Incorporating predictive analytics in an actuarial curriculum: some preliminary experience at the Chinese University of Hong Kong

Wai-Sum Chan, PhD, FSA, CERA, FRSS
Professor of Finance
The Chinese University of Hong Kong

SOA's 2018 ASA Curriculum Changes

SOA's 2018 ASA Curriculum Changes

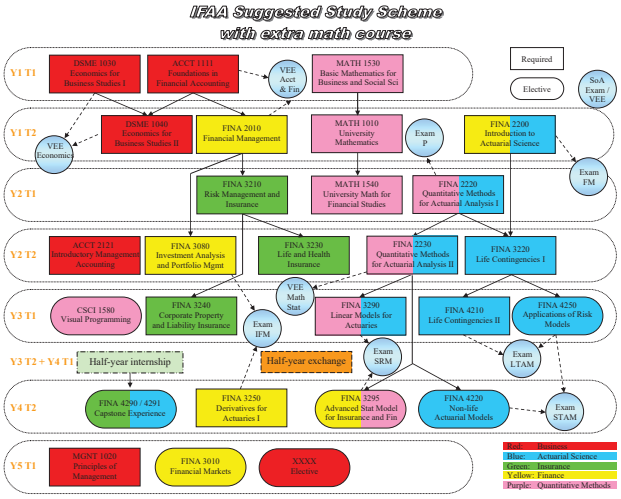
← At CUHK, we do not have a specific course for Predicitive Analytics yet

SOA's 2018 ASA Curriculum Changes

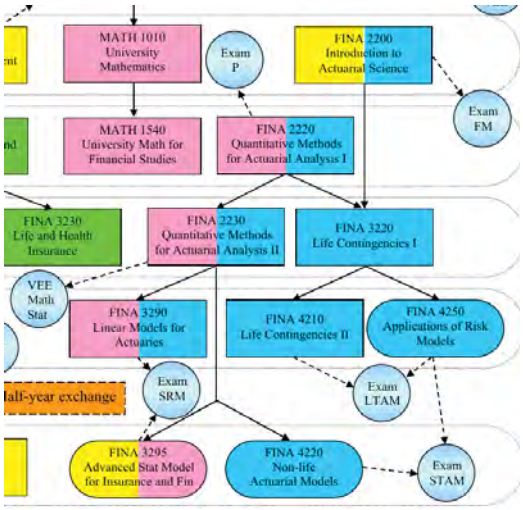
← We have two courses here to equip students with skills of R programming, regression & time-series predicition

← At CUHK, we do not have a specific course for Predicitive Analytics yet

CUHK Students' 2017 Study Plan



CUHK Students' 2017 Study Plan



A Pilot Case Study for FINA3290

- The National Data Buoy Center (NDBC) is a part of the National Oceanic and Atmospheric Administration's (NOAA) National Weather Service (NWS) of the US government.

A Pilot Case Study for FINA3290

- The National Data Buoy Center (NDBC) is a part of the National Oceanic and Atmospheric Administration's (NOAA) National Weather Service (NWS) of the US government.

- NDBC deploys weather buoys which are instruments which collect weather and ocean data within the world's oceans.

A Pilot Case Study for FINA3290

- The time-series weather data for each buoy are publicly available from the NDBC website (www.ndbc.noaa.gov).

A Pilot Case Study for FINA3290

- The time-series weather data for each buoy are publicly available from the NDBC website (www.ndbc.noaa.gov).

- These data have been used for research (e.g., Chen, Ruf and Cleason, *Journal of Geophysical Research: Oceans*, April 2016) and teaching purposes.

Part (A) - constructing the dataset

- Students are asked to locate the data webpage of the Weather Station buoy 46035 at 57.026 N 177.738 W from NDBC.

