Outlier Detection of Coastal Vertical Observation Data using AI

Department of Ocean Science, Inha University
Young min Choi
(ymchoi@mpl.inha.ac.kr)

Introduction

Study Area



	Position	Tide range	Mean Depth (MSL)	Temperature	Salinity
New Port, Incheon	37.30~37.36°N 126.50~126.67°E	7~11m	11.9m	15.42°C	28.87

Background

CTD



ADCP



Multibeam

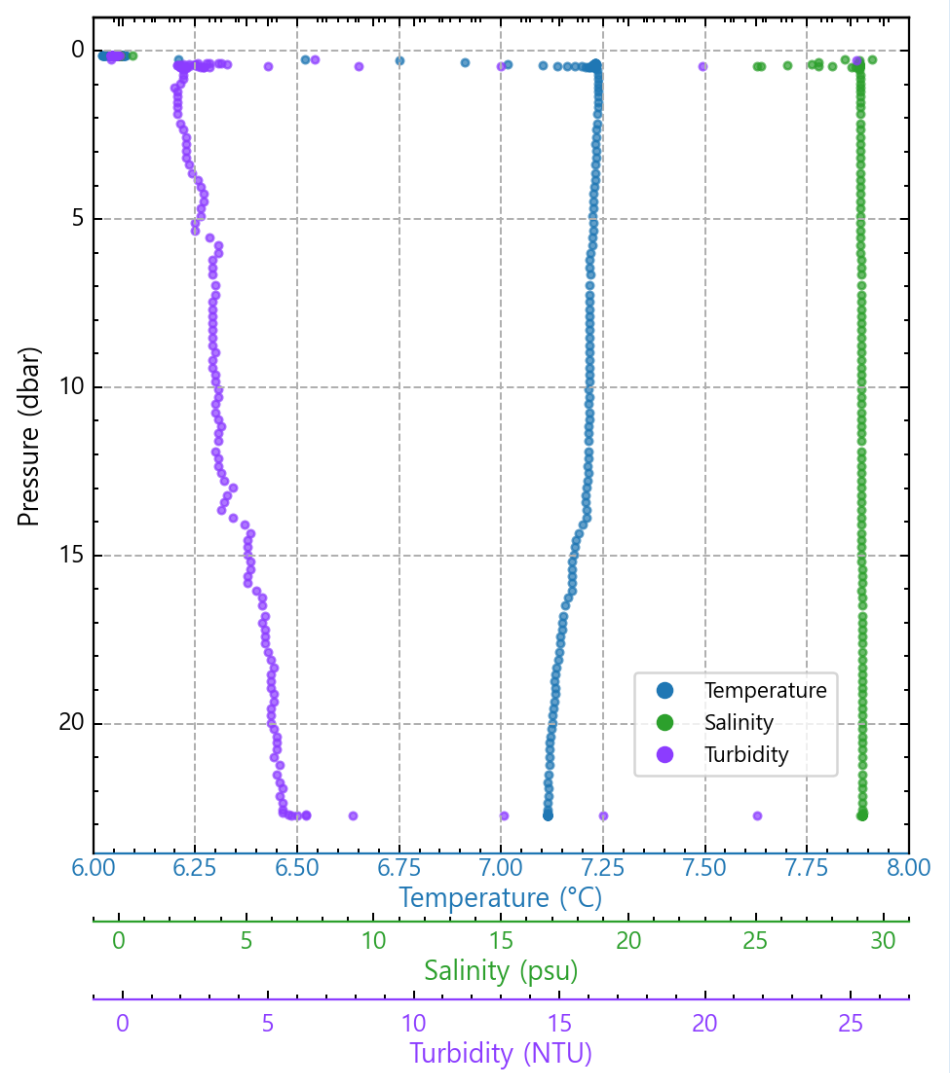


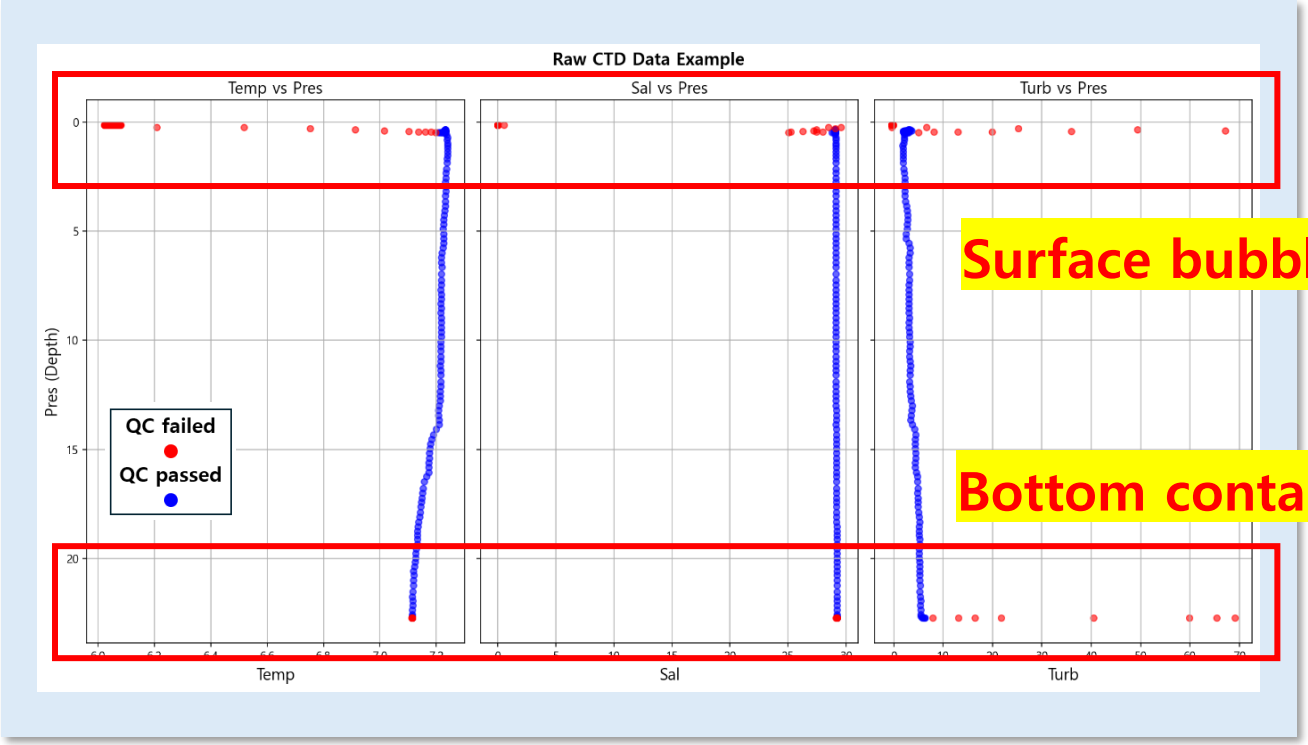
Korea Ministry of Oceans and Fisheries

→ **Ocean observation are conducted with a variety of sensor-based methods.**

→ **Because sensors operate in field environments, they are bound to contain errors.**

Coastal observation outlier





→ Raw data profiles reveal repeated spikes at the surface and near-bottom

→ The main causes are surface stabilization delay and bubbles, and bottom contact and resuspended sediment.

Previous study (Advances in QC)

① Standard QC (Tradition)
- Range / Spike / Gradient Test
- Real time flag
- Fast & Consistent

② Statistics QC
- Linear/curve fitting
- Cubic spline QC
- GTSPP / EuroGOOS

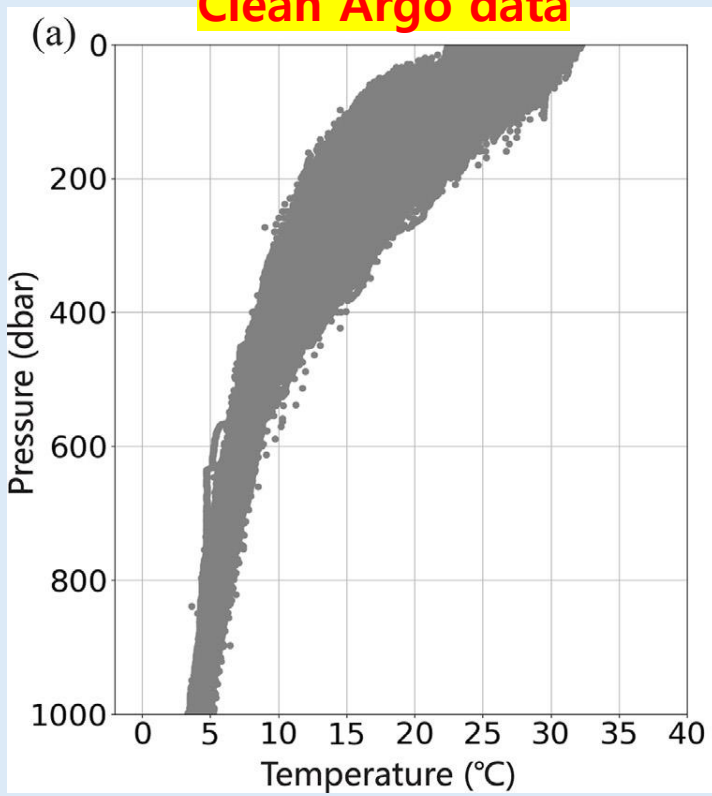
③ AI QC
- Unsupervised AI
- Multivariate ML / DL
- Active Learning

Joint IOC/ICES (2010)	→ Standard real-time quality control for profiles defining rule-based test.
Argo Data Management Team (2010)	→ Two-tier QC for Argo profiles : real-time and delayed review.
Good et al. (2013)	→ Global EN4 T-S dataset with monthly analyses and strengthened QC
Zhou et al. (2018)	→ ARIMA-based automated QC for sea floor observatory data
Li et al. (2023)	→ ODEAL framework, active learning QC for Argo that boosts efficiency and accuracy.
Sui et al. (2024)	→ FFT-enhanced transformer for Argo profile that improves automatic outlier detection.

