Quality Assurance & Quality Control of Lake Sensor Data

Tony Liu & Ian Wang

Original Data

```
"Date","Time","","Temp","","SpCond","","LDO%","","LDO","","IBatt",""
"M/D/YYYY","HH:MM:SS","","°C","","µS/cm","","Sat","","mg/l","","%Left",""

5/29/2014,09:00:00,"","20.94","","0","","100.3","","8.36","","84",""
5/29/2014,09:15:00,"","23.09","","0","","102.5","","8.20","","94",""
5/29/2014,09:30:00,"","25.01","","0","","102.4","","7.89","","93",""
5/29/2014,09:45:00,"","26.18","","0","","100.5","","7.59","","91",""
5/29/2014,10:00:00,"","26.35","","0","","100.6","","7.57","","86",""
5/29/2014,10:15:00,"","23.16","","0","","98.1","","7.83","","91",""
5/29/2014,10:30:00,"","22.94","","0","","97.1","","7.79","","82",""
5/29/2014,10:45:00,"","22.03","","0","","97.6","","7.96","","86",""
5/29/2014,11:00:00,"","21.58","","0","","97.3","","8.01","","91",""
5/29/2014,11:15:00,"","22.36","","0","","98.8","","8.01","","93",""
5/29/2014,11:30:00,"","22.19","","0","","99.1","","8.06","","93",""
5/29/2014,11:45:00,"","21.60","","0","","98.8","","8.13","","93",""
5/29/2014,12:00:00,"","21.48","","0","","98.4","","8.12","","93",""
5/29/2014,12:15:00,"","21.53","","0","","99.1","","8.17","","93",""
5/29/2014,12:30:00,"","21.69","","0","","99.3","","8.15","","91",""
```

“Obvious” Data Cleaning

- Temperature range [-50,100]
- Conductivity abnormal 0
- Dissolved oxygen percentage range (0,∞]
- Battery percentage range (0,100]

QA Standard

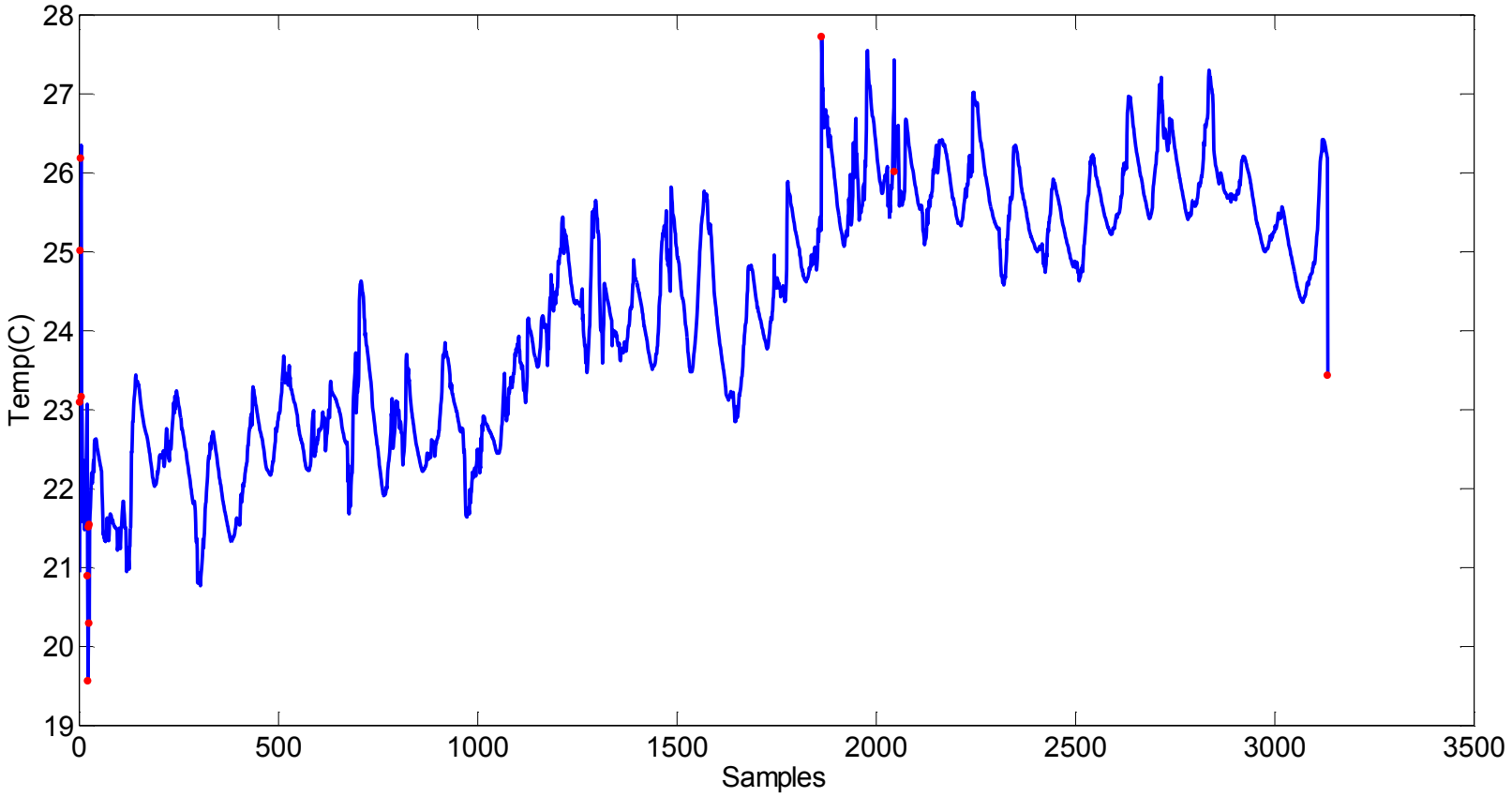
- Temperature variance ± 1
- Conductivity fluctuation ± 20%
  - If <10 then define fluc as ± 50%
- Dissolved oxygen percentage fluc ± 20%
  - If <10% then define fluc as ± 50%
- Dissolved oxygen value fluc ± 20%
  - If <10 then define fluc as ± 50%
- Battery percentage fluctuation ± 5

