Internship Report

Jin Kweon

Fall 2016

Supervisor: Sergio Castellanos / Deborah Sunter

Berkeley Energy and Climate Institute

SWITCH Mexico

Contents

1. Summary

Going through all the application and interview process, it was an honor for me to get accepted in Daniel Kammen's lab to work in the Energy and Climate Institute. First day I went to the office, I met four other undergraduate peers and two supervisors. To tell the truth, it was really hard for me to get familiar with the topics they talk about for our research. Frankly speaking, I am not the one who are super into the Energy and Climate issues, so it was very challenging for me to know the "topics," so I can do better job in the team.

My first assignment was to read the former visiting researcher's report, so I could have an idea what our research group is working on. Also, my supervisor, Sergio assigned me some of the readings to find which method I had to go for to predict the future hydro-plant powers. Since one of my biggest questions to answer for this research is "can we forecast future hydropower generation with the data that we currently have," I had spent most of the times collecting the articles and analyze other people's thoughts. I had read around five big articles/reports to summarize what we can do to predict the future hydropower generations based on the historical data our lab has. I had learnt so many different aspects of ideas different regions and people work on to get better friendly environmental energy. These processes definitely attract me into the energy and climate issues. Although I could not contribute so many things other for researches, I firmly believed that the experience and lessons I learned from this research will be enriching and will help me forever.

Another big thing I worked on was managing the data. I downloaded so much data (big data) from the specific website my supervisor gave to me, and I scraped and cleaned the data by using the knowledge I am learning in Stat 133 (R class). It was so fun for me to realize that I can use the skill and computer language skills I learned from the school in real life.

I am assigned around 4-6 hours of works a week, and our research group usually have a weekly meeting every Monday. Actually, this is my first semester in UC Berkeley, and I could not imagine that I could join this great research team and learn a lot from the other peers and supervisors. I learned that there are so many global issues I need to contribute in the future, and to do that, I also realized that I should find a field I am interested in and passionate of. This semester was so enriching for me. This research experience is my first time utilizing my knowledge and skills into the real-world issues.

2. Activities

The big two parts I did for this semester were: 1. Forecast future hydro-power generation 2. Manage the data.

Forecasting the hydro power generation is the biggest and the most time-consuming task I had had for this time (but, this is very important and I am going to explain it later this report). There are several different literatures I read to research about it, but I had picked the four most important out of them, and summarized them. The names of them are “Applying a Correlation Analysis Method to Long-Term Forecasting of Power Production at Small Hydropower Plants,” “Internship Report (written by previous researcher, Alejandra),” “Projecting changes in annual hydropower generation using regional runoff data,” and “The Role of Rainfall Variability in Reservoir Storage Management at Hydropower Dam.” All of the literatures are explaining their trials to predict the future generations. I found that they are all helpful for our team to predict the future’s, because they are all trying different ways and regions, so our team could learn different perspectives from their works. So, I am going to explain their works from now on.

There are two categories for hydro-plants: Short term hydro-plants (SHP) and Long term hydro-plants (LHP). For SHP, there are so many disadvantages. For example, it is hard to collect / manage information since not much data has been accumulated. Second, SHP is typically derived from run-off river plants with little regulating capacity. Because the installed capacity of a single plant is very small, it is difficult to forecast its power production because of strong uncertainty and fluctuation. Third, the power production of SHP plants shows great spatiotemporal diversity, and thus, it is difficult to establish a commonly used model. If wanting to establish it, we are required to attain so much data. Last, because SHP is widely distributed and large in quantity, building forecasting models for all SHP plants is unnecessary and cannot satisfy the demand of precision. Different data for each model requires too much work. To be more specific about the disadvantages of SHP, here are the diagrams and explanations of them below:

First characteristic of SHP is “randomness of a single plant.”