→ **Most QC advances are Argo-centric; coastal CTD needs a tailored pipeline that understands surf/bott artifacts.**

Necessary of QC

Clean Argo data



Argo data containing error

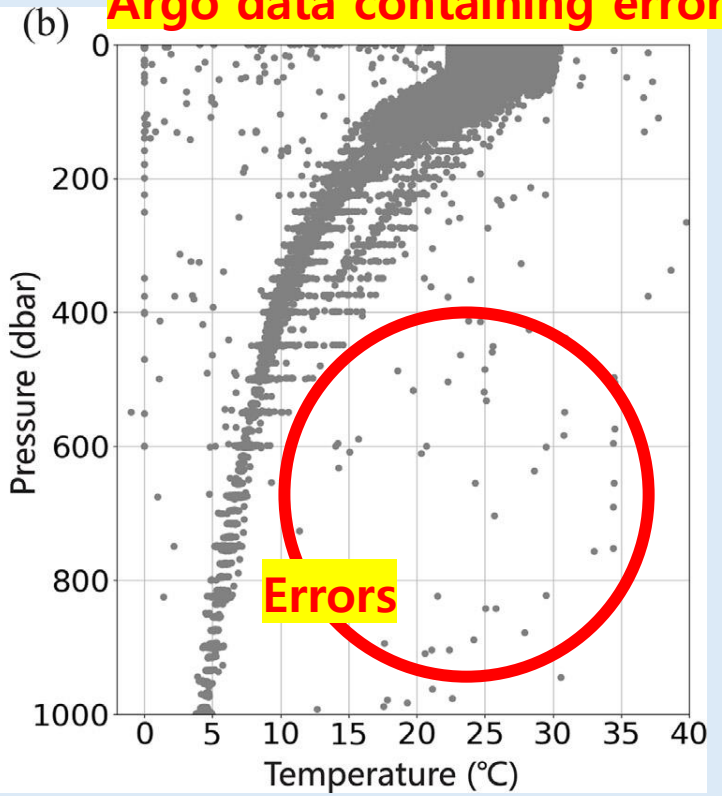
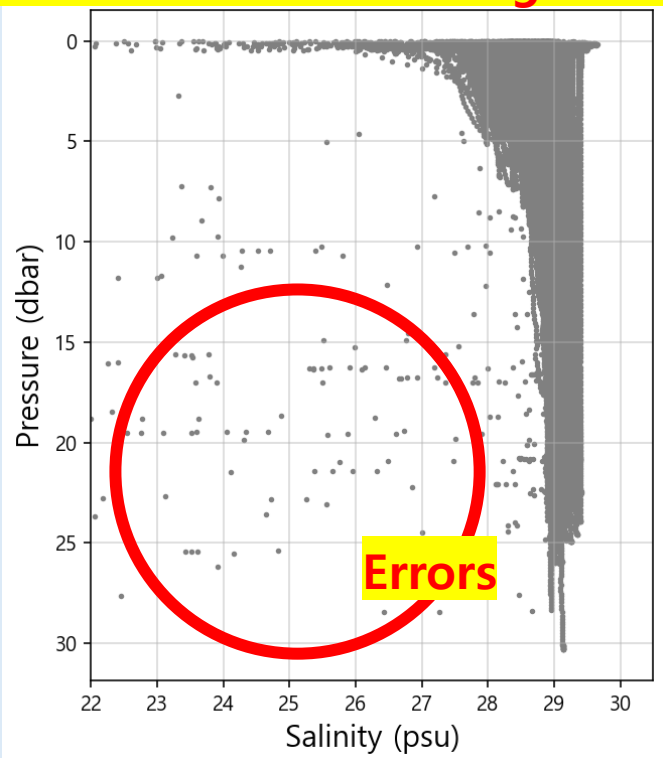


Fig 1. Zhang et al. (2023)

Coastal data containing error



→ The Argo float is a well-known platform, and even with Argo floats, false readings are frequent.

→ Coastal observations that demonstrate physically impossible natural phenomena.

→ QC is essential

Limit of Tradtion QC

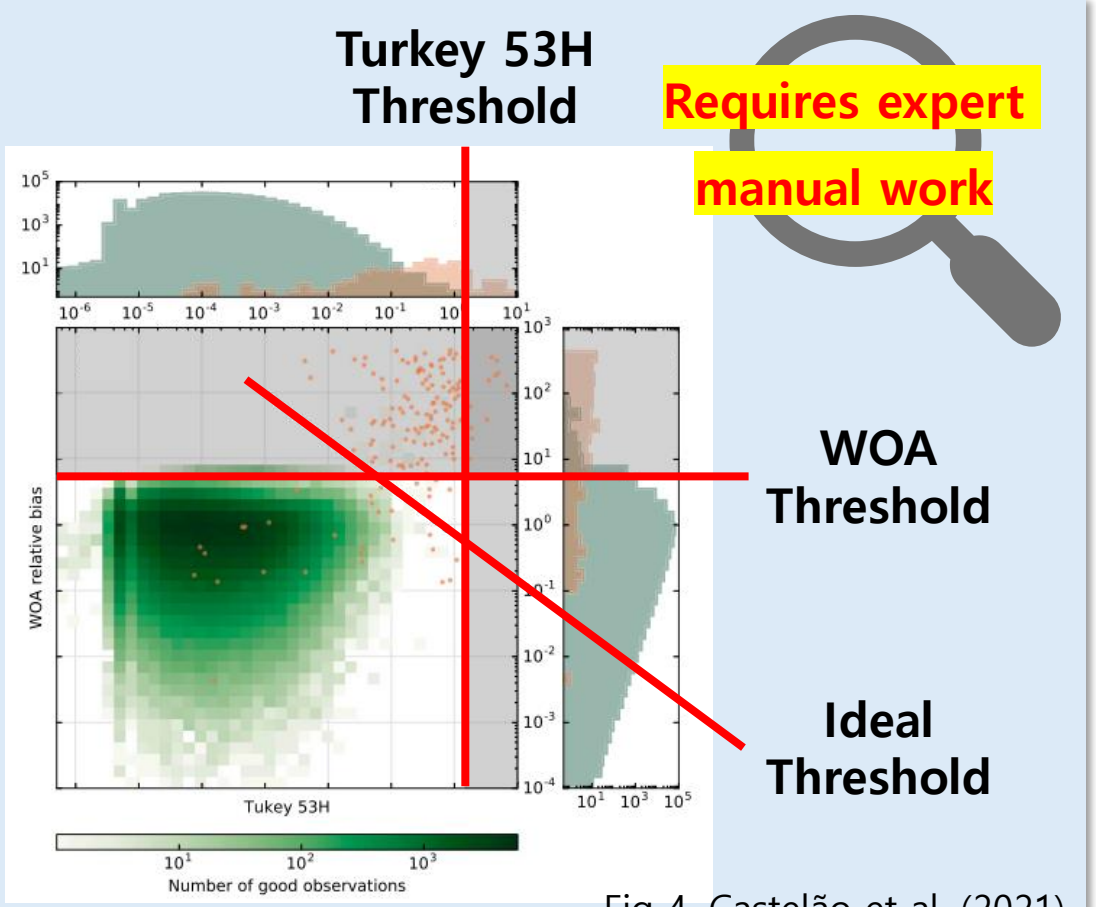


Fig 4. Castelão et al. (2021)