QA Standard

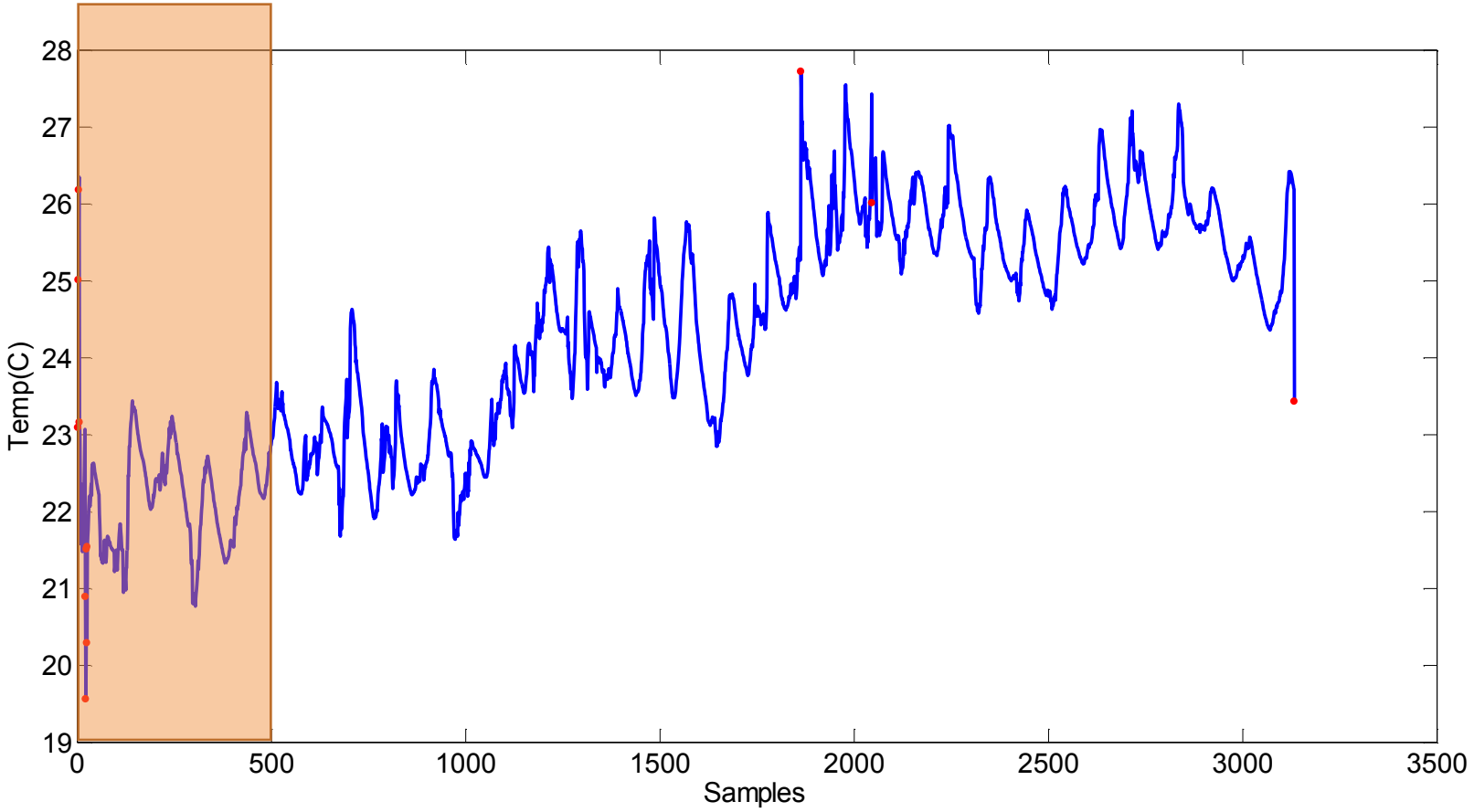
- Temperature variance ± 1
- Conductivity fluctuation ± 20%
  - If <10 then define fluc as ± 50%
- Dissolved oxygen percentage fluc ± 20%
  - If <10% then define fluc as ± 50%
- Dissolved oxygen value fluc ± 20%
  - If <10 then define fluc as ± 50%
- Battery percentage fluctuation ± 5