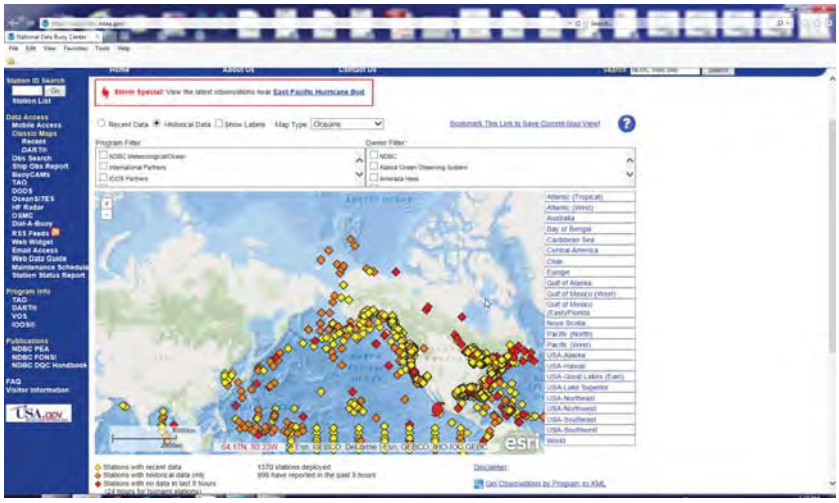
Part (A) - constructing the dataset

- Students are asked to locate the data webpage of the Weather Station buoy 46035 at 57.026 N 177.738 W from NDBC.
- Examine the data format for each yearly data file.

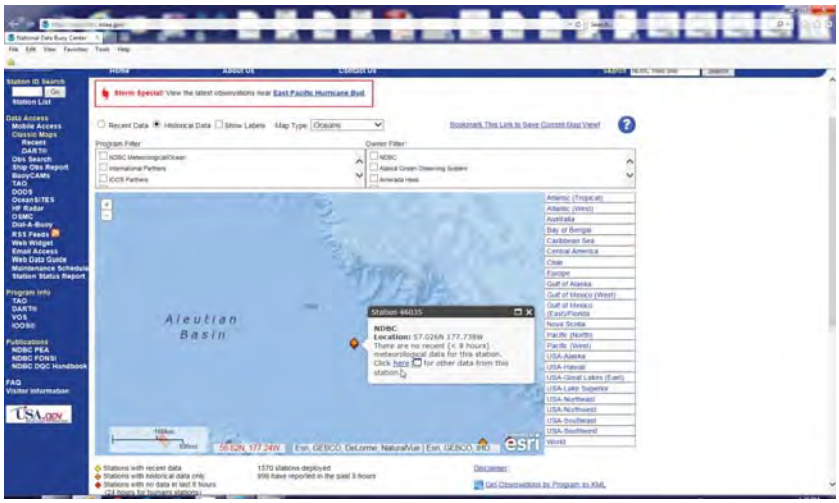
Part (A) - constructing the dataset

- Students are asked to locate the data webpage of the Weather Station buoy 46035 at 57.026 N 177.738 W from NDBC.
- Examine the data format for each yearly data file.
- Write an R program to extract and patch the data into two time-series of daily **Air Temperature** and **Sea Temperature** readings recorded at noon.