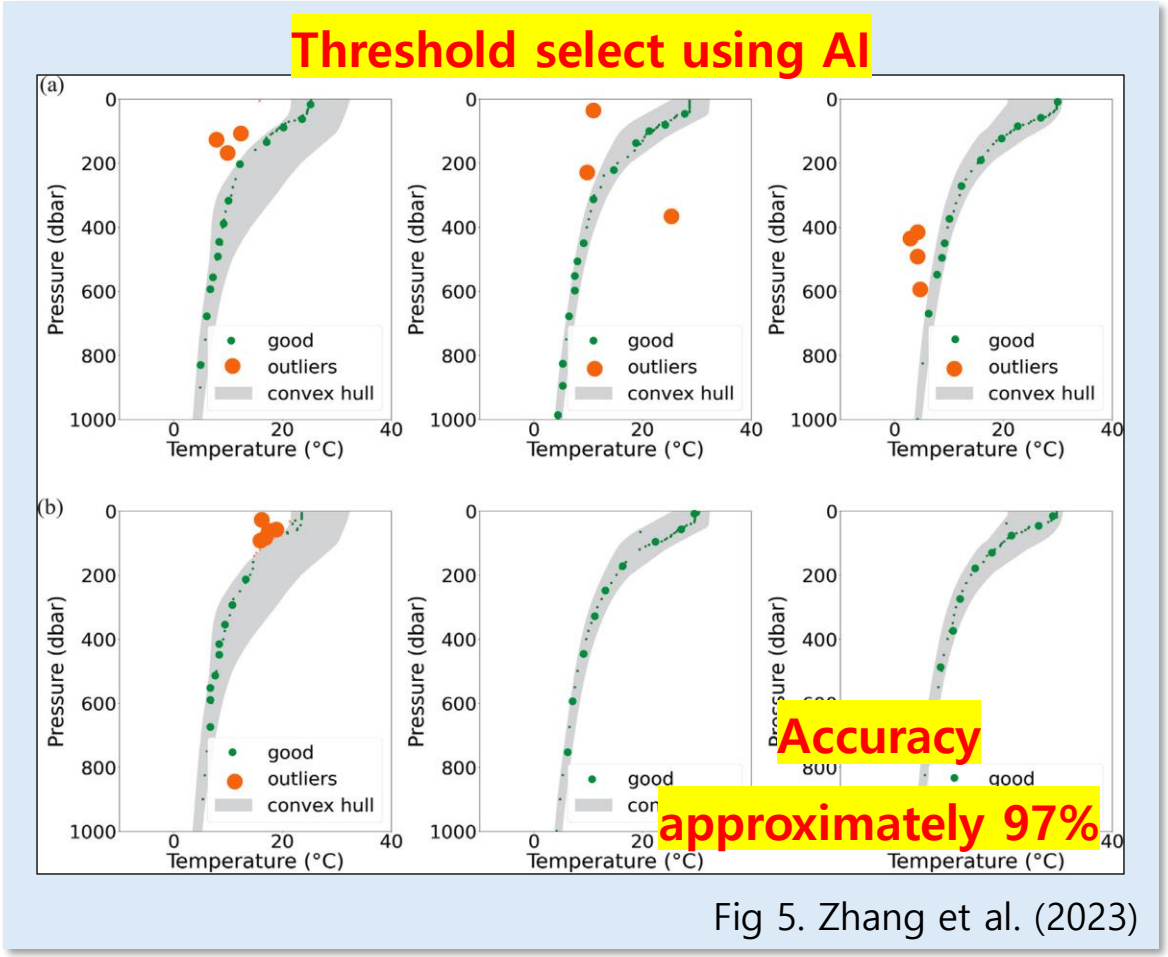


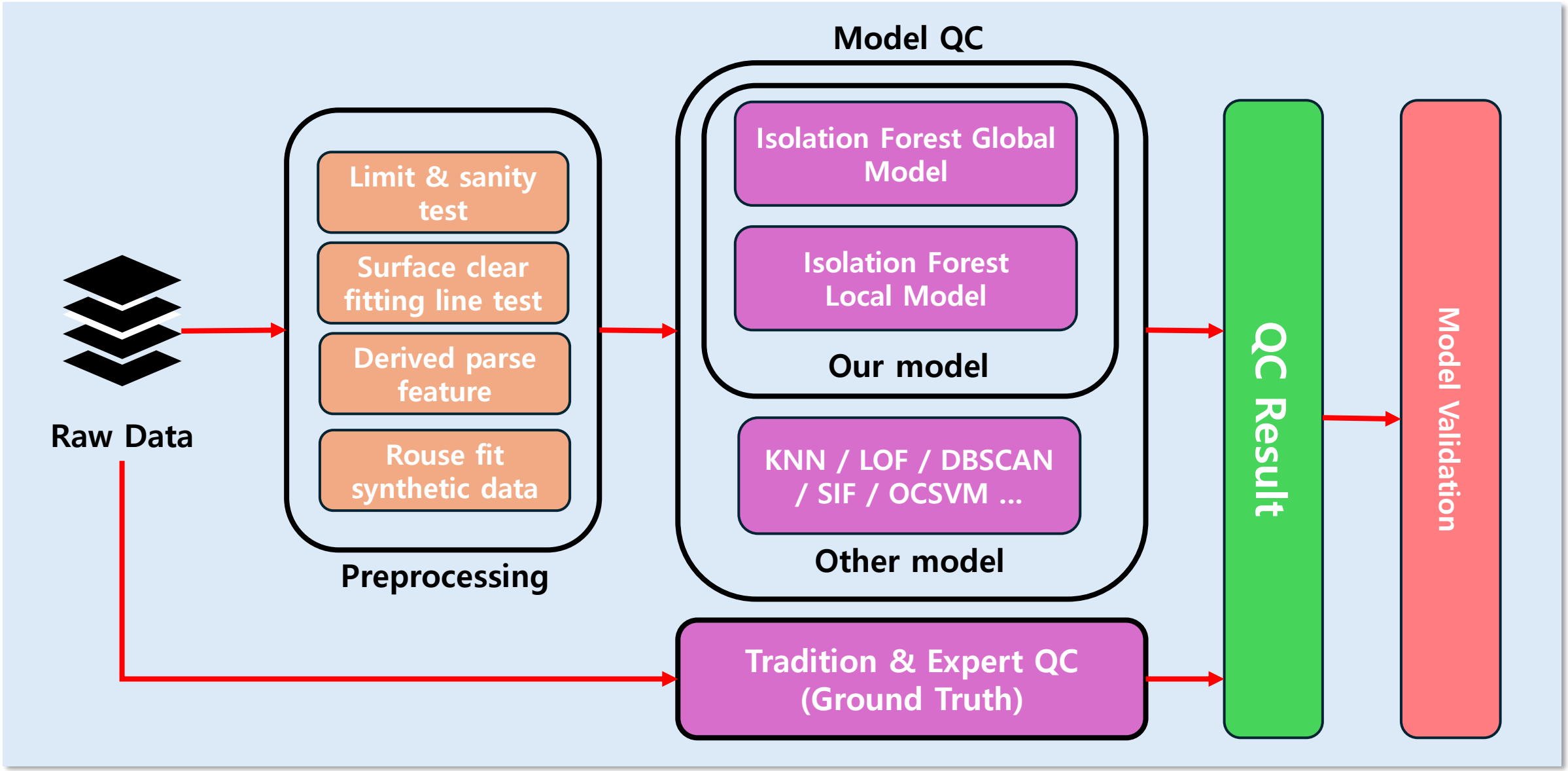
Fig 5. Zhang et al. (2023)

→ Traditional QC method have limitations in detecting errors that are similar to the correct data.

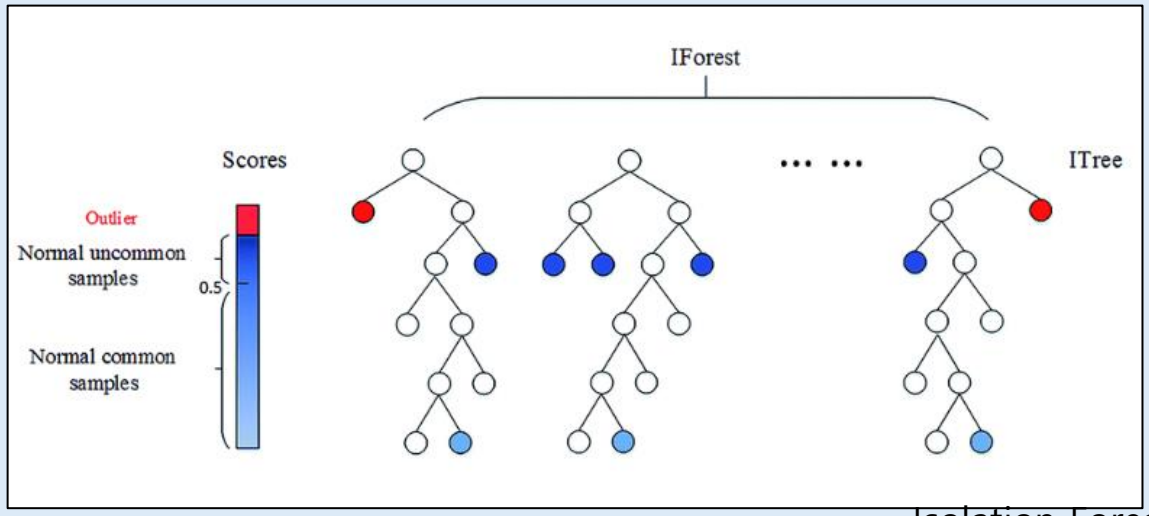
→ AI-based QC can make non-linear thredshold and detect unique error more sensitively than traditional Qcmethods, eliminating subjectivity and providing greater efficiency.

