



Accurate forecasts begin with accurate observations: How improved data production systems can lead the way



Stuart Hamilton, Dave Hutchinson, Touraj Farahmand and Derek Forsbloom



Smith Creek Saskatchewan, May 2011
Are the data accurate enough?



Timely data is essential but...

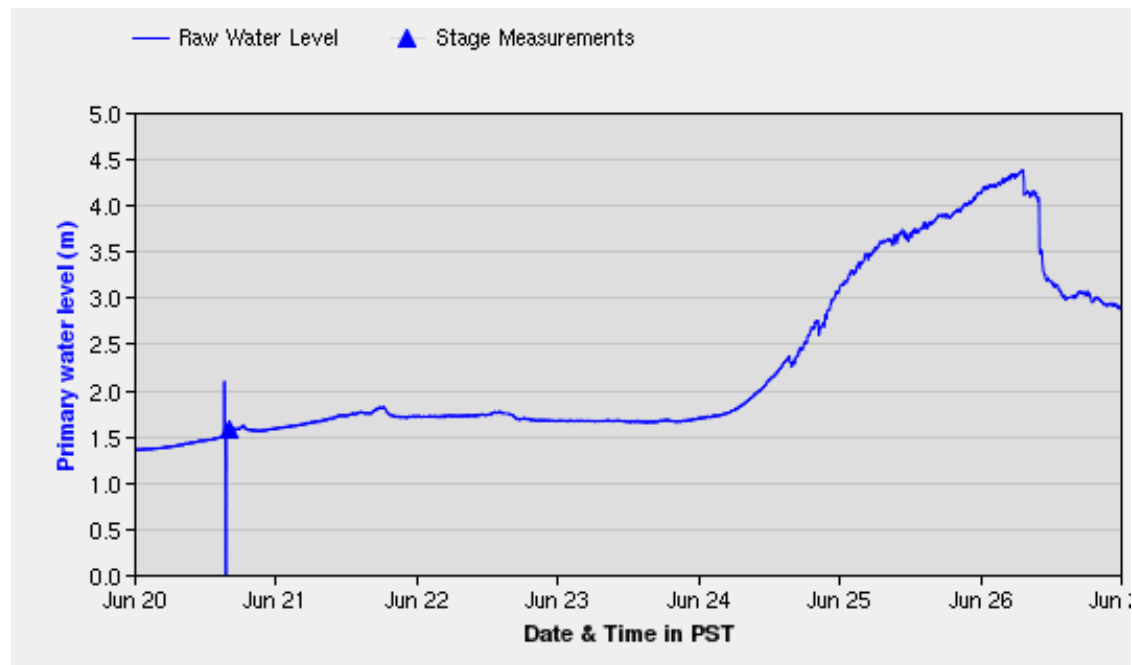
Potential for Disinformation

- ▶ Spurious data that looks real
- ▶ Valid data that looks spurious
- ▶ Valid stage, invalid discharge

Requires interpretation and analysis:

- ▶ Site-specific knowledge of weather conditions
- ▶ Controls on runoff processes (including, for example, operating rules for flow regulation)
- ▶ Hydraulic characteristics of control section
- ▶ Currently deployed technology
- ▶ Fault history