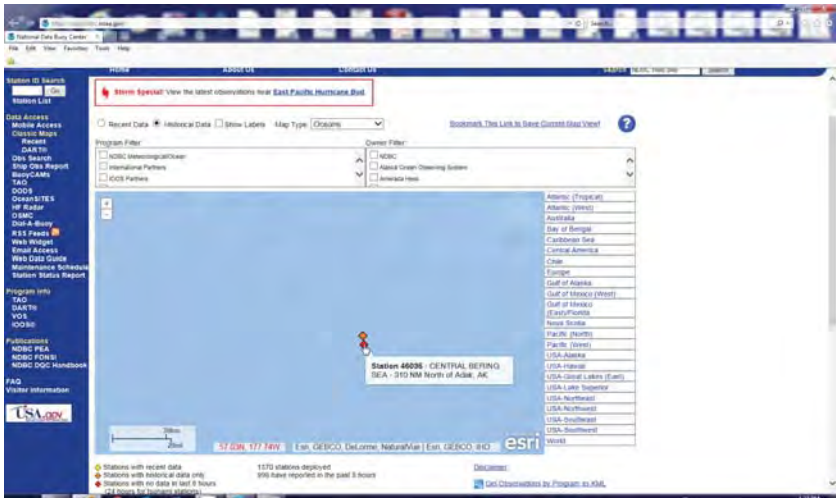
Part (A) - constructing the dataset



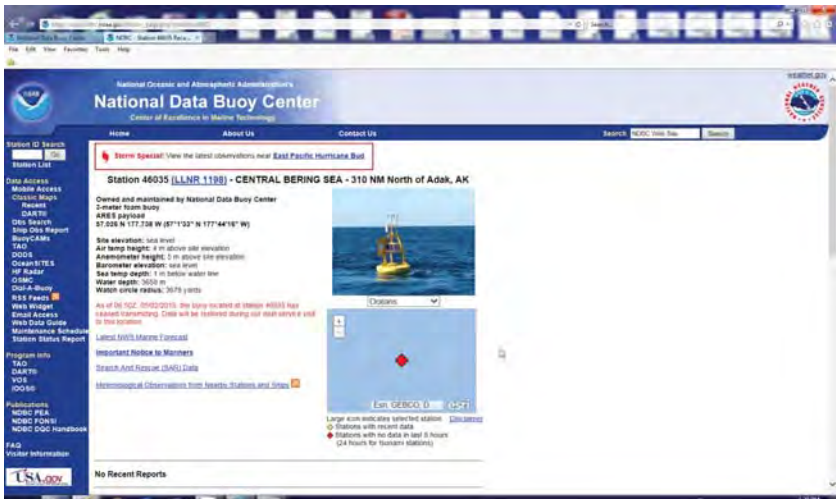
Part (A) - constructing the dataset



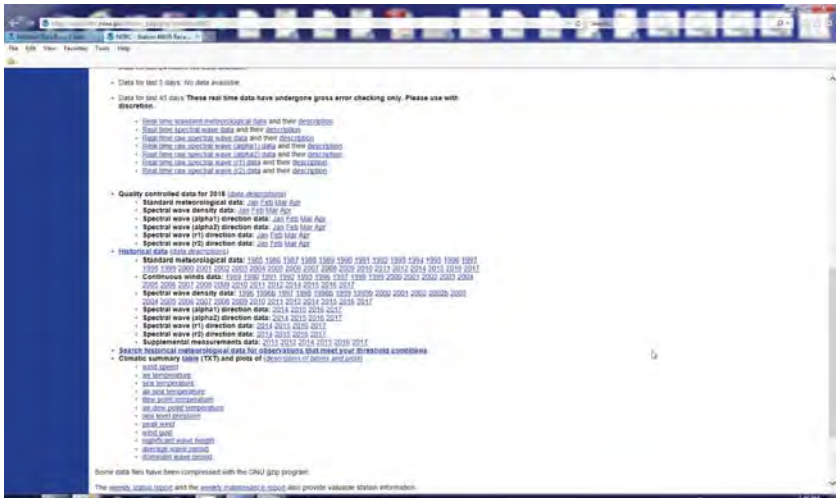
Part (A) - constructing the dataset



Part (A) - constructing the dataset



Part (A) - constructing the dataset

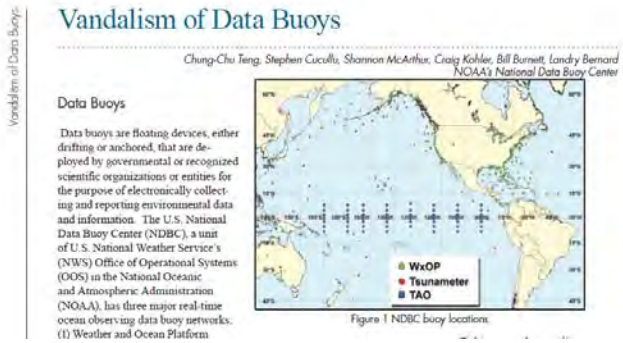


Part (B) - data cleaning

- Students are asked to plot and clean the data.

Part (B) - data cleaning

- Students are asked to plot and clean the data.
- Messy data: outliers, missing values, lost of data – due to vandalism/stolen of data buoys



Vandalism of Data Buoys

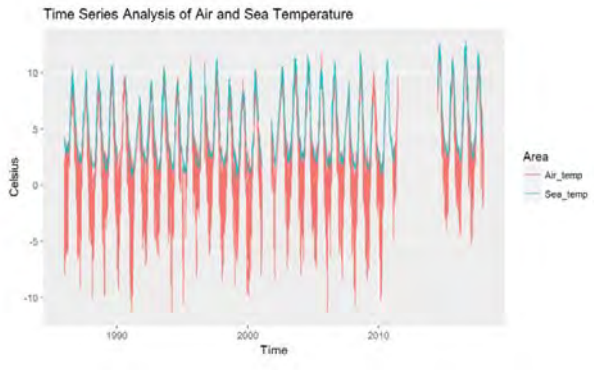
Chung-Chu Teng, Stephen Cucullu, Shannon McArthur, Craig Kohler, Bill Burnett, Landry Bernard
NOAA's National Data Buoy Center