If stays the same for 3 samples, flag them!

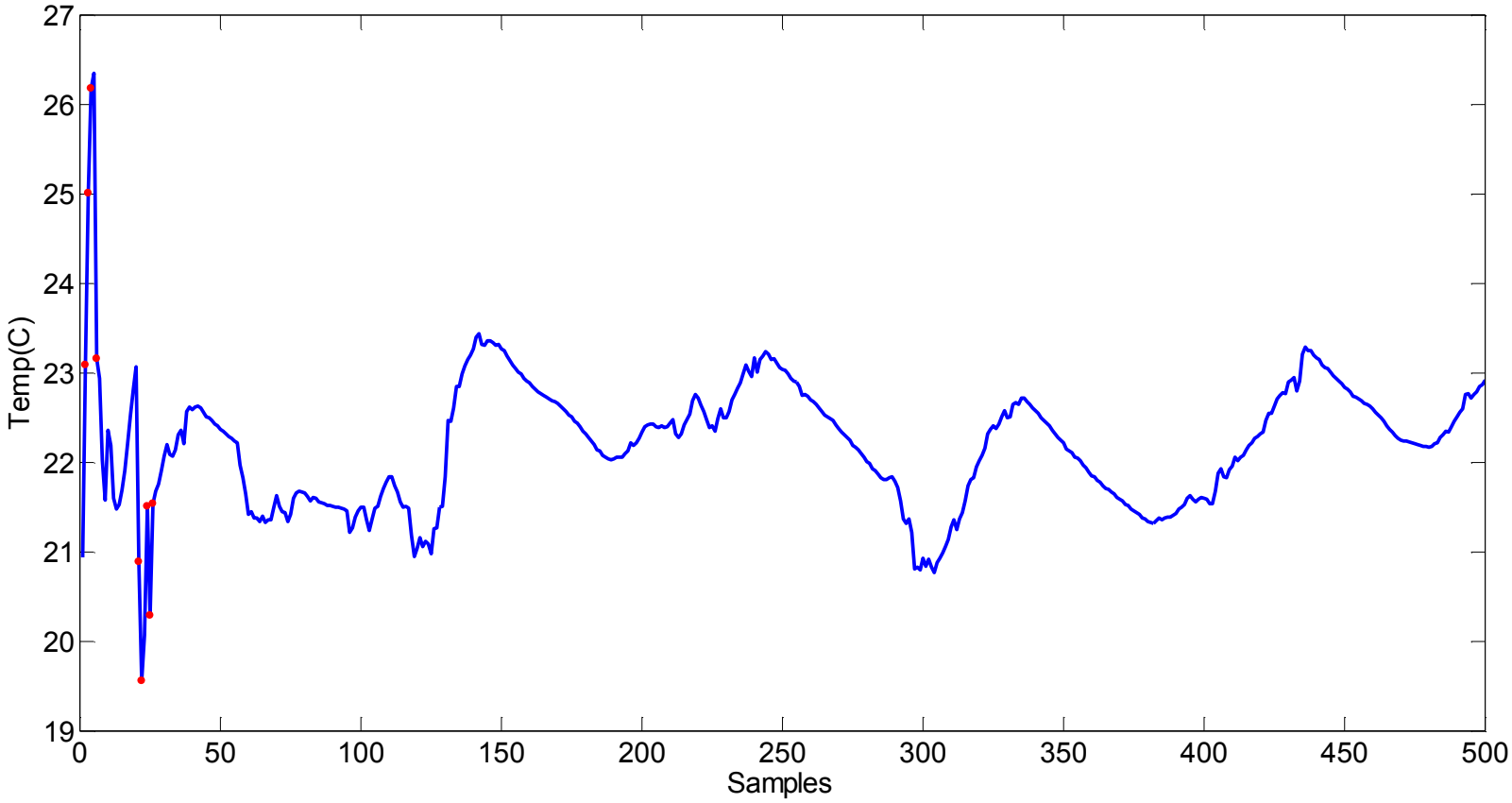
Example of Temperature Data



