general form

$$p(x_k|Y_{1:k}) \propto p(y_k|x_k)p(x_k|Y_{1:k-1})$$

where Y_{1:k} denotes the observation history and p(x_k|Y_{1:k-1}) is the predictive distribution obtained from the process model. The proportionality hides the normalizing constant, which is needed for absolute likelihoods but not for comparing relative posterior mass over state values. This formulation accommodates nonlinear measurement functions through local linearization or sampling. It also provides a natural place to insert reliability information: a suspect sensor should increase the covariance of its likelihood rather than being discarded outright, because even a drifting sensor may carry information about direction of change.

For redundant measurements that have been transformed to a common state component, a local fusion step can be written as a precision-weighted combination,

$$\hat{x}_k = (\sum_{i=1}^{m} w_{i,k} R_{i,k}^{-1} y_{i,k})(\sum_{i=1}^{m} w_{i,k} R_{i,k}^{-1})^{-1}$$

where m is the number of available measurements for that component, R_{i,k} is the measurement covariance, w_{i,k} is a scalar credibility weight, and $\hat{x}_k$ is the fused estimate. The weight is not a fixed expert score. It is reduced when a sensor has a large normalized residual, an expired calibration, a recent cleaning interval that has not yet stabilized, or a communication delay that makes its observation temporally inconsistent with the state prediction. The covariance R_{i,k} should be inflated jointly with a decrease in w_{i,k}; otherwise the fusion equation can overstate precision even when the weight is small. This distinction prevents a common failure mode in which a degraded sensor is down-weighted in the numerator but still contributes an unrealistically narrow covariance in the denominator.

Floating marine systems also require spatial reasoning. If two dissolved oxygen sensors on the same cage disagree, the disagreement may reflect vertical stratification, wake effects, or sensor fouling. A residual that is large for one sensor but consistent with a known current direction should not be treated identically to an isolated residual under well-mixed conditions. The framework therefore permits a context vector c_k containing current speed, wind direction, feeding state, and sensor depth to modulate the prior probability of a sensor fault. This modulation is a modeling choice, not an empirical result; its justification is physical. A fouling event on the upstream side of a cage is more plausible than a simultaneous identical bias on all sensors, whereas a genuine hypoxic plume may affect multiple depths with a lag consistent with advection.

3.3 Anomaly Detection and Calibration-State Tracking

Anomaly detection separates three cases: a sensor fault, a genuine water-quality event, and a model inadequacy in which the process model fails to represent observed change. The normalized innovation for sensor i is

$$d_{i,k}^2 = r_{i,k}^T S_{i,k}^{-1} r_{i,k}$$

where r_{i,k} = y_{i,k} - h_i($\hat{x}_k^-$, θ_i) is the innovation, $\hat{x}_k^-$ is the predicted state before measurement update, and S_{i,k} = H_i P_k^- H_i^T + R_{i,k} is the innovation covariance with H_i the measurement Jacobian and P_k^- the predicted state covariance. Under linear-Gaussian assumptions, d_{i,k}^2 follows a chi-square distribution with degrees of freedom equal to the measurement dimension. In practice, the threshold should be set conservatively and interpreted alongside cross-sensor consistency [20]. A single large innovation is evidence for further checking, not proof of a fault. Two or more innovations that remain large after covariance inflation indicate either a real event or a shared model error.

Reliability is tracked as a bounded quality penalty q_{i,k}. A proposed update is

$$q_\{i,k+1\} = min\{1, max\{0, q_\{i,k\} + \eta_1 1(d_\{i,k\} > \tau_i) - \eta_2 u_\{i,k\}\}\}$$

where q_{i,k} is the penalty, 1(·) is an indicator function, τ_i is a sensor-specific residual threshold, u_{i,k} is a cleaning or recalibration action, and η_1 and η_2 are positive adaptation rates. The equation is a bookkeeping device for maintenance state, not a claim that fouling follows a universal law. Its value is architectural: it forces the fusion layer to remember that a sensor was cleaned recently, that a calibration certificate has aged, or that repeated residuals have occurred. The penalty should increase the effective measurement covariance and reduce the credibility weight in the local fusion step. When a cleaning action is performed, the penalty decreases, but the sensor should not immediately return to full credibility because post-cleaning stabilization and electrode repolarization can take time.