Method

Method schematic diagram



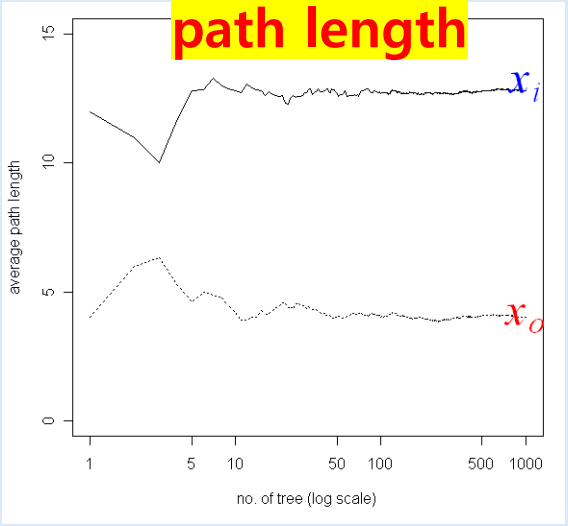
Main Algorithm (Isolation Forest)

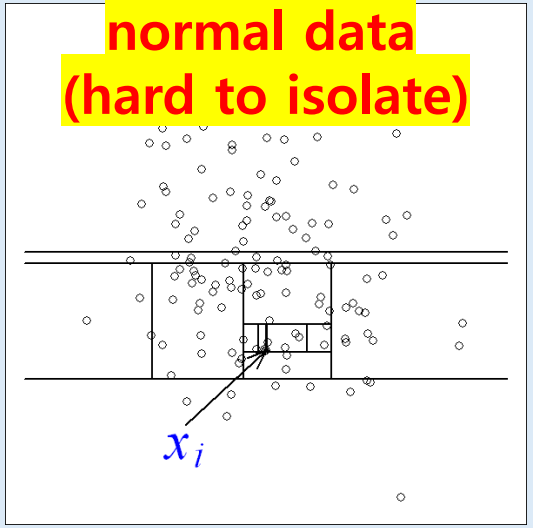


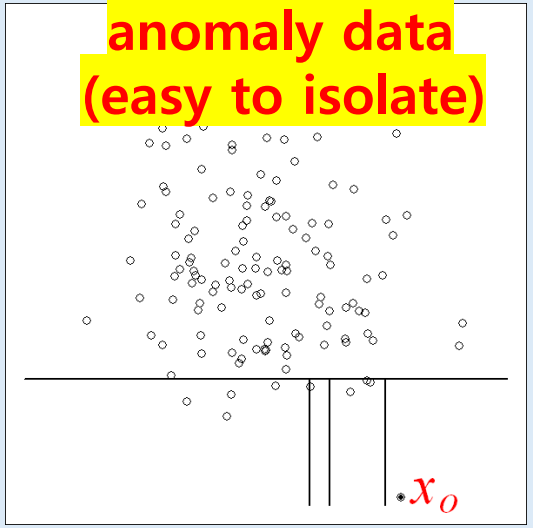
Isolation Forest

→ Isolation Forest's main idea is that anomaly can be more easily isolated compared to normal data.

→ Each decision tree in the forest splits the data to isolate observations, and the number of steps required is used to measure how much of an outlier a point is.







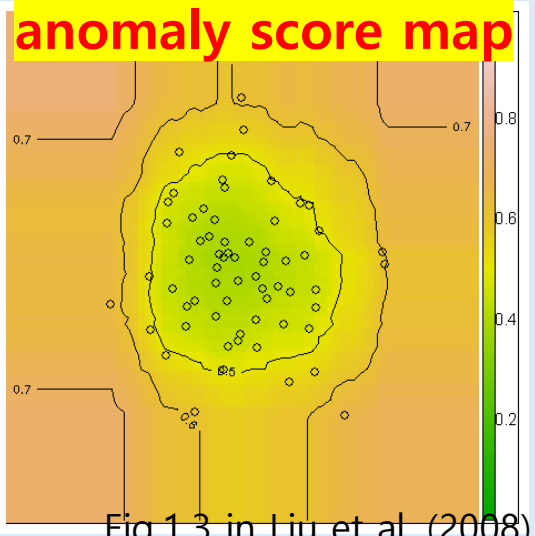


Fig.1,3 in Liu et al. (2008)

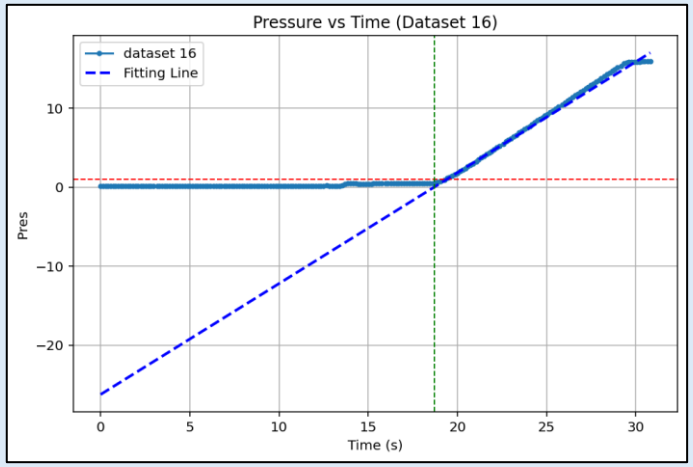
Preprocessing method

Limit & sanity test

Temperature	-2~40°C
Salinity	0~42
Turbidity	0~1000
Depth	0<
Time	$T_{i+1} > T_i$

Mask if condition is not met

Surface clear fitting method



$$z(t) = \alpha t + \beta, \qquad t^* = \frac{\beta}{\alpha}$$

**Masking range : t<t*

Parse feature

1. feature diff

$$\Delta f_i = f_{i+1} - f_i$$

2. feature grad

$$\nabla f_i = \frac{f_{i+1} - f_i}{P_{i+1} - P_i}$$

3. feature curv

$$\kappa_i = \frac{\nabla f_{i+1} - \nabla f_i}{P_{i+1} - P_i}$$

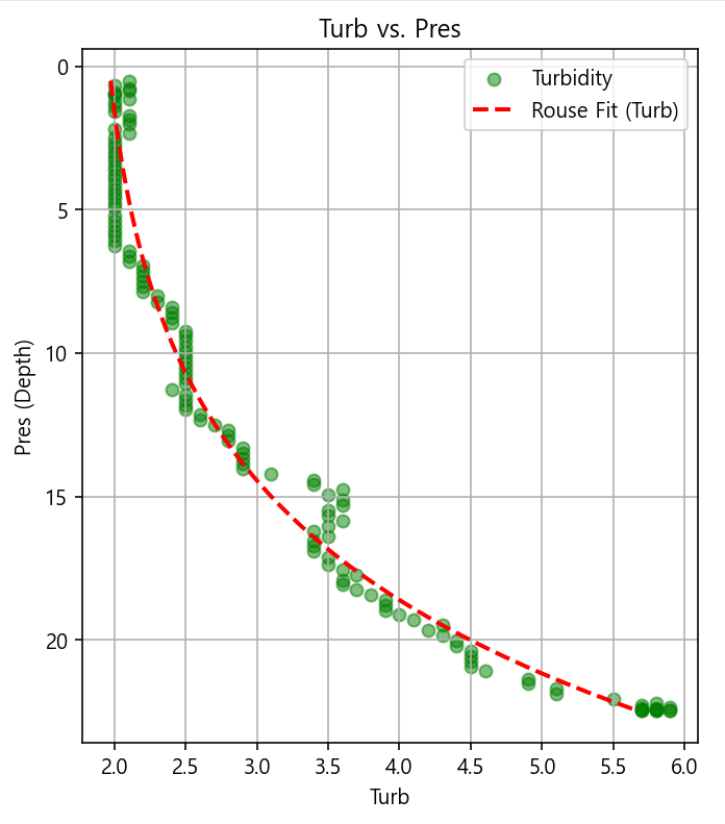
Step 1 : Values outside physical ranges or violation monotonic time/pressure are **masked immediately when conditions are not met.**

