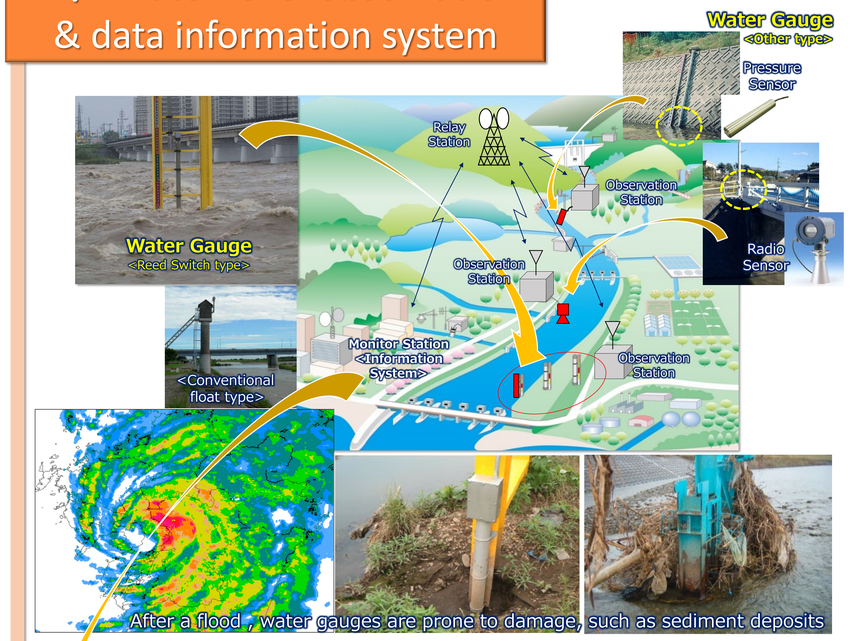
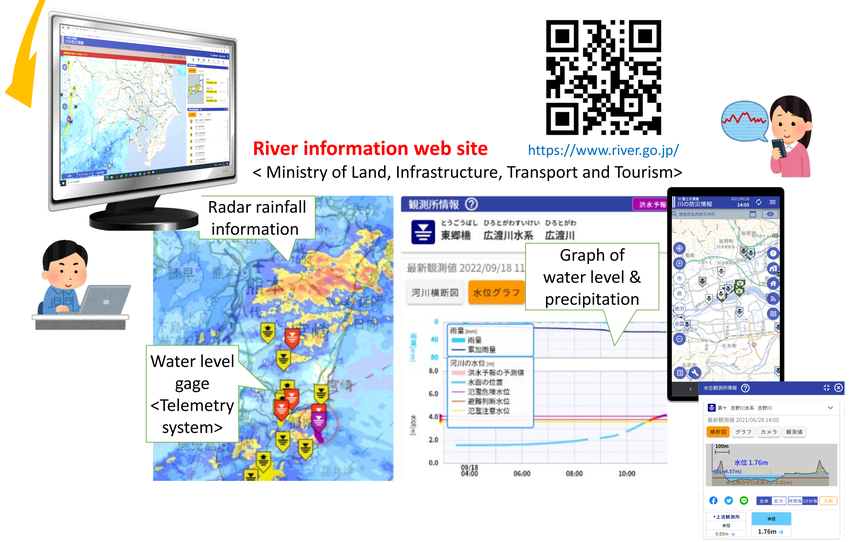
Development and actual use of
Real-time anomaly detection system
for river water level data

Tadashi TADOKORO / Osamu OKADA / Hisaya SAWANO / Arata SEI
Foundation of River and Basin Integrated Communications (FRICS) , JAPAN

Ⅰ. Water level observation & data information system

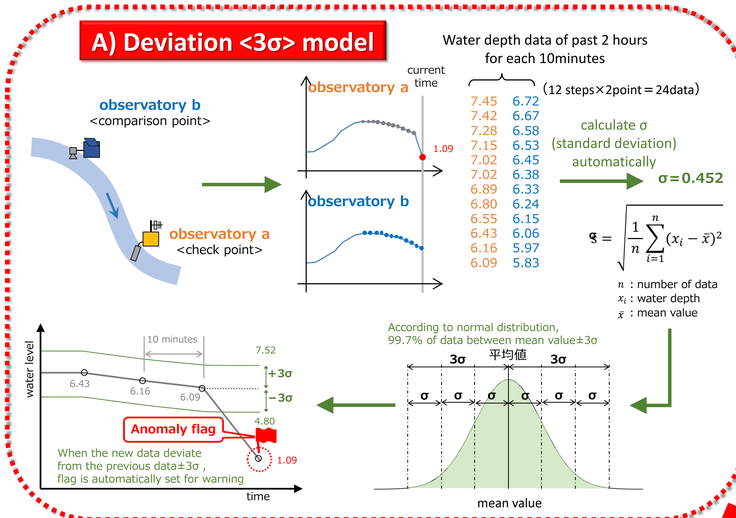




- In Japan, river water levels are observed at over 14,000 points
- **Real-time data** from these stations are integrated into the information system and delivered on **web sites**
- However, water gauges often produce **irregular data**, so, we need a precise data verification system at the monitoring station