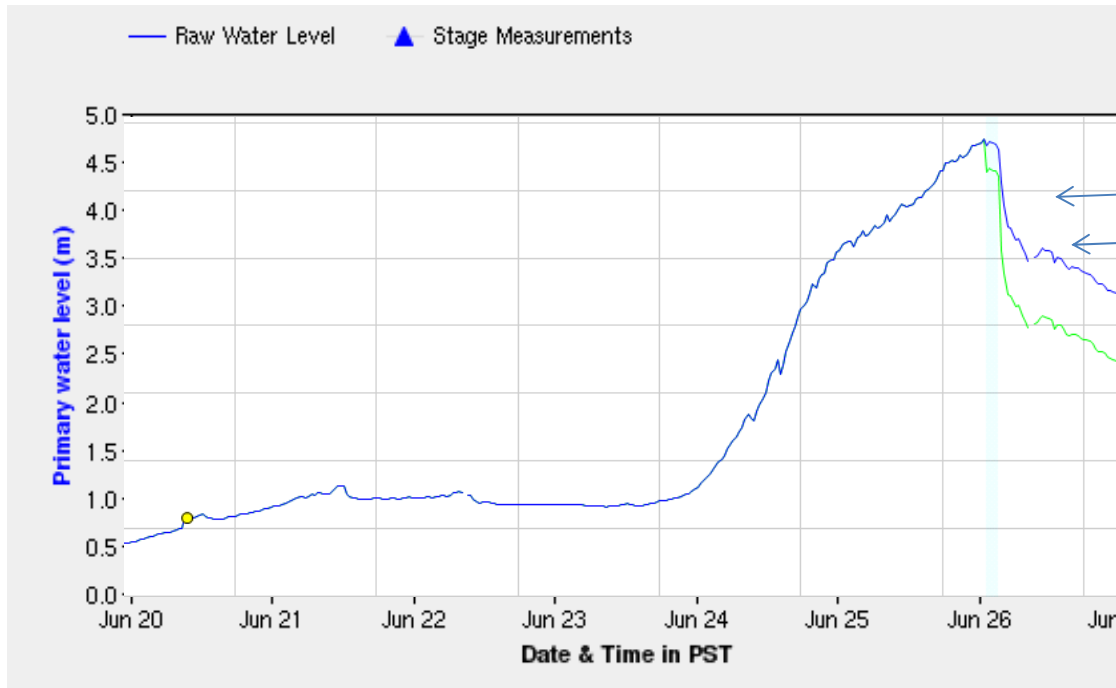
Potential for disinformation...



Halfway River

- Peak flow of 2011
- 3x as high as snowmelt peak
- Potential gauge for Site C reservoir inflow forecasting

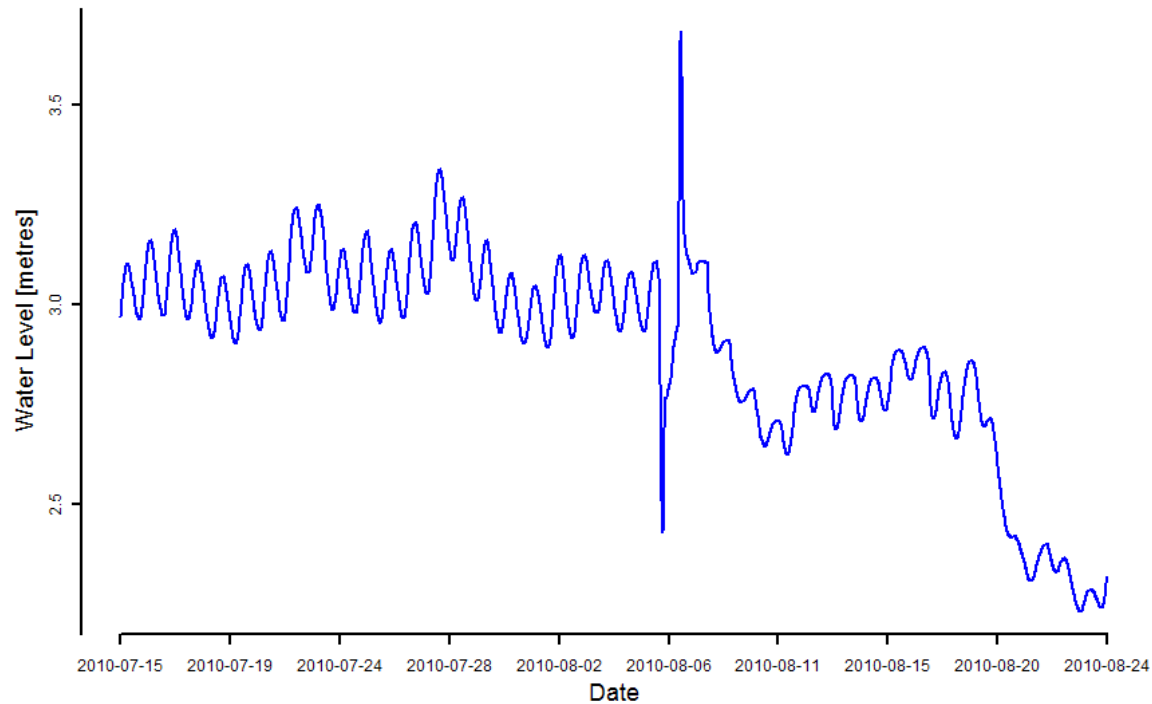
Potential for disinformation...