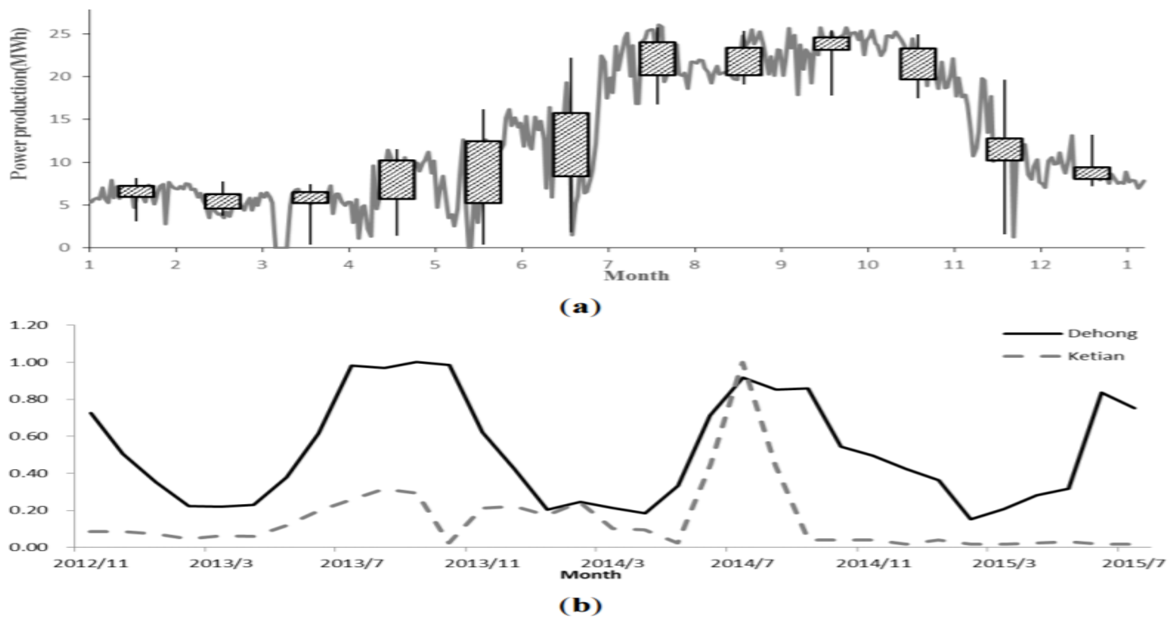

Figure 1 - Randomness of a single plant. Image from Ref [1]

- Influenced by rainfall and it is difficult to achieve stable regularity of power production
- More outliers, stronger randomness and greater fluctuation

Second characteristic is “spatial differences.”

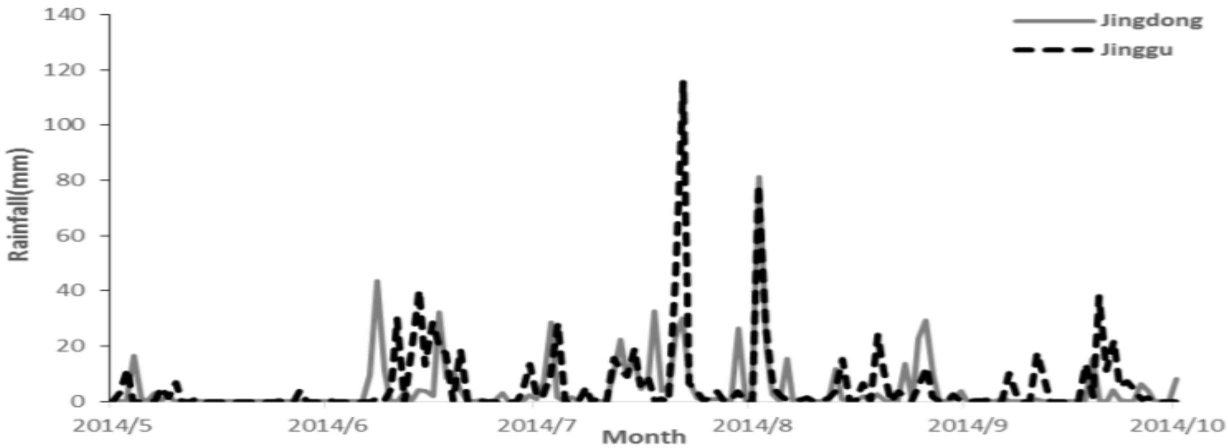

Figure 2 - Variation of rainfall across different months in Jinggu and Jingdong. Image extracted from Ref [2]

- Each plant has different topograpy and land forms, the power production presents spatial differences.
- Significant differences in rainfall and runoff formation occur because of Wuliang Mountain, Ailao Mountain and other mountains in this area

Last characteristic is "similarity of regions."