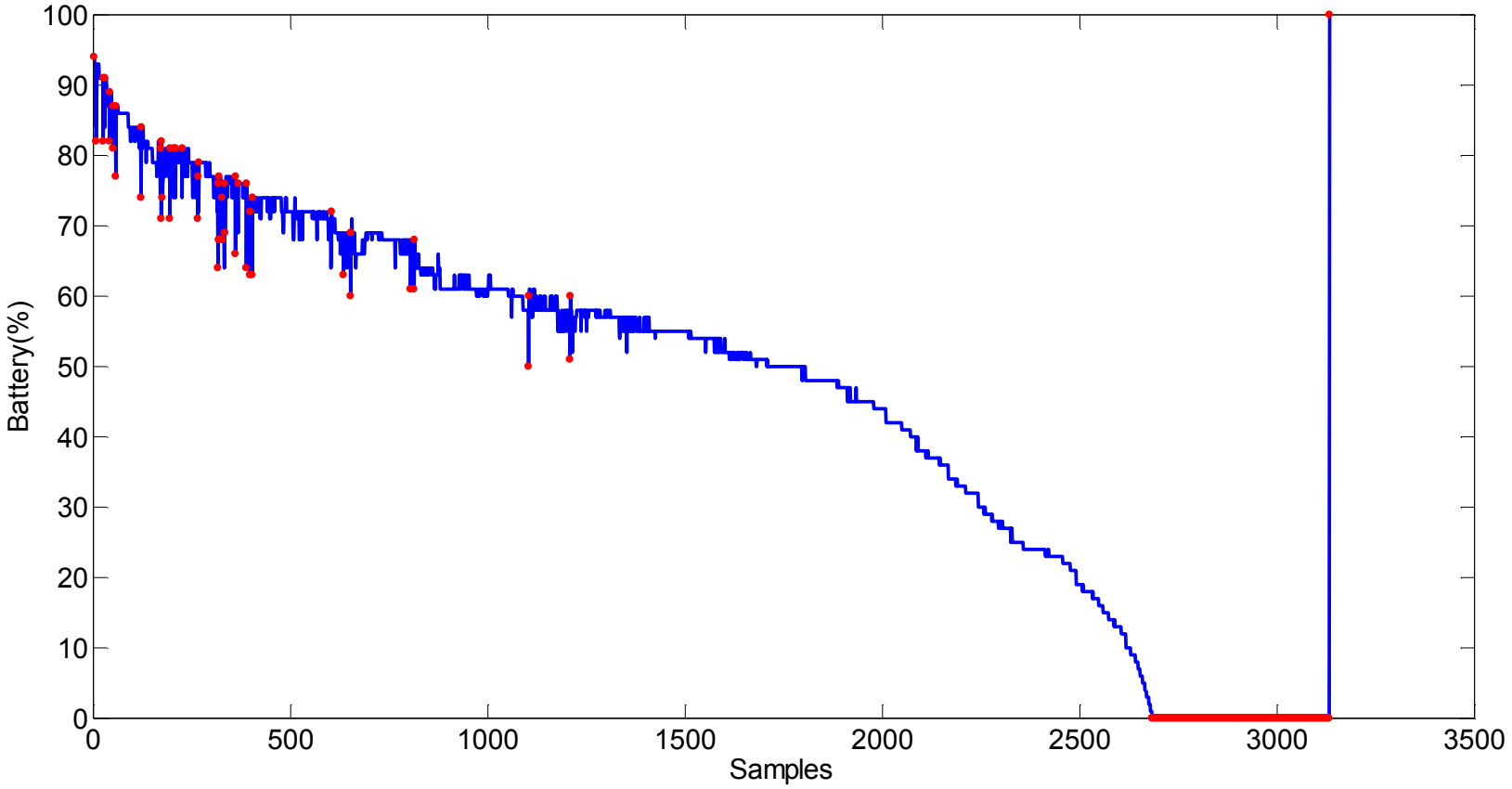
Example of Temperature Data



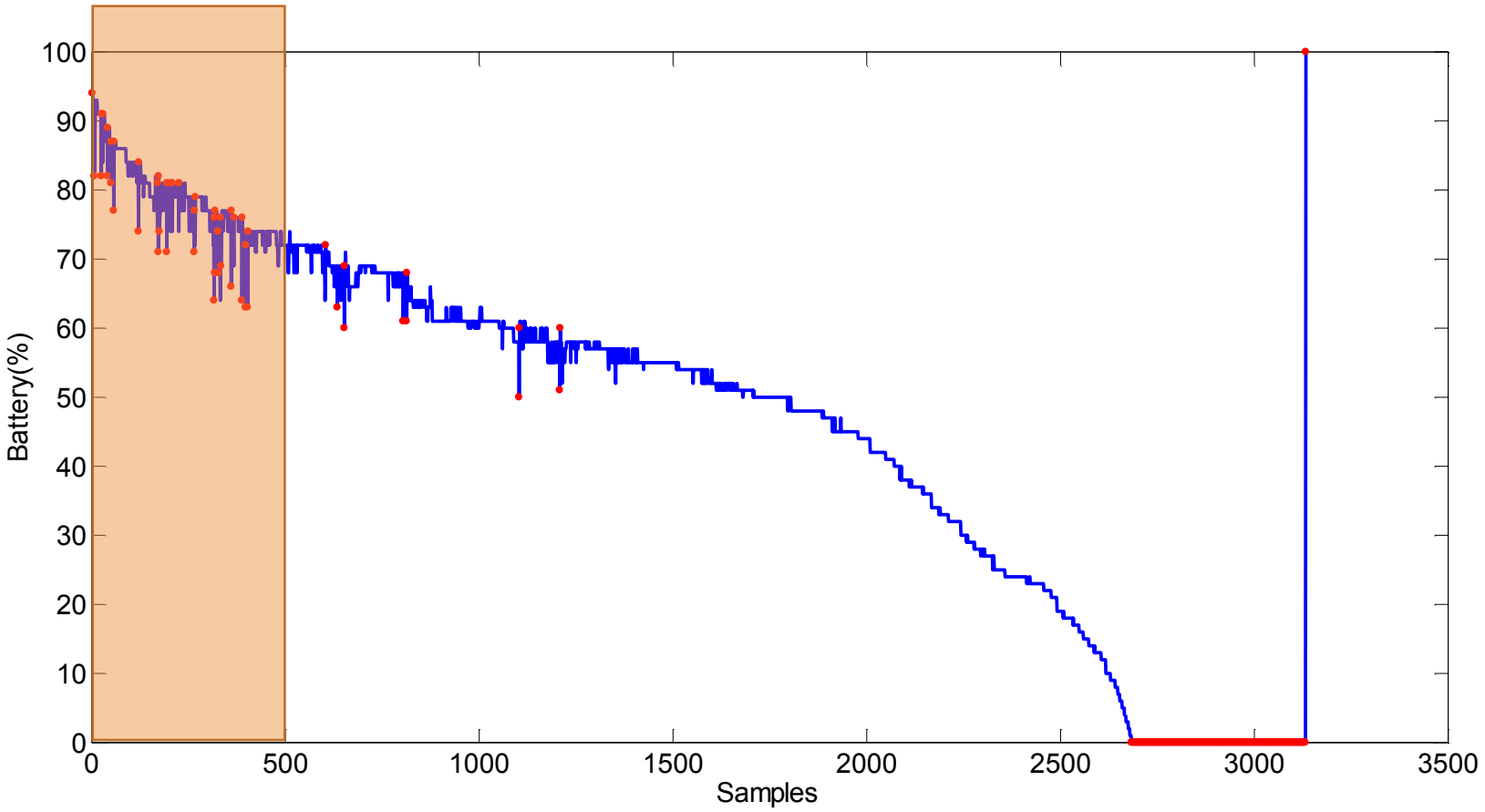