Orifice movements x2

- Net effect on stage about 0.7m
- Movement confounded by steep recession
- Readily identifiable by rate of change criteria

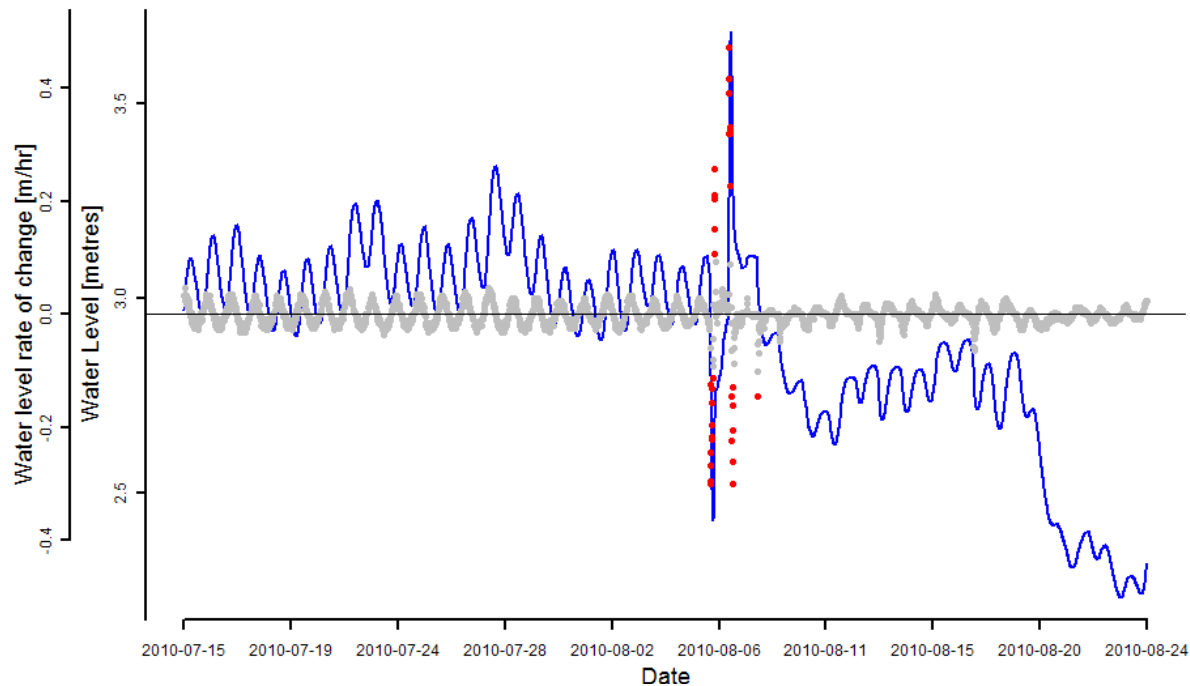
Valid data that looks spurious...



Lillooet River

- Diurnal fluctuation from glacial melt
- Could this be an event of interest?

Valid data that looks spurious...