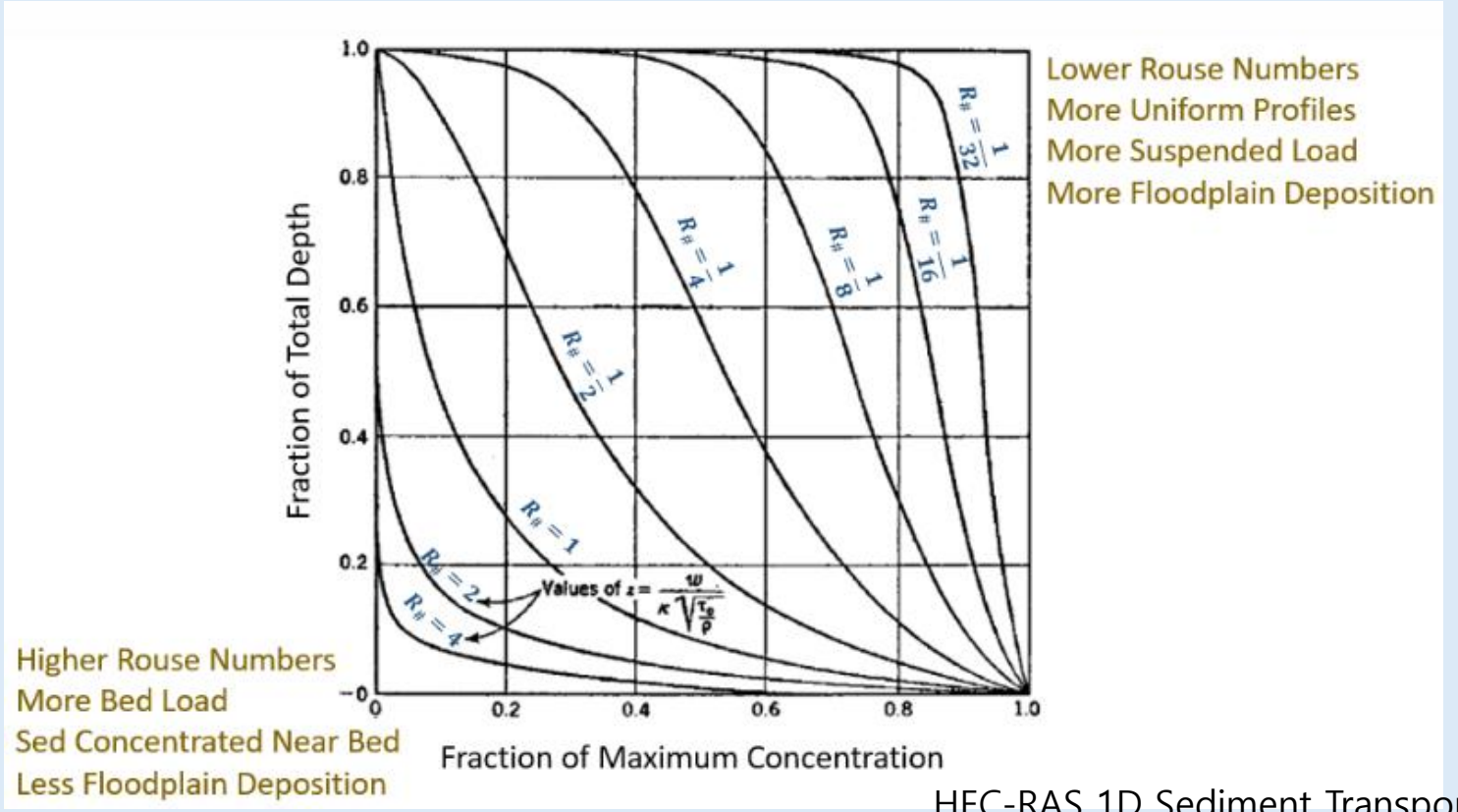
Step 2 : Fit line to estimate the true descent start(t^*) and **remove the surface stabilization segment.**

Step 3 : Compute difference(Δf), gradient(∇f), and curvature(κ) for each variable to **encode physical shape changes** and feed these features to the Isolation Forest.

Preprocessing method (Rouse fit)

Step 4 : To **encode the bottom-increasing turbidity pattern**, ideal turbidity function is fit and residuals are used as feature.





HEC-RAS 1D Sediment Transport

Rouse function

$$C(z) = C_a \left[\frac{z_a/(H - z_a)}{z/(H - z)} \right]^P, \qquad P = \frac{\omega_s}{\kappa u_*}$$

Parse feature method

4. feature turb residual

$$r_i = Turb_{z(i)} - C(z)$$

Validation method

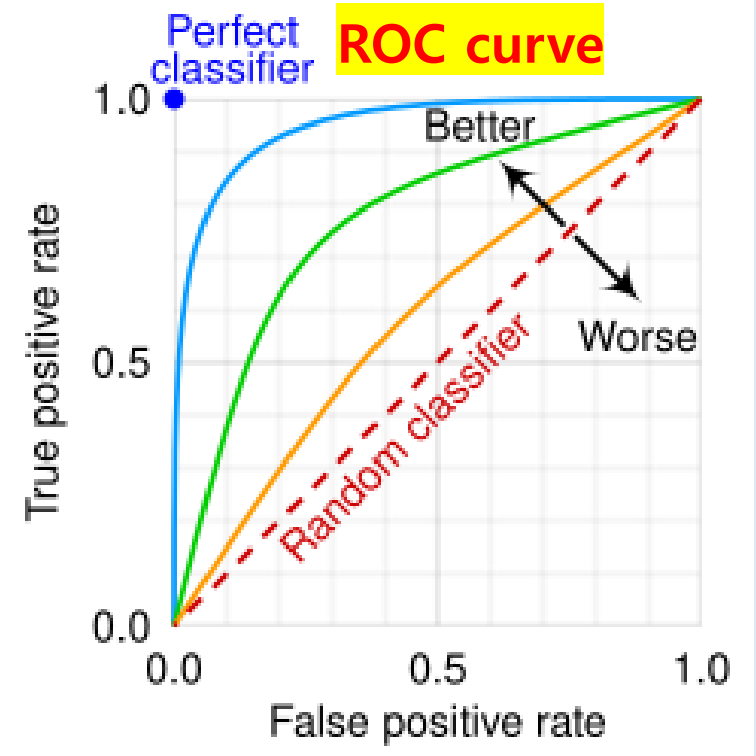
Confusion matrix

		Ground Truth (expert QC)	
		True	False
Model Result (IF QC)	True	True Positive (TP)	False Positive (FP)
	False	False Negative (FN)	True Negative (TN)

$$(Precision) = \frac{TP}{TP + FP} \quad (Accuracy) = \frac{TP + TN}{TP + FN + FP + TN}$$

$$(Recall) = \frac{TP}{TP + FN} \quad (F1\ score) = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