Example of Temperature Data



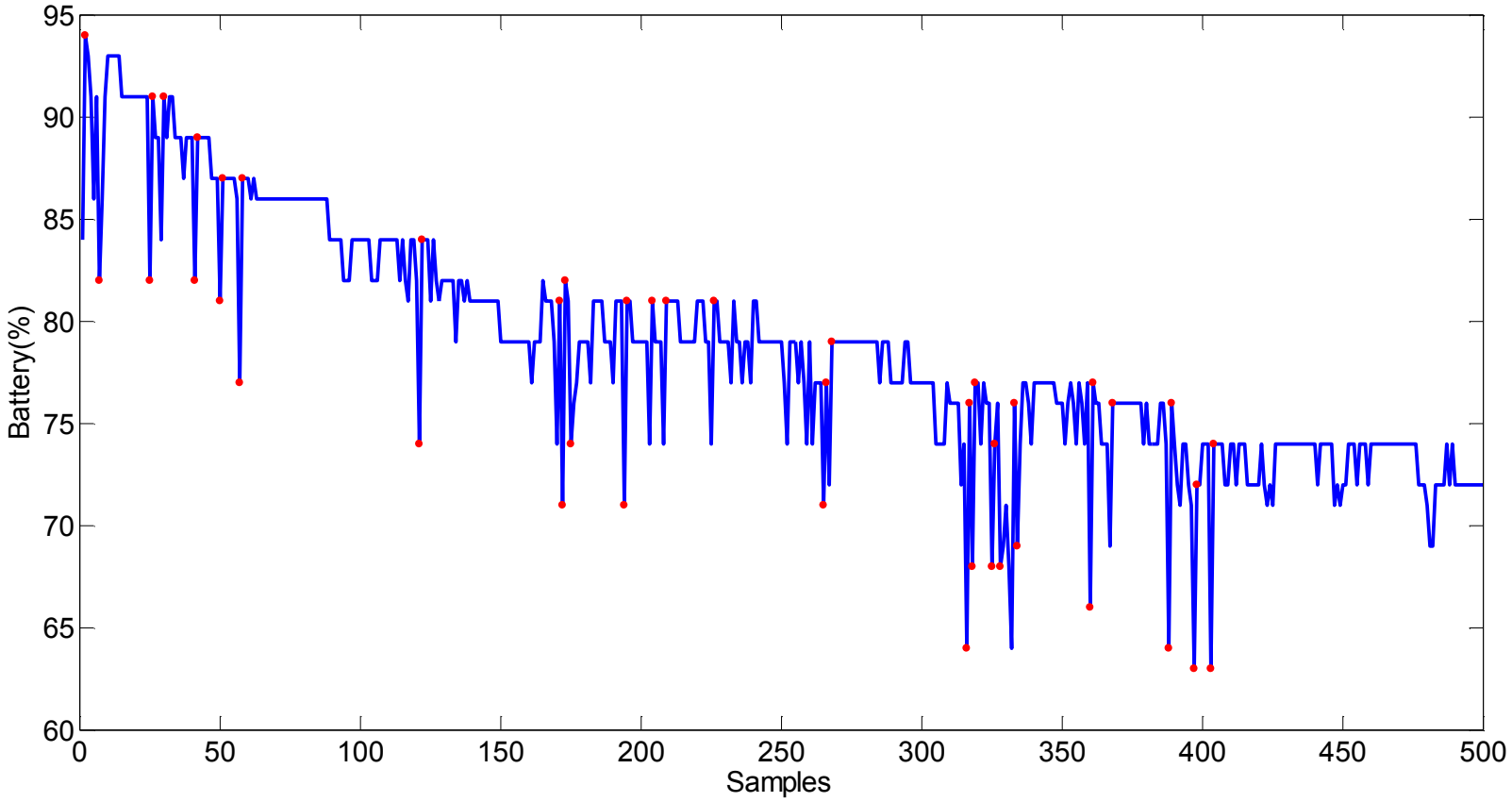
Example of Battery Data



Example of Battery Data