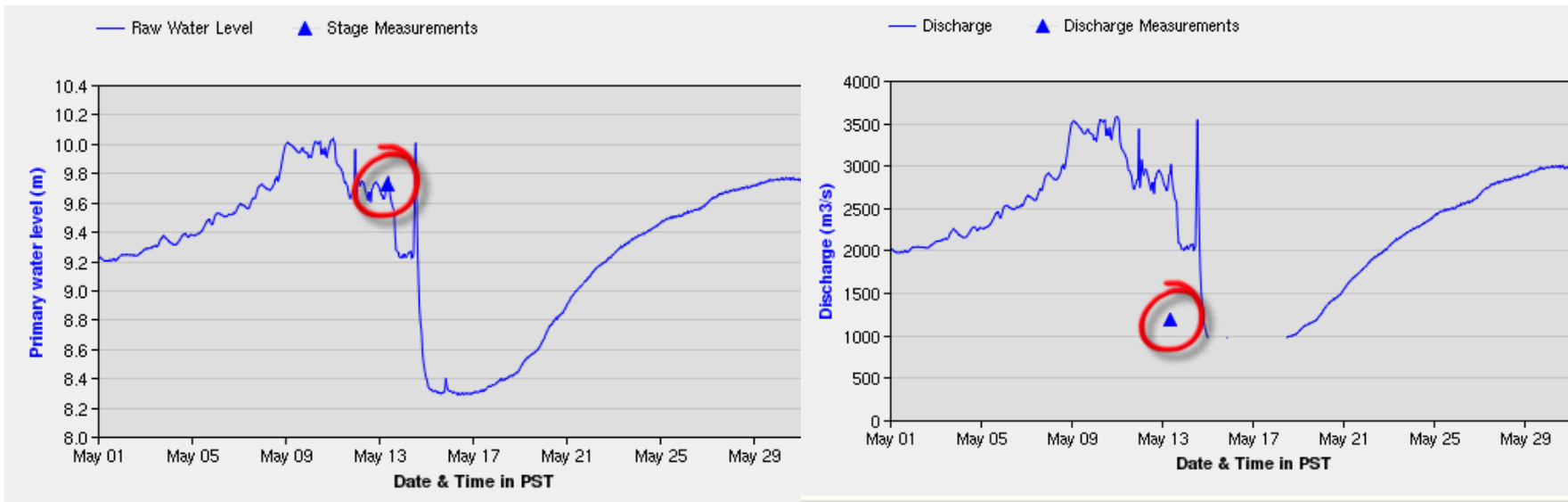
Lillooet River

- ▶ The event is readily identifiable by rate of change criteria
- ▶ timely investigation of anomalous data is always a good idea...

Meagre Creek Slide



Valid Stage – Invalid discharge



Yukon River

- ▶ Valid (ice affected) stage data
- ▶ Discharge from rating curve is almost 3x the measured discharge in mid-May

Real-Time Quality-Enhanced Discharge (RTQ) application

1. Developed by the WSC

- ▶ Mackenzie River Water Level Forecast Model
- ▶ On behalf of Canadian Coast Guard
- ▶ To assist in navigation of the Mackenzie River
- ▶ 3 day forecasted water level
- ▶ All available corrections in Water Survey's computational software supported
- ▶ Expanded to cover over 500 real-time gauges

2. Low Error Rates

- ▶ False Positives (where valid data are treated as pathological)
- ▶ False Negatives (where pathological data is treated as valid)
- ▶ Design tempered by years of trial-and-error experimentation with live data by skilled hydrographers

3. Efficiencies gained

- ▶ Set up of filters and corrections – automates processes that would otherwise be tedious manual tasks
- ▶ Alerts and warnings draw attention to anomalies that require interpretation or some other action
- ▶ Computer does what a computer does best – Hydrographer does what a hydrographer does best

4. Automation impacts:

- ▶ Products and services
- ▶ Response expectations
- ▶ Management of Field Operations
- ▶ Accountability
- ▶ Standard Operating Procedures (SOP)
- ▶ Training

Parameterized Data Quality Control

Value-by-Value QC

- ▶ Manual, human-driven, process
- ▶ Every new 'batch' of data requires inspection and custom correction.
- ▶ Data interpretation is subjective → Difficult to pass on accumulated knowledge about how a given signal should be corrected.
- ▶ Focus is on the information content of the current batch of data

Parameterized QC

- ▶ Automated, machine-driven, process
- ▶ Can be evaluated point-by-point, no loss of efficiency/effectiveness in processing small batches
- ▶ Data interpretation is preserved in parameter values. Comments on parameter history facilitate transfer of accumulated knowledge and experience
- ▶ Focus is on prediction. Effective parameter calibration is informed by interpretations of hydrology, hydraulics and statistical theory