ROC curve



ROC curve

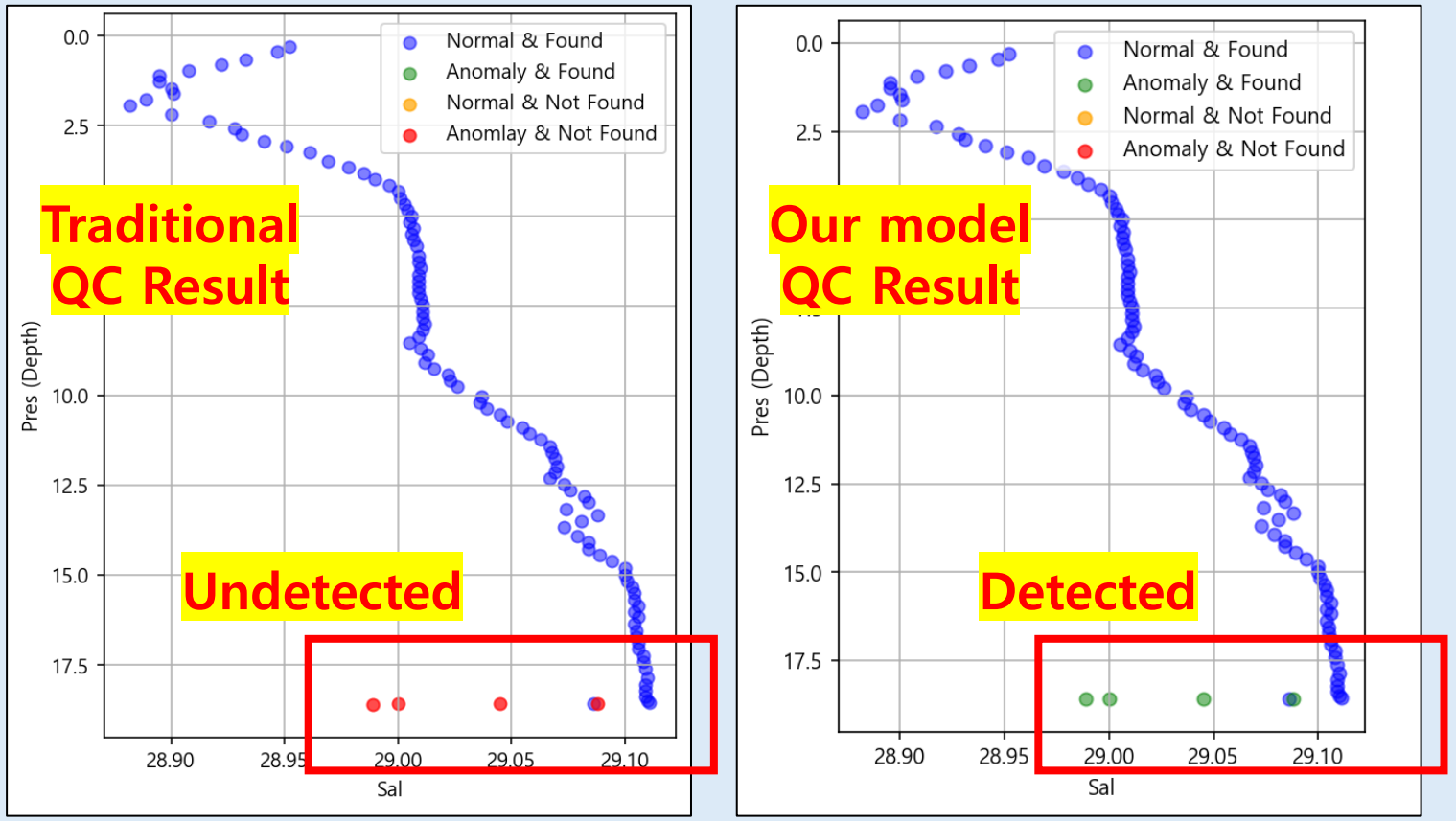
→ Use expert QC as ground truth and build a confusion matrix against model outputs.

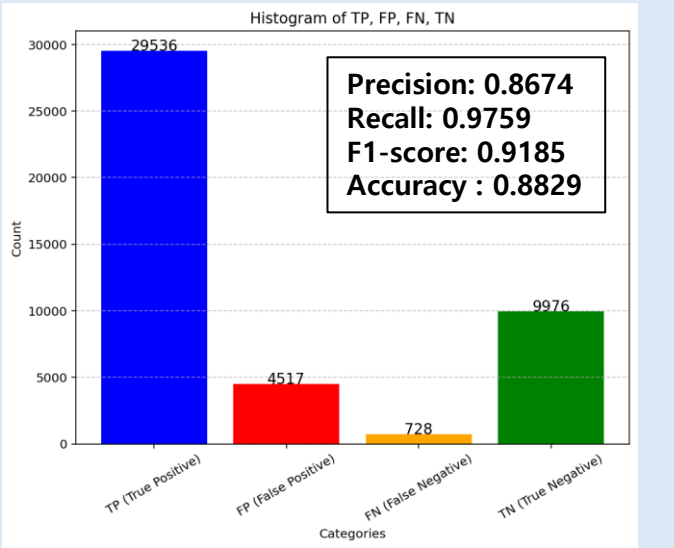
→ Report Precision(controls false alarms), Recall(controls misses), F1(balance), and Accuracy(overall).

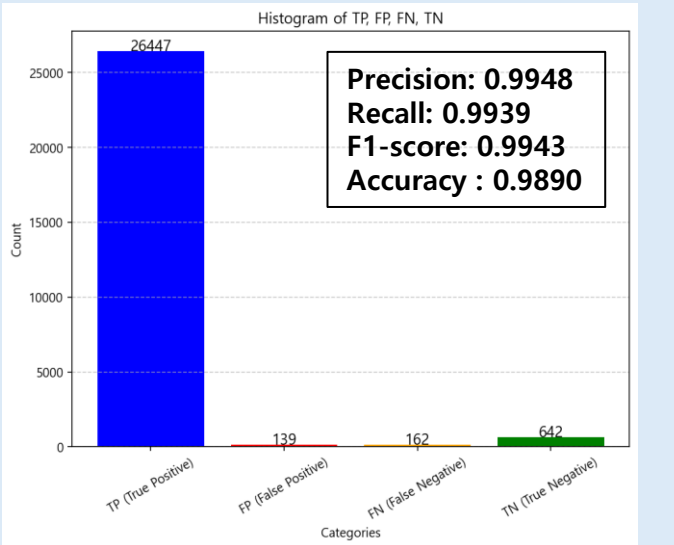
→ Sweep the threshold to plot TPR-FPR and summarize separability with AUC(Area Under Curve).