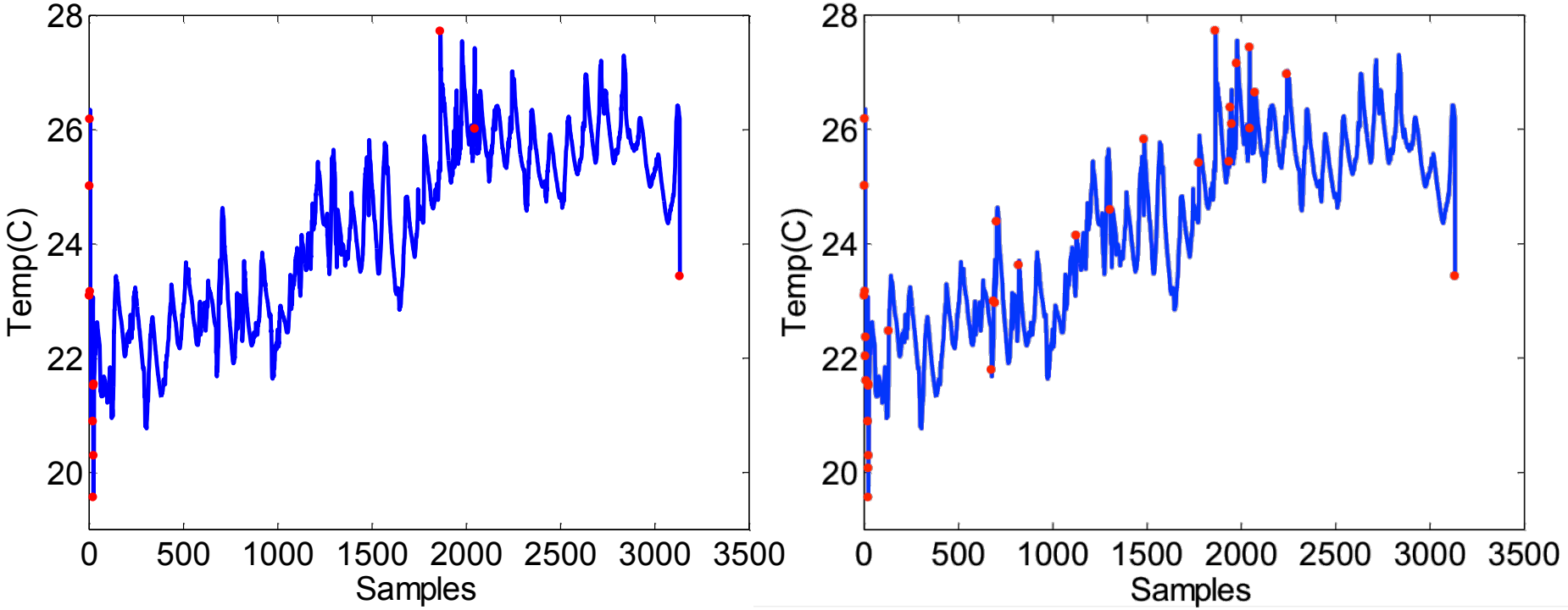


Example of Battery Data



Effect of Different Variance