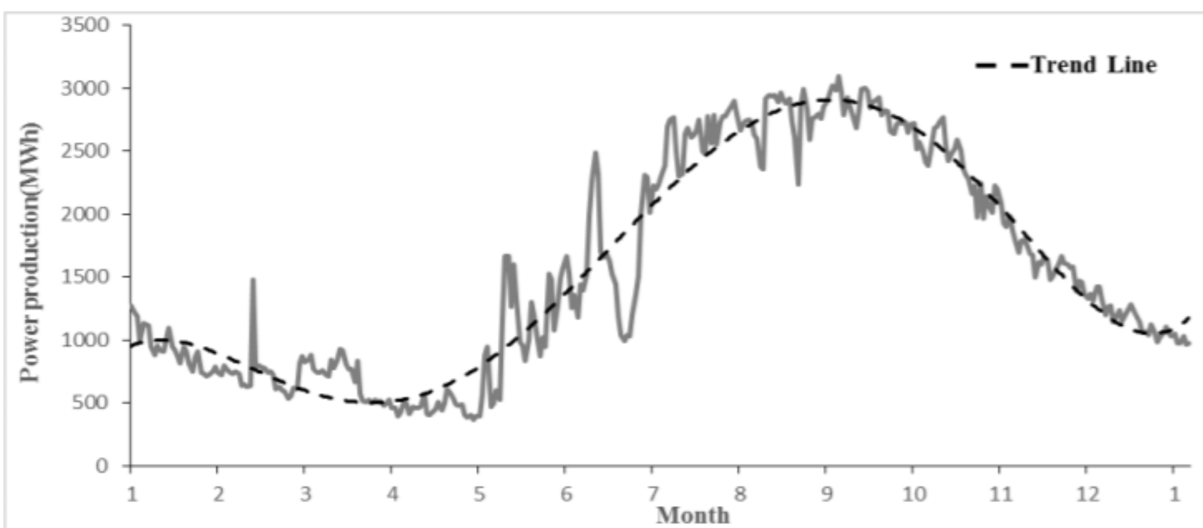

Figure 3 - Show the trend line for similar regions. Image from Ref [1]

- Similarities in terms of hydrological and meteorological conditions

- Monthly power production and trend line of SHP plants in county.

Based on these characteristics, I read four articles I mentioned above. Alejandra, the former researcher [2] tried to find which factors are linked to the production of certain hydro power stations. Her goal was to predict the behaviors of hydro stations given input variables as amount of precipitation and drought levels. So, she looked into the precipitation levels and correlate them with energy production. (http://forecast.io) In order to find a relationship between climate data and the net production of all reservoirs Hydro power stations, she applied a linear regression using the ordinary least squares method. She concluded that those two variables have something to do with hydro power production, yet the correlations are too low to be considered meaningful. The challenges she faced were lack of data, so she could find more factors that have relationships with hydro power productions.

Second report [1] was trying to apply a correlation analysis method to long-term forecasting of power production at small hydropower plants (I personally found that this method was very interesting, since it is really realistic way). SHP (small hydro-plants) are hard to be estimated compared to LHP (large and medium hydro-plants) due to difficulty of information collection, lack of available data accumulation, difficulty to forecast its power production because of strong uncertainty and fluctuation. The power production of SHP indicates the maximum generation capacity of an SHP plant under certain meteorological and hydrological conditions, which are easily influenced by factors, such as hydrology, climate, and installation capacity. Therefore, forecasting the power production of an SHP plant is a *nonlinear, multi-factor complicated problem*. SHP and LHP plants in the same region or neighborhood have similarities in some aspects of hydrology, meteorology and geography. As result, some correlation exists between SHP and LHP in some aspects. LHP plants can provide long-term historical data and accurate forecast values. Thus, historical data are used as a reference for the study of forecasting long-term SHP power production. LHP plants can provide long-term historical data and accurate forecast values. Thus, historical data are used as a reference for the study of forecasting long-term SHP power production. Making a regression model is developed to forecast the power production of SHP through the predictive value of LHP. Here are steps to make a correlation:

1. SHP plants are considered to belong to the same area as a whole
2. Interval flow of the LHP plant is screened as the correlation factor.
3. Correlation analysis of LHP and SHP is performed, and the significance of the correlation is tested.
4. A regression model is developed to forecast the power production of SHP through the predictive value of LHP.

In here, I found that stronger correlation between power production of SHP plants and the interval flow of the LHP plant through the numerical simulation analysis is needed. The geographical and hydrological conditions of SHP plants are similar to that of LHP plant in the same region or neighborhood. Interval flow of a set of candidate factors of LHP plants can reflect the runoff in the region where SHP plants are located, and there must be some correlation between the interval flow of LHP plants and the power production of SHP plants in the same region. Then the correlativity of LHP and SHP in the region can be established. One case study is executed in one district in China.

Figure 4 - SHP and LHP plants that are closed to each other. Image from Ref [1]

Correlation exists between the power production of SHP plants in a district and the interval flow of the LHP plant. Power production of SHP plants shows significant correlation with the LHP plant of the same basin, whereas the correlation with the LHP plant in other basins is not significant. Power production of SHP plants shows spatial differences, the further the distance in space is, the less significant the correlation relationship is. Here are the pictures of the correlations below:

LHP Plant\District	Dehong	Puer	Kunming
Dayingjiang III	0.94 **	0.81 **	0.82 **
Dachaoshan	0.77 **	0.92 **	0.75 **
Ketian	0.46 *	0.39	0.89 **

otes: ** denotes significantly correlated at the 0.01 level, * denotes significantly correlated at the 0.05

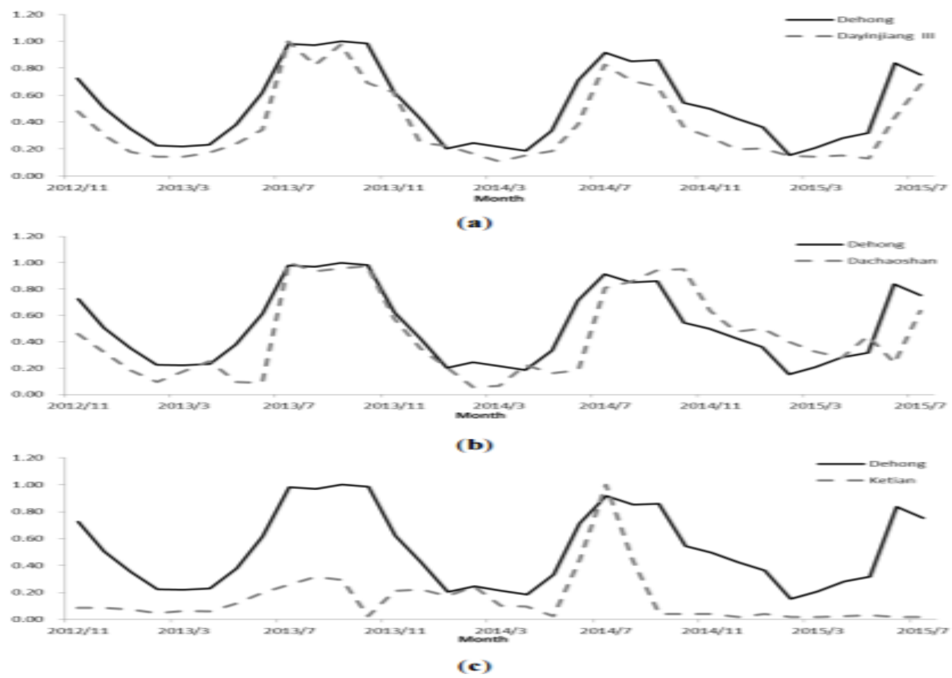

Figure 5 - Correlation results for each region. Image from Ref [1]

Here is the result of correlation analysis method.