Result

Result

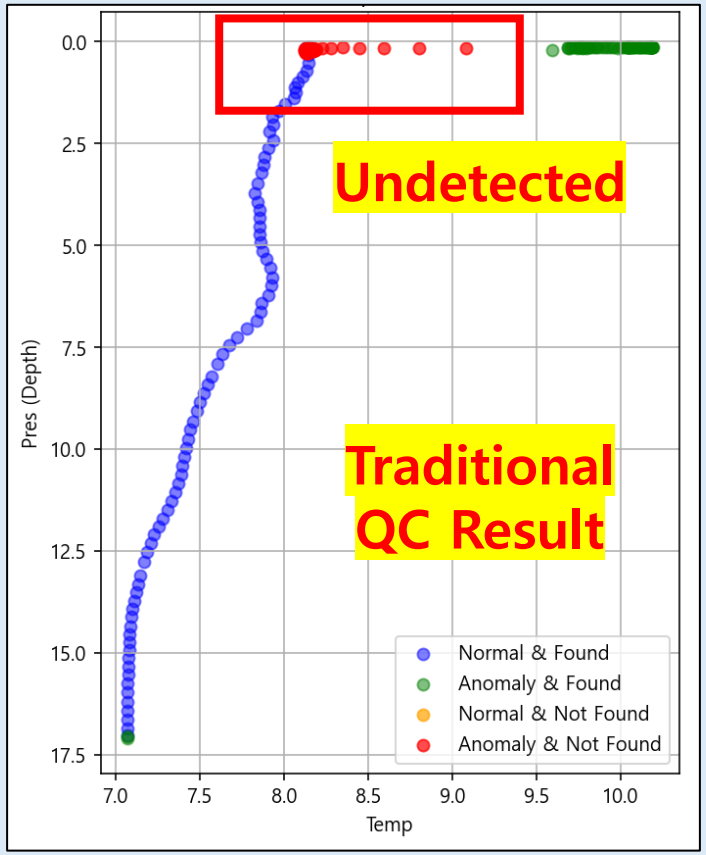


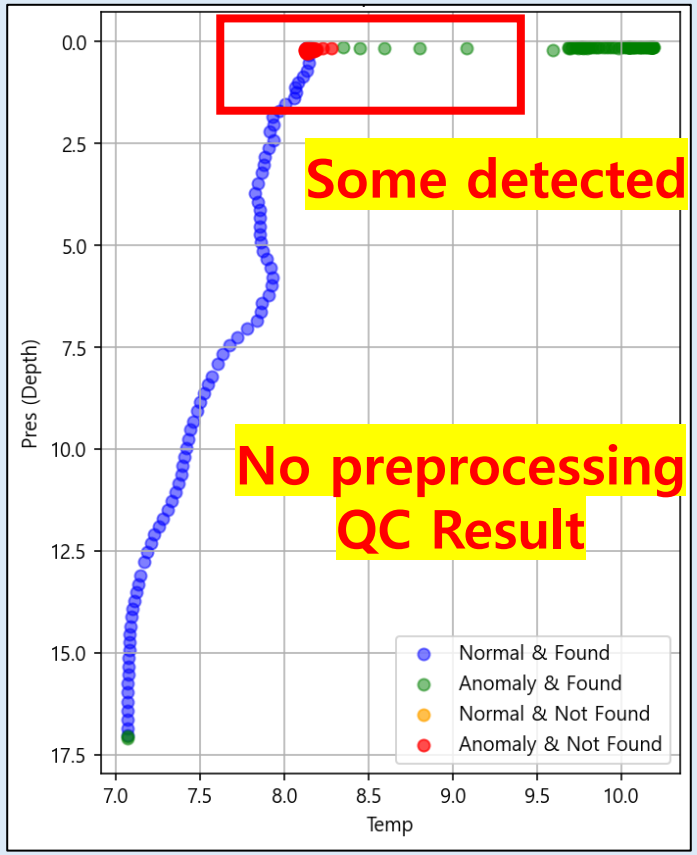


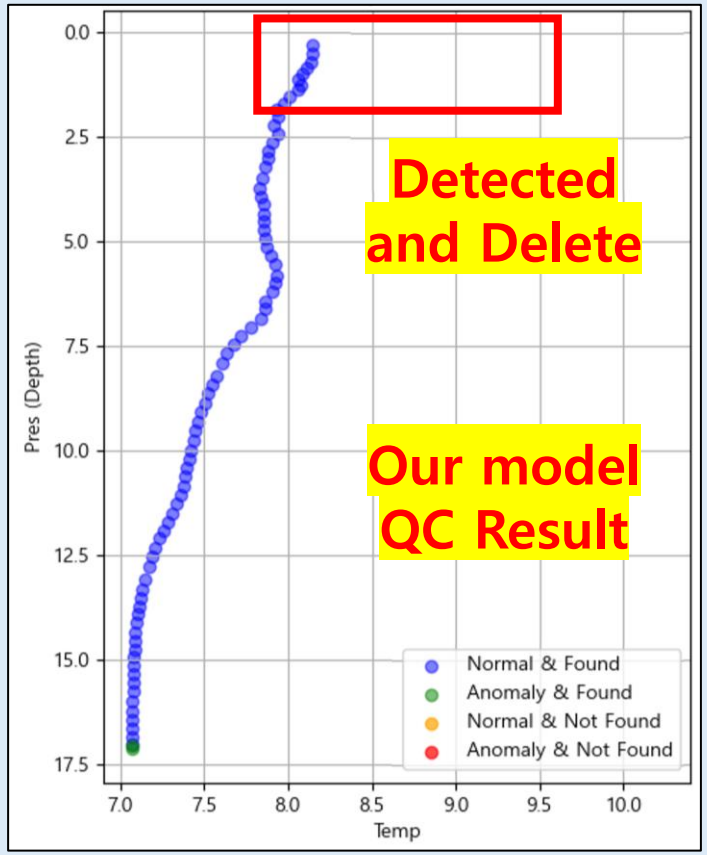


→ **The results of the QC method using AI detect false errors more effectively than the traditional QC method, and are more sensitive to low-level specific errors occurring in coastal areas.**

Result



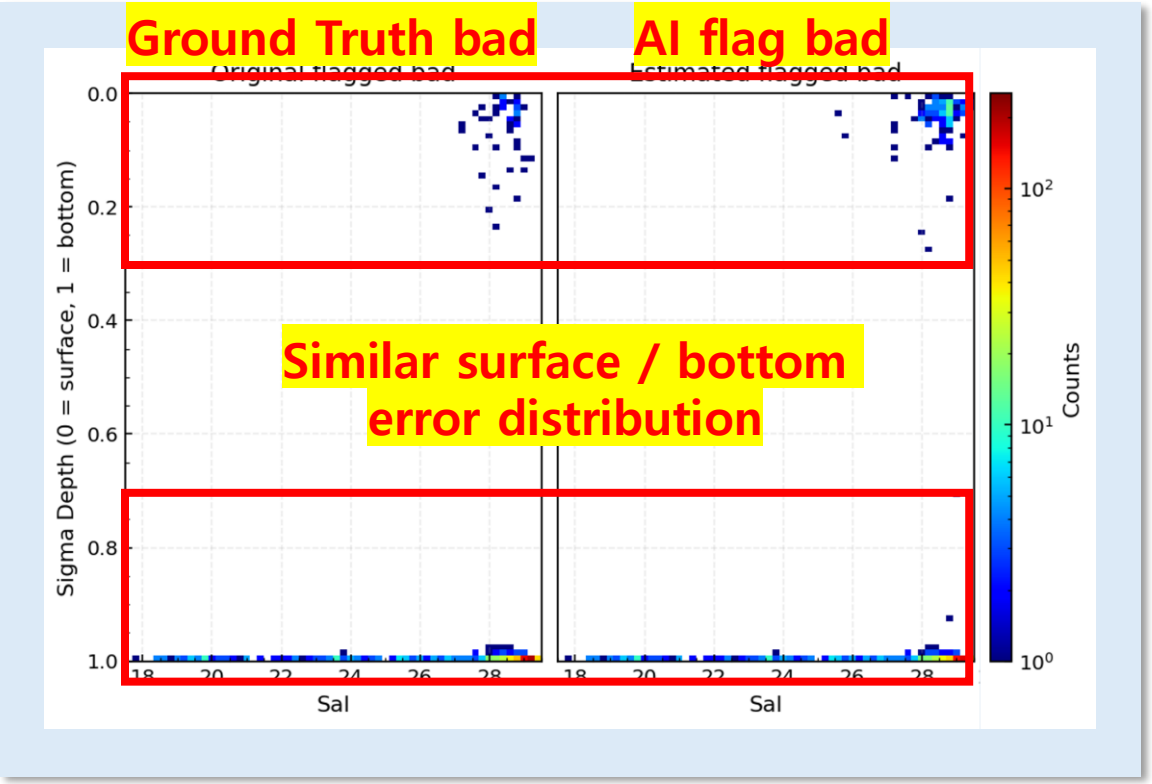


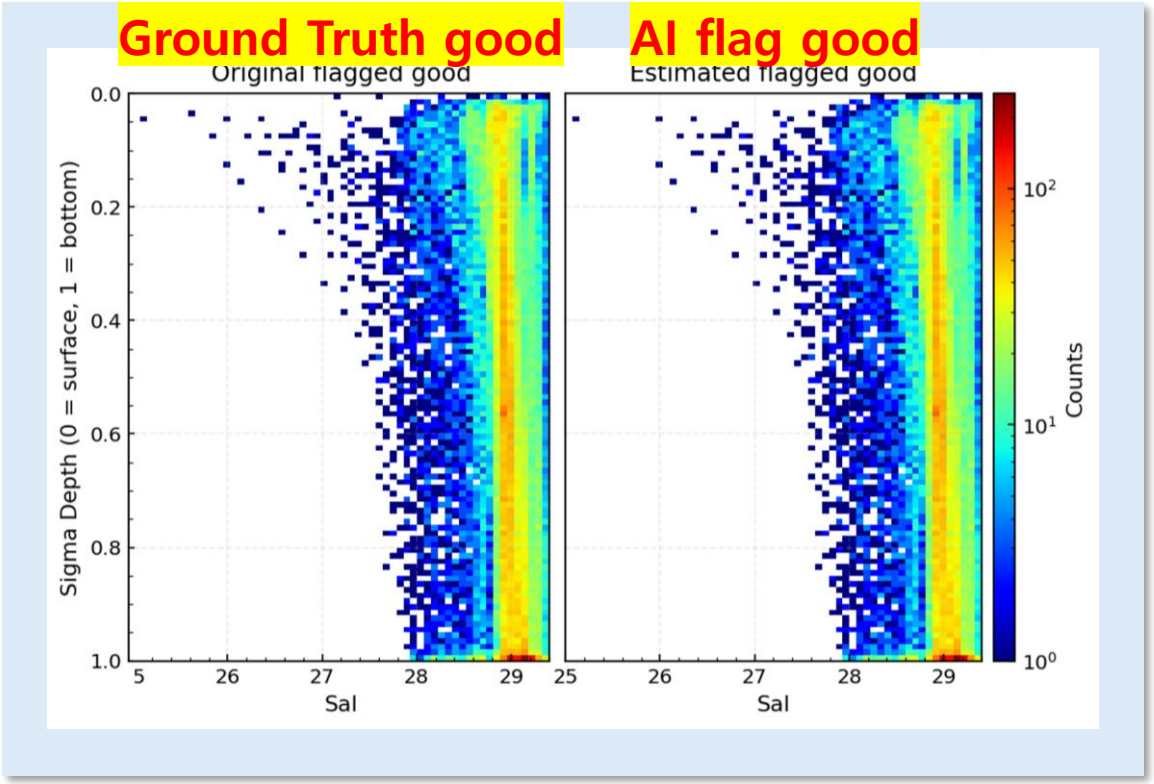


→ Traditional QC method rarely detect errors in the initial section of the surface, and QC method without preprocessing only dtect a portion of them.