± 1 VS ± 0.5



Generalization

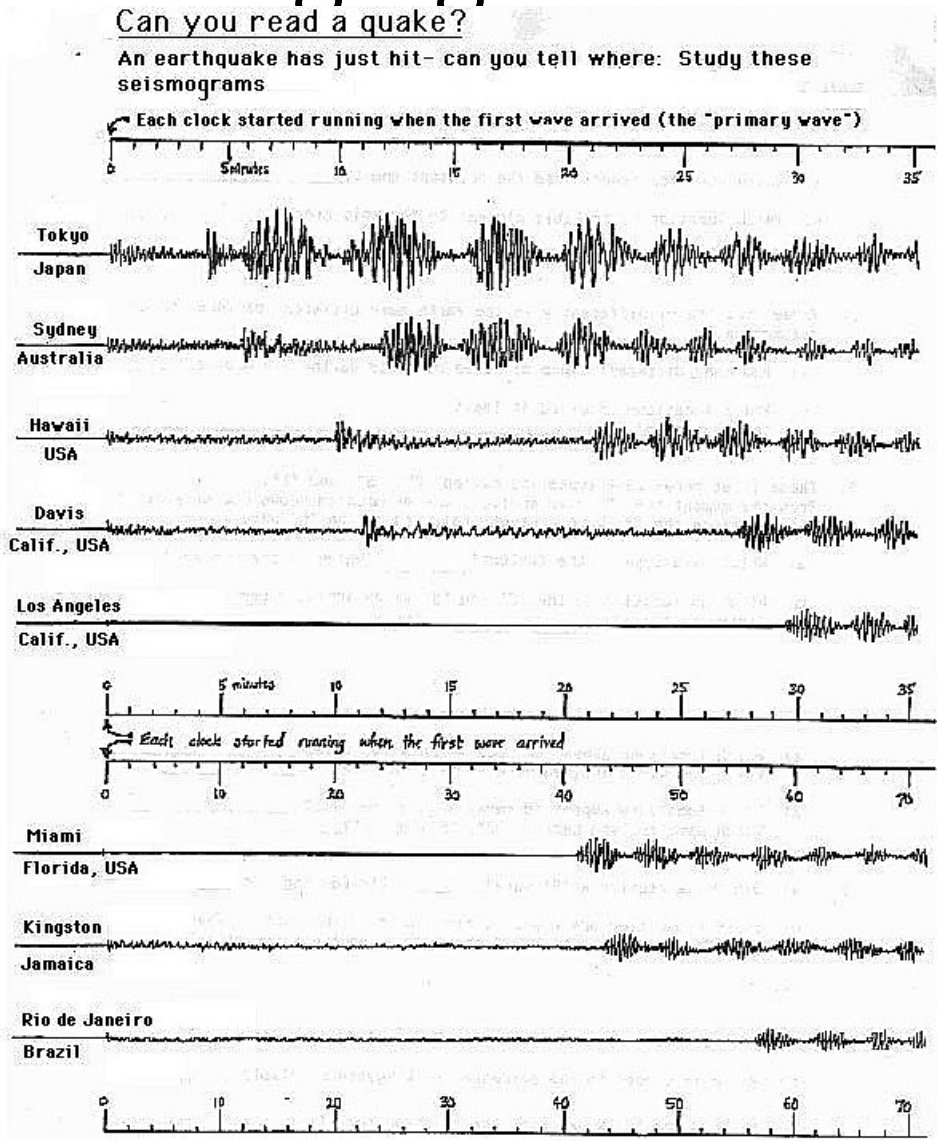
- Each data point
  - Null
  - Too high, too low