4. Aquaculture Management Decisions Under Uncertainty

The fusion layer is useful only if it supports decisions that are timely and defensible. Let a denote a management action, such as increasing aeration, reducing feed, initiating a partial harvest, or taking no action. Let s denote a latent risk state, such as adequate oxygen, incipient hypoxia, ammonia stress, or sensor-fault uncertainty. The posterior p(s|Y_{1:k}) is derived from the state posterior by thresholding or by a learned risk classifier that uses fused state, trend, and uncertainty. A decision rule can be written as

$$a* = argmin_a \sum_s C(a,s)\, p(s|Y_\{1:k\})$$

where C(a,s) is the loss incurred when action a is taken in state s. The loss matrix is management-specific and should reflect stock value, energy cost, animal welfare, and regulatory constraints. The equation makes the asymmetry of aquaculture risk explicit. For example, the loss of failing to aerate during a high-confidence hypoxia event is much larger than the cost of brief preventive aeration. By contrast, repeated unnecessary harvest actions in response to isolated turbidity spikes are undesirable and should be penalized through the loss function.

The posterior state also supplies an uncertainty gate for action. Aeration may be triggered when the upper credible bound of the hypoxia probability exceeds a management threshold, while a harvest decision should require both a high probability of severe risk and evidence that the event is spatially coherent across sensors. This gating reduces the effect of single-sensor artifacts without ignoring them. The decision layer can be optimized subject to operational constraints such as maximum aerator power, feed inventory, and crew availability; convex formulations are appropriate when the loss and constraints are expressed in linear or quadratic form [17]. Statistical learning can then be used to estimate event classifiers or loss-sensitive thresholds from historical records, but the final decision rule remains interpretable because its inputs are fused state probabilities and explicit costs [18].

The management layer should also expose maintenance decisions. A high quality penalty q_{i,k} that persists across several sampling intervals implies that cleaning or recalibration has greater expected value than further algorithmic down-weighting. The same posterior machinery can compare the expected loss of operating with a degraded sensor against the cost of an inspection. In floating marine systems, maintenance scheduling is constrained by weather windows and vessel availability, so the practical output of fusion is not only a water-quality estimate but also a ranked list of sensor and intervention needs.

5. Technical Discussion and Boundary Conditions

The proposed framework differs from static weighted averaging in its treatment of reliability as a state variable. Static averaging performs well when sensor errors are independent and stationary, but it fails when biofouling creates correlated bias across sensors mounted at similar depths. Dynamic covariance inflation and quality penalties reduce the influence of such sensors without discarding them completely. The approach also differs from purely data-driven prediction because it retains a physically interpretable process model. Learned models are most defensible as residual classifiers, event detectors, or short-horizon forecast corrections rather than as unconstrained replacements for water-quality dynamics. This distinction matters when a model encounters conditions outside its training distribution, such as an unusual tide, a storm, or a change in feed formulation.

Several boundary conditions limit the framework. The process model F and input matrix B must be specified from local hydrodynamics and farm operations; a poorly specified F can cause the filter to attribute real environmental change to sensor bias. The loss matrix C(a,s) is a management instrument and cannot be inferred from sensor data alone. The anomaly detector assumes that innovation covariance is at least approximately calibrated; persistent underdispersion or overdispersion requires covariance adaptation. Finally, the quality-penalty update is a heuristic maintenance model, not a validated fouling law. Its parameters should be chosen conservatively and reviewed against service records rather than tuned to maximize short-term alarm suppression.