Automated Hydrometric Data Processing Concepts

1. Filtering

- ▶ Preliminary Sanity check
- ▶ 'Sticky First'; 'Normal'; 'Sticky Last' Rule-based to take out technological faults e.g. spikes and flat lines
- ▶ Derived data filtering

3. Traceability, Reversibility

- ▶ Flagging, Grading, Commenting, Authorization
- ▶ Event logs, audit trail
- ▶ Source tracking
- ▶ Undo/modify/redo
- ▶ Continuous learning – parameter optimization

2. Corrections

- ▶ 'Sticky First'; 'Normal'; 'Sticky Last'
- ▶ Time-limited or 'open-ended'
- ▶ Trend following
- ▶ Time series and/or derivation model based

4. Notifications and reporting

- ▶ On append – e.g. did something happen that I need to know about?
- ▶ Scheduled – e.g. did something not happen that I need to know about?
- ▶ Production – e.g. what are my priorities today?
- ▶ Station Health diagnostics – e.g. How close am I to failure?
- ▶ Current Conditions – e.g. What is going on at my gauges?

Generic Bayesian inference framework

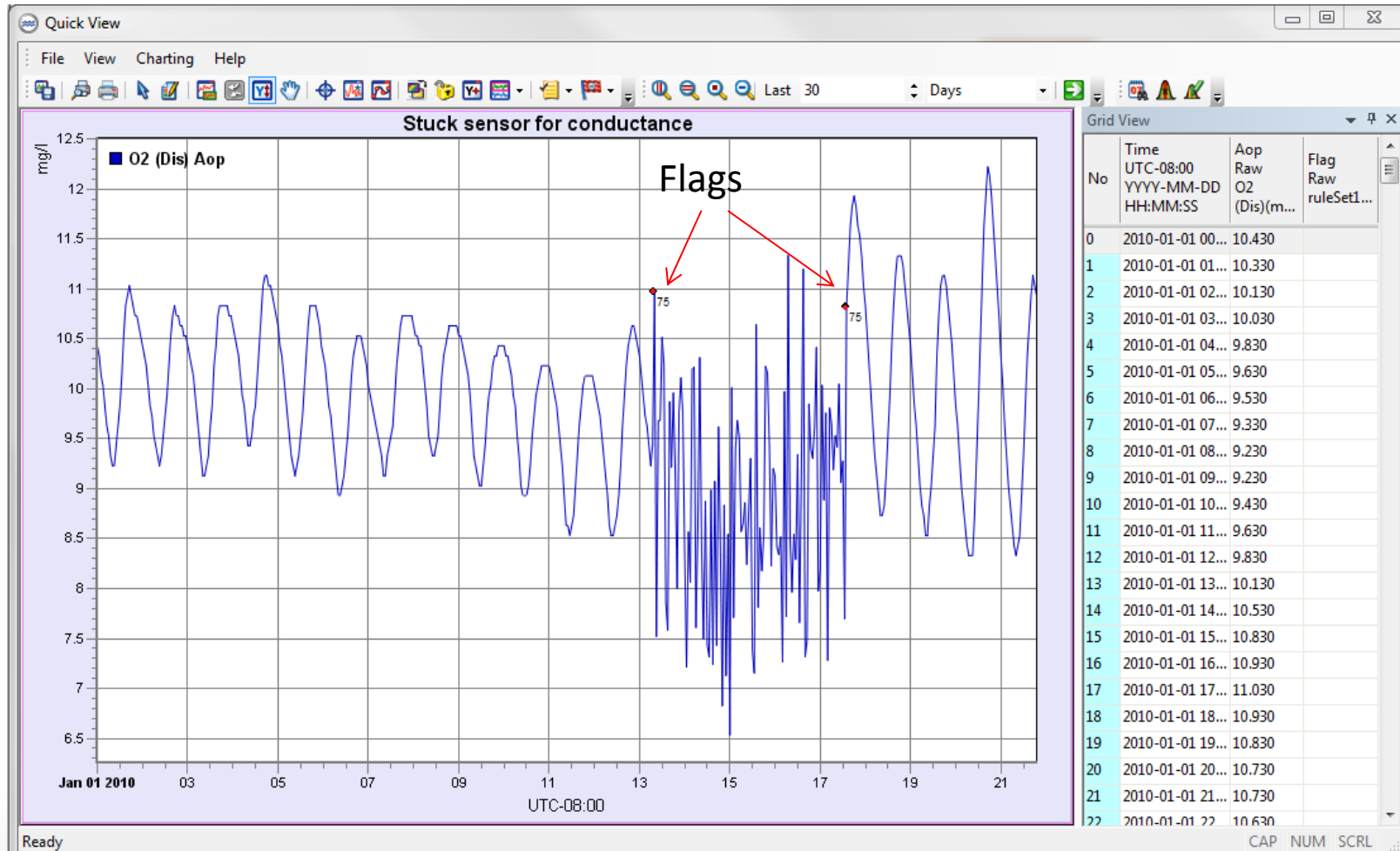
▶ A fast and efficient (real-time) linear and piecewise (switching) linear dynamical machine learning model has been developed:

- Sensor fault/anomaly detection. E.g. outlier, stuck sensor, offset,...
- Data correction and estimation. E.g. gap filling
- Short term prediction
- Smoothing
- Minimal user interaction since it learns all parameter from data

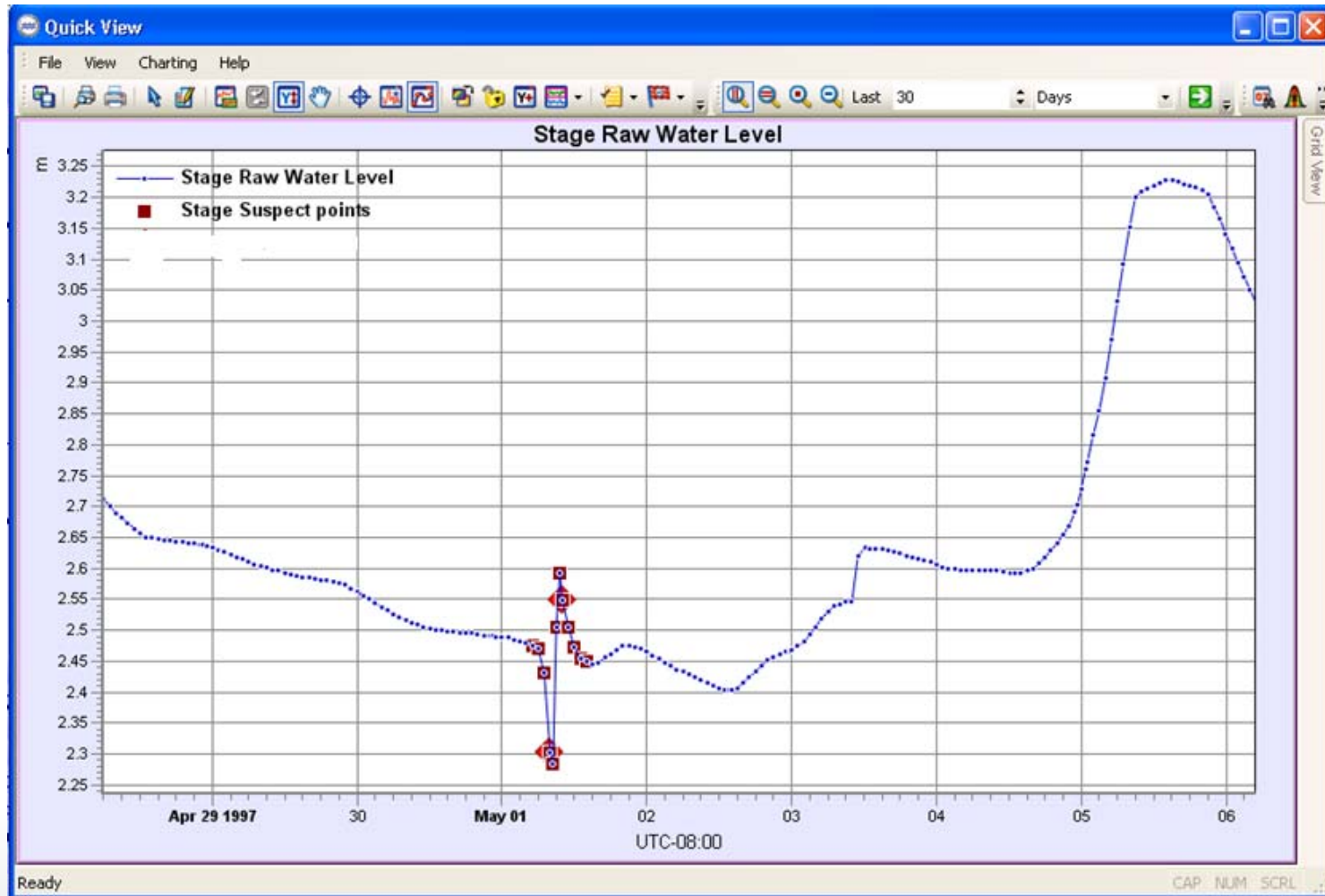
▶ Nonlinear dynamical machine learning models is under research:

- They are more accurate for modeling highly chaotic signals
- The big challenge is computational complexity and speed of training and inference

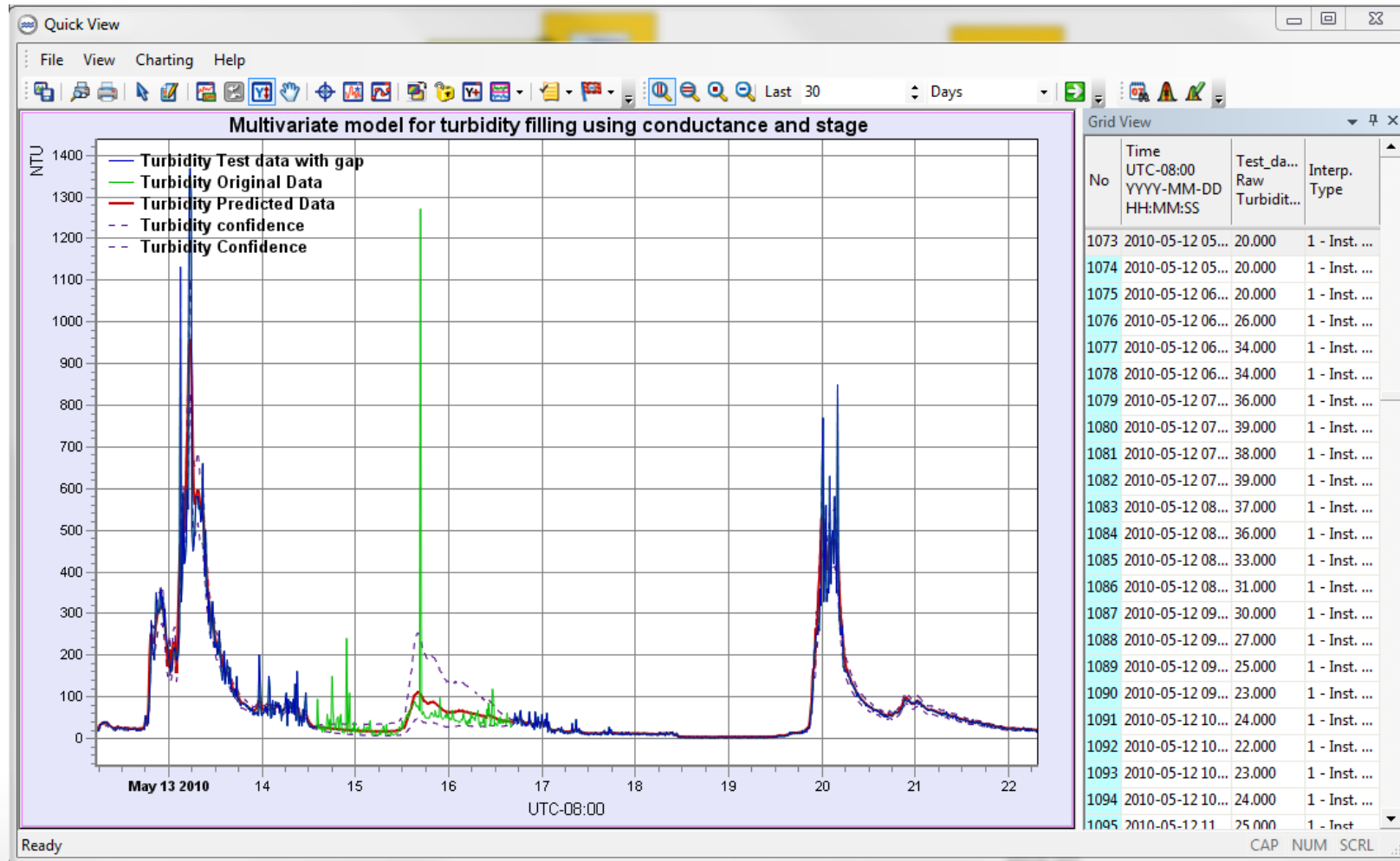
Univariate Model Results: Sensor fault detection