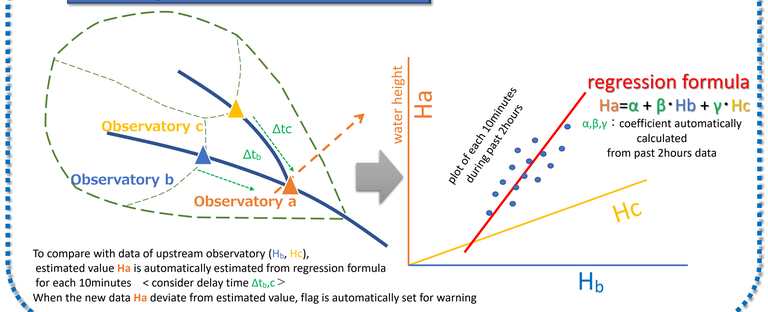
Ⅱ. Concept of anomaly detection models

A) Deviation <3σ> model



B) Regression <VAR> model

VAR : Vector Auto Regression



- At first, we devised two types of anomaly detection model to compere with neighboring station data.
  - **A) Deviation <3σ> model**
  - **B) Regression <VAR> model**
- When compared with **C) Existing difference model** (comparison with previous data from the same station) , both models showed a significantly better precision ratio.
- **A) Deviation <3σ> model** recorded highest Precision ratio. Therefore, we adopted this model as a real-time check system for the monitoring offices.

Verification

Result of checking some representative observation data during one rainy season

Precision rate : P=TP/(TP+FP)

A) Deviation <3σ> model	P = 0.60
B) Regression <VAR> model	P = 0.40
C) Existing differential model	P = 0.21

Classification		Predicted	
		outlier	normal
Actual	outlier	TP	FN
	normal	FP	TN

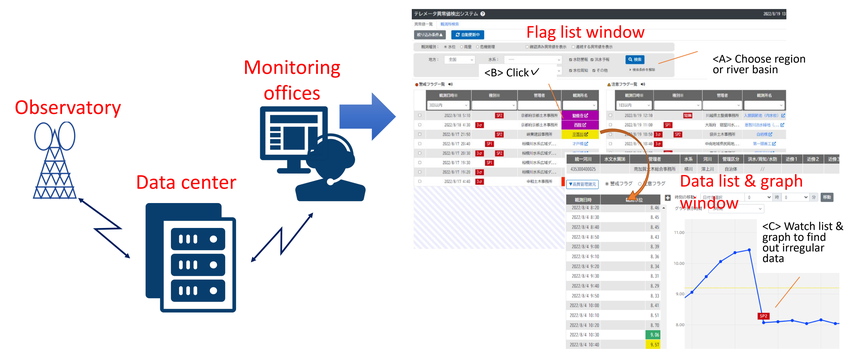
A) Deviation <3σ> model		Predicted	
		outlier	normal
Actual	outlier	6	0
	normal	4	20

B) Regression <VAR> model		Predicted	
		outlier	normal
Actual	outlier	6	0
	normal	9	15

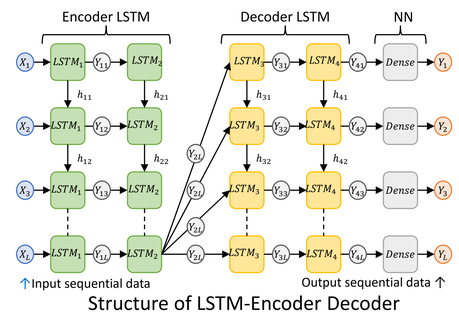
C) Existing difference model		Predicted	
		outlier	normal
Actual	outlier	6	0
	normal	22	2

Ⅲ. Construction of real-time data check system

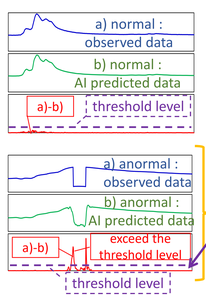
- Two years ago, we constructed a real-time anomaly detection system hosted on a cloud server in a data center.
- Today, data inspectors in **24-hour monitoring offices** utilize this system to identify and correct irregular data.
- This has significantly enhanced the precision and efficiency of our operations.



- We aim to further increase the accuracy by implementing an **AI model** using deep learning technology, specifically an **LSTM-Encoder Decoder**.
- Our prototype model has shown promising results, but we need to enhance its accuracy by training it with a larger dataset from across Japan.



Structure of LSTM-Encoder Decoder



- We have tested the AI model using over 50 flood data series.
- anormal data exceed threshold level★

★minimize F=2*P*R/(P+R)

Precision: P=TP/(TP+FP)

Recall : R=TP/(TP+FN)