The framework also has implications for edge implementation. The local fusion step is computationally light because it requires only weighted matrix operations and residual tests. Full Bayesian updating may be performed at the edge using linearized or sampling-based filters, while heavier learned event classifiers can be run intermittently or at a shore node. Communication should transmit sufficient statistics rather than raw high-frequency streams: fused state, covariance, innovation summaries, and quality penalties. This design reduces bandwidth and preserves the information needed for audit. It also makes the system more resilient to link interruption, because the edge can continue operating with the last validated model and a conservative uncertainty inflation.

6. Conclusion

Floating marine aquaculture requires water-quality decisions that remain reliable when sensors are exposed to fouling, drift, wave disturbance, and intermittent communication. This paper developed a reliability-aware hierarchical Bayesian fusion framework that represents water-quality state, sensor bias, and sensor credibility in a coordinated inference architecture. The framework combines measurement-level redundancy with model-based state estimation, updates covariance and credibility from residual and maintenance information, distinguishes sensor faults from genuine hydrochemical events, and converts posterior distributions into management actions through explicit loss functions. Its main analytical contribution is the treatment of reliability as a dynamic state that must be estimated and maintained, not as a fixed weight assigned at deployment. The design is conservative where aquaculture risk is asymmetric, interpretable where management decisions must be justified, and light enough for edge implementation. Future work can evaluate specific filter approximations, learned residual classifiers, and maintenance policies under controlled conditions, but the methodological logic presented here provides a coherent basis for real-time fusion in floating marine systems.

References

1. Akyildiz, I. F., Su, W., Sankarasubramaniam, Y., & Cayirci, E. (2002). Wireless sensor networks: A survey. Computer Networks, 38(4), 393–422.

2. Glasgow, H. B., Burkholder, J. M., Reed, R. E., Lewitus, A. J., & Kleinman, J. E. (2004). Real-time remote monitoring of water quality: A review of current applications, and advancements in sensor, telemetry, and computing technologies. Journal of Experimental Marine Biology and Ecology, 300(1–2), 409–448.

3. Hall, D. L., & Llinas, J. (1997). An introduction to multisensor data fusion. Proceedings of the IEEE, 85(1), 6–23.

4. Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. Journal of Basic Engineering, 82(1), 35–45.

5. Julier, S. J., & Uhlmann, J. K. (1997). A new extension of the Kalman filter to nonlinear systems. Proceedings of SPIE, 3068, 182–193.

6. Shafer, G. (1976). A mathematical theory of evidence. Princeton University Press.

7. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780.

8. Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.

9. Rasmussen, C. E., & Williams, C. K. I. (2006). Gaussian processes for machine learning. MIT Press.

10. Liu, J., Hu, H., Xu, W., & Luo, D. (2024). Internet of things challenges and future scope for enhanced living environments. In Advances in computers (Vol. 133, pp. 201-246). Elsevier.

11. Stumm, W., & Morgan, J. J. (1996). Aquatic chemistry: Chemical equilibria and rates in natural waters (3rd ed.). Wiley.

12. Colt, J. (2006). Water quality requirements for reuse systems. Aquacultural Engineering, 34(3), 143–156.

13. Timmons, M. B., & Ebeling, J. M. (2013). Recirculating aquaculture (3rd ed.). Ithaca Publishing.

14. Føre, M., Frank, K., Norton, T., Svendsen, E., Alfredsen, J. A., Dempster, T., Eguiraun, H., Watson, W., Stahl, A., Sunde, L. M., Schellewald, C., Skøien, K. R., Alver, M. O., & Berckmans, D. (2018). Precision fish farming: A new framework to improve production in aquaculture. Biosystems Engineering, 173, 176–193.

15. Bar-Shalom, Y., Li, X. R., & Kirubarajan, T. (2001). Estimation with applications to tracking and navigation. Wiley.

16. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.

17. Boyd, S., & Vandenberghe, L. (2004). Convex optimization. Cambridge University Press.

18. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: Data mining, inference, and prediction (2nd ed.). Springer.

19. Snoeyink, V. L., & Jenkins, D. (1980). Water chemistry. Wiley.

20. Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. ACM Computing Surveys, 41(3), 1–58.