Univariate Model Results: Anomaly detection

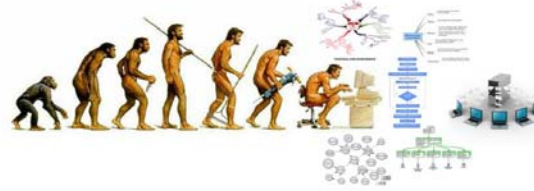


Multivariate Model Results: Gap Filling/Prediction



Multivariate Model Results: Gap Filling/Prediction

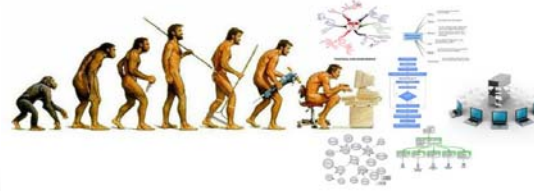




Hydrometric Evolution

Existing workflow and SOP:

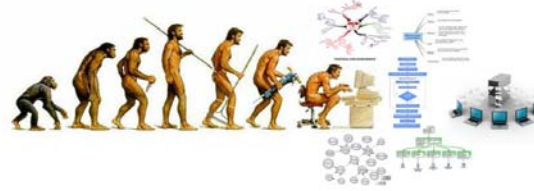
- ▶ Designed for an annual production cycle
- ▶ Designed for publication of annual statistics and daily aggregates
- ▶ Designed for a limited version history



Hydrometric Evolution

The role of the hydrographer:

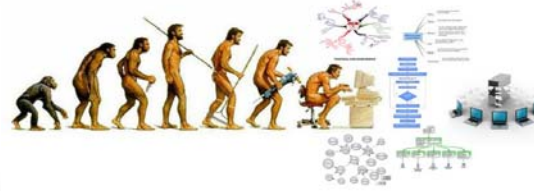
- ▶ Predictive rather than reconstructive; analytics rather than manipulations
- ▶ Focus on verifying assumptions (e.g. level surveys, rating curve applicability)
- ▶ Data-driven work planning and decision making – responsive and accountable actions



Hydrometric Evolution

Field operations focus:

- ▶ Preventive maintenance and early interventions
- ▶ Technological optimization - station health monitoring and fault history analysis
- ▶ Redundancy as a strategy for fault mitigation
- ▶ Model selection – site-and-season suitability



Hydrometric Evolution

Expanded suite of products and services:

- ▶ Quality controlled unit values
- ▶ Event-triggered alerts and reports
- ▶ Expanded metadata
- ▶ Version history

Conclusion

- ▶ Change in focus from archive to real-time data production process
- ▶ Predictive rather than reactive monitoring ethic
- ▶ Efficiencies gained by automation offset the time required for predictive analysis

Questions?