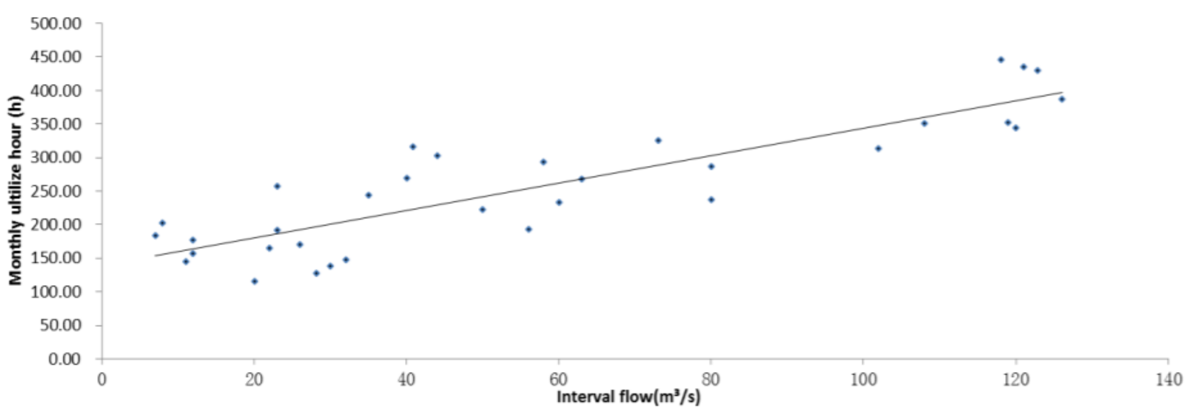

Figure 7. Linear curve fitting of LHP and SHP in Puer County.

Figure 6 - Linear regression of LHP and SHP. Image from Ref [1]

Here are the lists of the benefits of the method:

1. Great significance for realizing the coordination with LHP plants, solving the problem of the waste water and abandoned electricity and ensuring the safe operation of the power grid.
2. This method could achieve an accurate result, even during the flood season, which can provide the reliable forecasted data for arranging the coordinate power generation schedule between SHP plants and LHP plants.
3. Not much data is required. 1) only spatiotemporal characteristics, 2) the correlation between the power production of SHP plants and the interval inflow of LHP plants, 3) correlation characteristics in space
4. Generality of the correlation analysis method is demonstrated.
5. Finally, prediction accuracy of this method has been verified.

However, there are some challenges against this method as well. This method worked to forecast for SHP power generation given the data for LHP. However, in Mexico, most of them are LHP and we do not need to predict for SHP. So, we need to confirm whether this method is working to forecast LHP as well. I concluded that power production of SHP plants shows significant correlation with the LHP plant of the same districts, whereas the correlation with the LHP plant in other districts is not significant.

Third report [3] was about projecting changes in annual hydropower generation using regional runoff data. The basic idea for this method came from that climate change impacts on selected hydropower plants. Given the complexity of surface water storage, management, and distribution systems, and the proprietary nature of existing hydropower models and data, it would likely be very costly and time consuming to develop a large-scale energy water model through a conventional reservoir-based approach. Simplified energy method has been used to evaluate climate change impacts. Regression is used to predict the average annual stream flow and hydropower generation from the winter/summer precipitation fraction and temperature change so that the assessment can be expanded to cover more climate change models. In the present study, a runoff-based alternative approach was developed to project the change in annual hydropower generation. To summarize the historic and modeled precipitation and runoff for a hydropower plant, the contributing upstream drainage area was delineated and used to compute the spatially averaged precipitation and runoff from existing data sets. Based on the geographical coordinates of the most downstream hydropower plants in each river system in each study area. Motivated by the high correlation between Water runoff and hydropower generation, they conducted a series of regression analyses based on the annual time series of temperature, precipitation, runoff, and hydropower generation for each of the study areas. Empirical relationships between generation and runoff are key tools because they enable projected changes in future runoff to be translated into projected changes in annual generation. Here are the challenges of the method:

- This generalized approach works for US federal hydropower plants because their total installed power capacity has not changed significantly during the past 20 years. (Mexico might be different)
- This linear relationship, while working properly between annual runoff and generation, cannot be extended to seasonal or monthly scales.
- This method works for SHP prediction.