Data Buoys

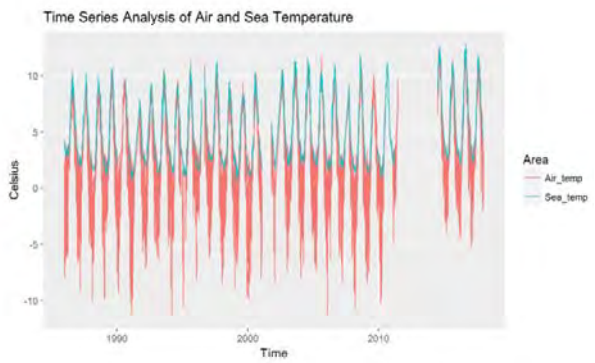
Data buoys are floating devices, either drifting or anchored, that are deployed by governmental or recognized scientific organizations or entities for the purpose of electronically collecting and reporting environmental data and information. The U.S. National Data Buoy Center (NDBC), a unit of U.S. National Weather Service's (NWS) Office of Operational Systems (OOS) in the National Oceanic and Atmospheric Administration (NOAA), has three major real-time ocean observing data buoy networks. (1) Weather and Ocean Platform

Figure 1 NDBC buoy locations

Part (B) - data cleaning



Part (B) - data cleaning



- Students have to research and decide on how to clean the data.

Part (C) - the research question

- Students are asked to answer the question: Global warming - have the temperatures (both sea and air) increased over the past 30 years?

Part (C) - the research question

- Students are asked to answer the question: Global warming - have the temperatures (both sea and air) increased over the past 30 years?

- Students can use any statistical methods learned in this course.

Part (C) - the research question

- Students are asked to answer the question: Global warming - have the temperatures (both sea and air) increased over the past 30 years?
- Students can use any statistical methods learned in this course.
- All computations have to be carried out in R.

Part (C) - the research question

- Students are asked to answer the question: Global warming - have the temperatures (both sea and air) increased over the past 30 years?
- Students can use any statistical methods learned in this course.
- All computations have to be carried out in R.
- Two students form a team.

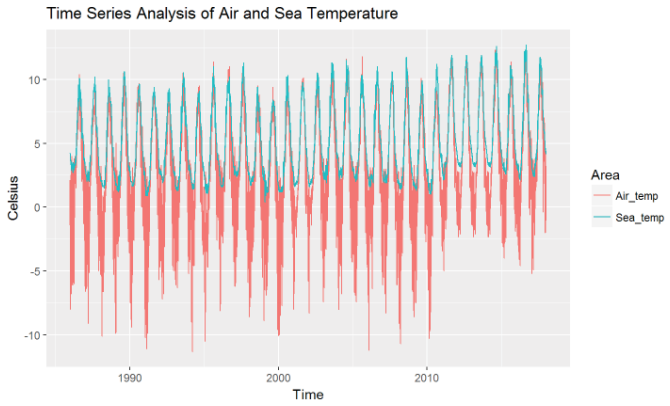
15 of 25

Part (C) - the research question

- Students are asked to answer the question: Global warming - have the temperatures (both sea and air) increased over the past 30 years?
- Students can use any statistical methods learned in this course.
- All computations have to be carried out in R.
- Two students form a team.
- Each team has to make a presentation and hand-in a final report (professionally written with proper conclusions and justifications).