Generalization

- Each data point
  - Null
  - Too high, too low

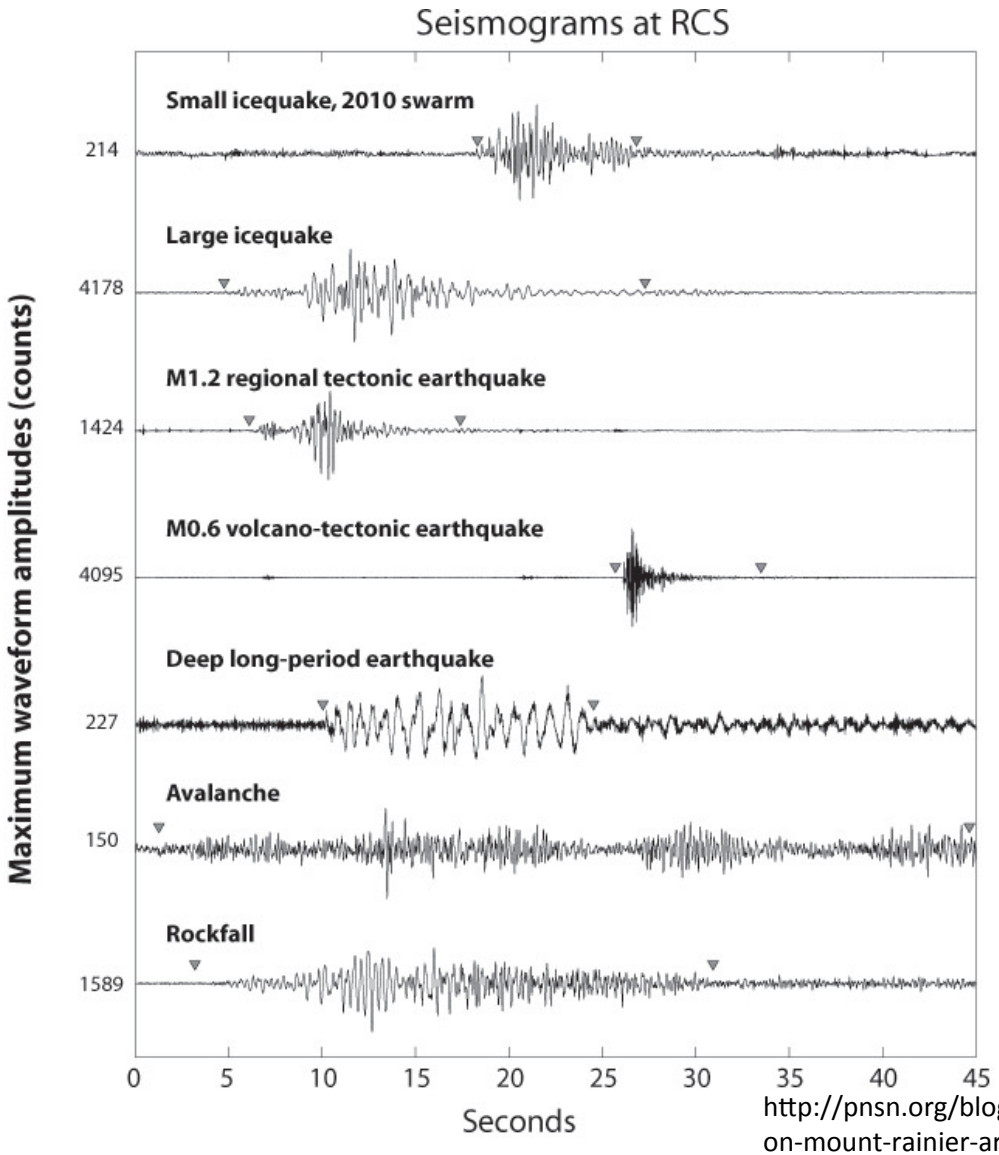
- Relation between different data points
  - Absolute variance too large
  - Fluctuation too large
  - Stays the same over certain period

More Challenging!



http://cse.ssl.berkeley.edu/img/earthquakes/seismogram000.jpeg

And more…



http://pnsn.org/blog/2012/01/06/repeating-earthquakes-on-mount-rainier-are-glaciers-the-culprit

Thanks!

Questions?