→ Our model accurately detects and removes surface errors, restoring the correct profile shape

Comparative analysis result

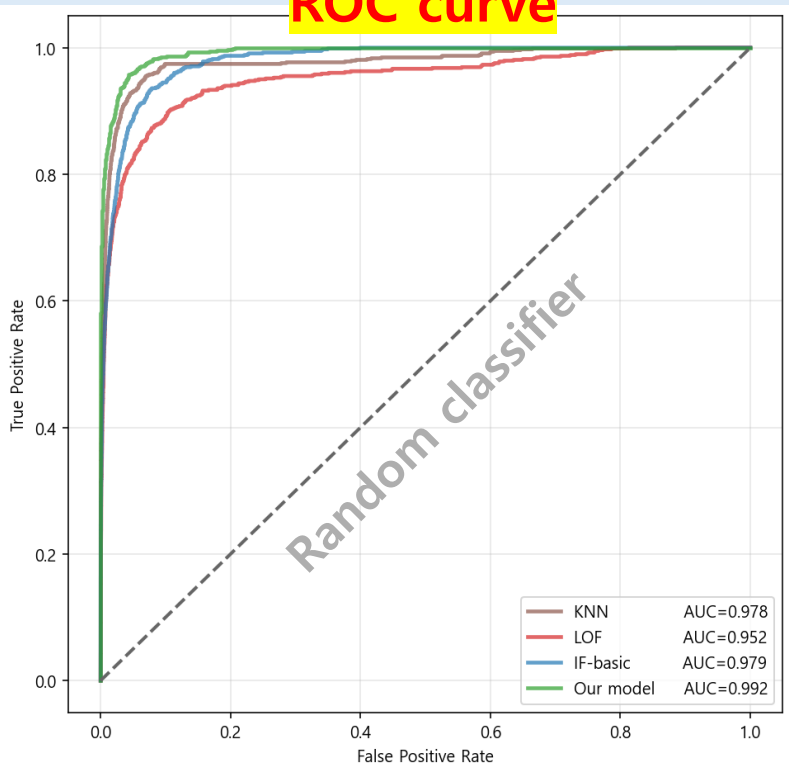


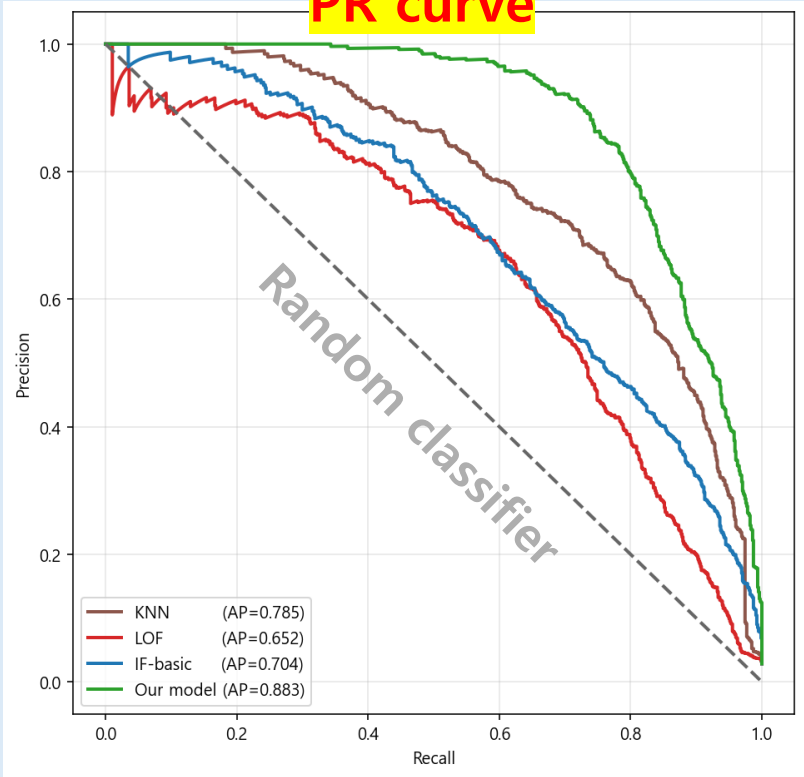


→ The AI-flagged bad distribution closely matches the ground truth(expert QC), with errors concentrated at the surface and near the bottom and a nearly identical spatial pattern.

→ The normal data distribution preserves the same boundary location as the ground truth(expert QC), maintaining a stable cluster of normal profiles.

ROC curve & PR curve





→ On the ROC plot, our model rises steeply from the very lsow-FPR region and stays farthest from the random baseline, indicating strong separability early on.

→ **In short, both views show that across practical thresholds the model keeps false alarms low while sustaining high detection, making it robust for operation**

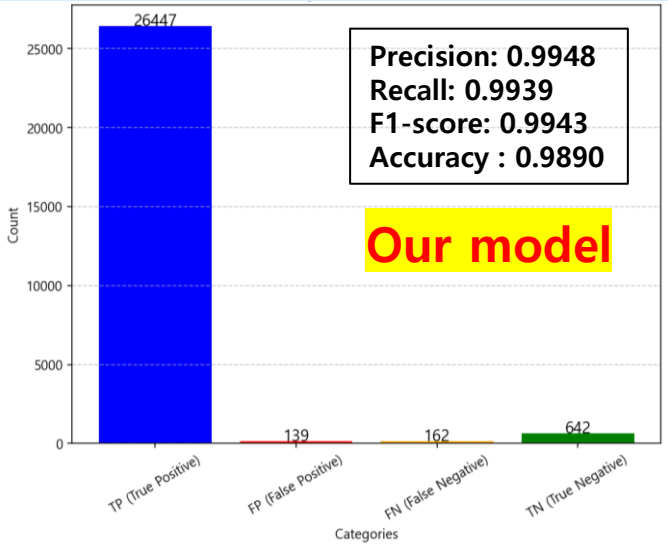
	Our model	KNN	LOF	IF-basic
ROC curve (AUC)	0.9923	0.9777	0.9518	0.9789
PR curve (AP)	0.8834	0.7850	0.6519	0.7038

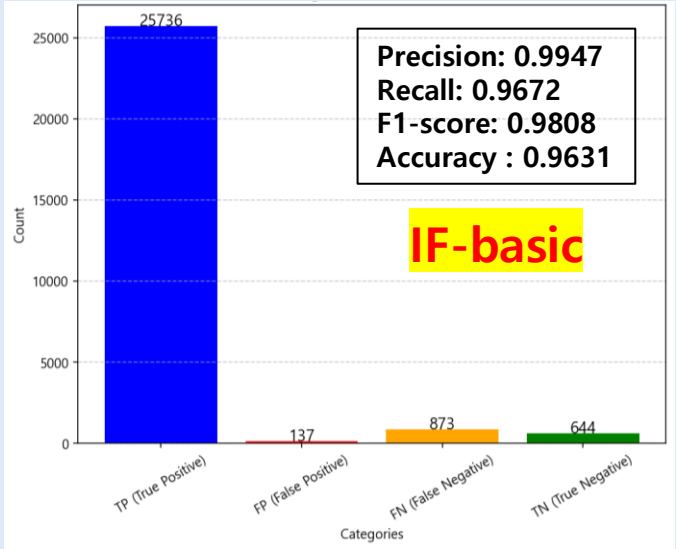
Final result

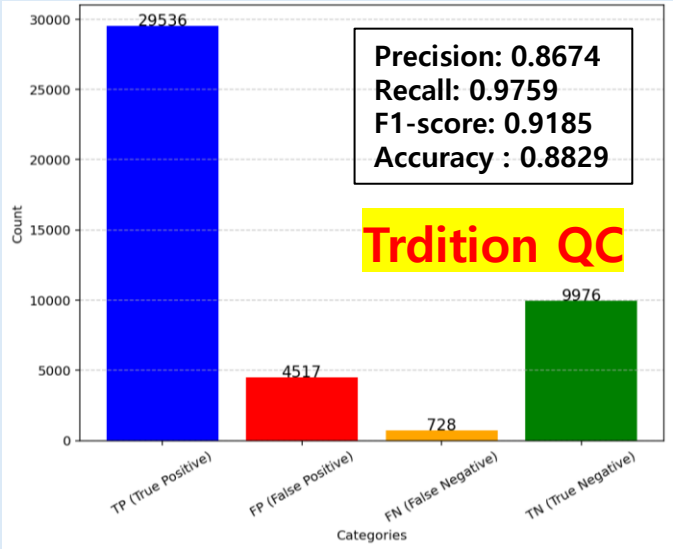
	Precision	Recall	F1 score	Accuracy	AUC	AP
Our model	0.9948	0.9939	0.9943	0.9890	0.992	0.880
IF-basic	0.9947	0.9672	0.9808	0.9631	0.979	0.704
Tradition QC	0.8674	0.9759	0.9185	0.8829	-	-

→ Our model maintains a high balance of precision and recall.

→ Versus IF-basic, FN drop substantially(smaller orange bar), improving recall; compared with traditional QC, FP are much lower, leading to more stable decisions.







Conclusion

Conclusion

Our model reliably separates surface and near-bottom outliers,
by cutting the unstable surface segment, masking bottom-contact artifacts,
and using an ideal turbidity-profile residual as a feature within IF score.

The method matches reference QC and outperforms alternatives,
achieving AUC = 0.992 and AP = 0.883,
with clear alignment of good/bad boundaries-especially in the surface and bottom zones.

The pipeline is lightweight and portable for coastal CTD QC, compatible with standard flag schemes
and requiring only minor tuning of limits, window size, and threshold for new sites.

Thank you