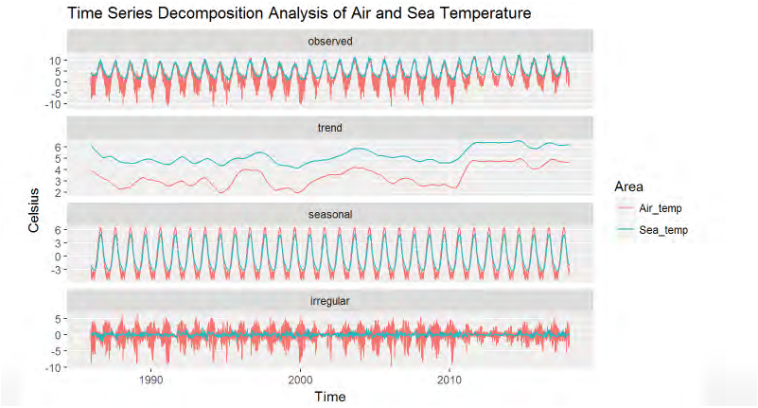
Report-1

Cleaned Data



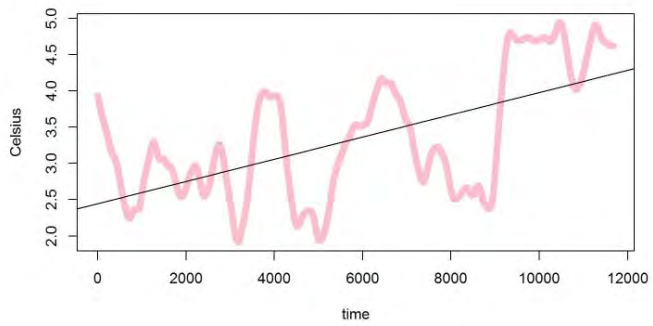
Report-2

Seasonal Decomposition



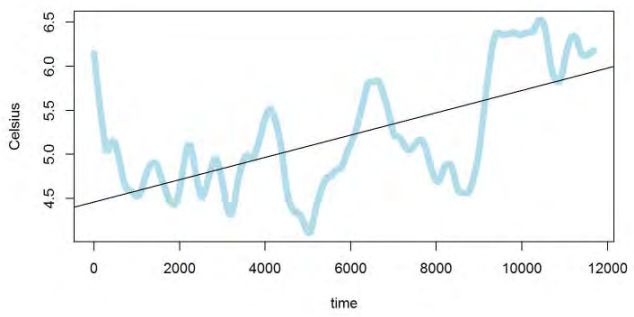
Report-3

Trend: Air temperature



Report-4

Trend: Sea temperature



Report-5

Robustness Check

To check whether sampling affected our evaluation of temperature change, lets conduct simple robustness check when temperature = 0:00, 6:00,18:00 and compare with 12:00.

```
##             tau       sl S        D        varS
## Air.Time_0  0.3382657 0  23083382 68240392 177205149696
## Air.Time_6  0.3505084 0  23918828 68240392 177205149696
## Air.Time_12 0.3590287 0  24500260 68240392 177205149696
## Air.Time_18 0.3716339 0  25360442 68240392 177205149696
```

```
##             tau       sl S        D        varS
## Sea.Time_0  0.3823255 0  26090036 68240384 177205149696
## Sea.Time_6  0.3794484 0  25893702 68240384 177205149696
## Sea.Time_12 0.3917121 0  26730586 68240384 177205149696
## Sea.Time_18 0.393612  0  26860232 68240384 177205149696
```

Remarks

- Students are forced to get familiar with R programming.

Remarks

- Students are forced to get familiar with R programming.
- Students generally like the case study, but complained the heavy workload and unclear instructions.

21 of 25

Remarks

- Students are forced to get familiar with R programming.
- Students generally like the case study, but complained the heavy workload and unclear instructions.
- Most students indulged too much in the R coding part of the project, without a good understanding of the statistical methods they used.

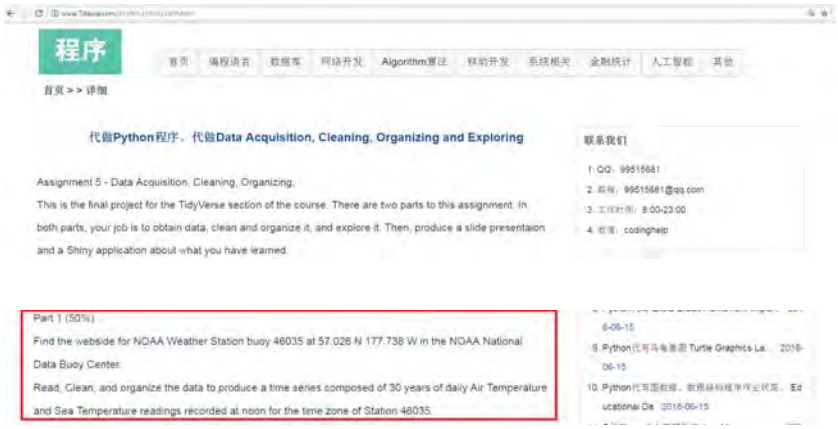
Remarks

- Students are forced to get familiar with R programming.
- Students generally like the case study, but complained the heavy workload and unclear instructions.
- Most students indulged too much in the R coding part of the project, without a good understanding of the statistical methods they used.
- Students need more training in report writing.

Other Issues: (I) Plagiarism

Other Issues: (II) Academic Honesty

Other Issues: (II) Academic Honesty



Other Issues: (II) Academic Honesty

THANK YOU!