Last report [4] was trying to find the role of rainfall variability in reservoir storage management at hydropower dam. (This is the only LHP method amongst four) This method was suggested from the fact that reservoir operation and management is usually patterned after the background of long standing water resources management experience. Records of rainfall-runoff relations for the study area suggests some level of variability in runoff due to rainfall variability; a condition which may impact on reservoir water balance. The objective is to develop functional hydrological relationship between (rainfall, inflow, reservoir storage and turbine releases) over the dam. This will provide scientific basis for operational decisions which can lead to optimum power plant utilization. Correlation coefficients of the main meteorological variable (rainfall) on hydrologic and reservoir variables were computed to determine the strength of the relationship between the variables. These variables are: rainfall, reservoir inflow, reservoir storage and turbine release. A correlation coefficient of rainfall of 0.770 reveals a very strong relationship. The moderate rise in inflow of about 0.439 (44%) indicated non-significant decrease in inflow into Shiroro dam. The turbine release of about 12% and the corresponding increase in reservoir storage of 11% has helped maintain all year round power generation at the dam.

The results indicate a significant positive uptrend in rainfall during the study period. A

Table 2. Correlation coefficient between rainfall and selected Reservoir variables.

Variable	Correlation Coefficient
Rainfall (mm)	0.770
Reservoir inflow (m^3/sec)	0.439
Reservoir storage (m^3/sec)	0.105
Turbine release (m^3/sec)	0.120

Table 1 - Correlation Coefficient between rainfall and three variables. Image from Ref [4]

correlation coefficient of rainfall of 0.770 reveals a very strong relationship. The moderate rise in inflow of about 0.439 (44%) indicated non-significant decrease in inflow into Shiroro dam (Fig 4). The turbine release of about 12% and the corresponding increase in reservoir storage of 11% (Figs 5 and 6) has helped maintain all year-round power generation at the dam. This has positive effect in the dam operations unlike the situation at Kainji dam, where Oluwatosin and Adeyemo (2011) asserted that the generous turbine release at Kainji hydropower dam has been propitious as it has helped maintain hydropower generation from the dam over the years, thus having an insidious effect on the operation of the dam as it resulted in the significant drop in the storage of the dam.

Table 3. Trend line equations.

Variable	Equations
Rainfall (mm)	$y = 0.007x + 6.03$
Reservoir Inflow (m^3/sec)	$y = 0.811x - 210.2$ $\{\mathbf{R^2} = 43.9\}$
Reservoir Storage (m^3/sec)	$y = 5.735x + 2248.0$ $\{\mathbf{R^2} = 10.5\}$
Turbine Release (m^3/sec)	$y = 0.811x + 160.2$ $\{\mathbf{R^2} = 12.5\}$

Table 2 - Linear fitting equation between rainfall and three variables. Image from Ref [4]

The three months, July, August and September, together represent the period of highest rainfall at Shiroro and consequently significant reservoir inflow harvest. This has significant positive impact on reservoir storage. Strong limitations are therefore, imposed on power generation at Shiroro after rainfall cessation in all hydrological year. Water use pattern always takes cognizance of this trend in the annual reservoir filling and emptying cycles.

It also reported a strong and positive association between rainfall characteristics and the examined reservoir variables (reservoir inflow, reservoir storage and turbine release. Therefore, it may be concluded that the non-significant variation in rainfall amount as a result of the uptrend in rainfall amount, favors increased reservoir inflow and thus increase in reservoir storage. The optimization of turbine release as part of the operation rule for careful water management also safeguarded reservoir storage that sustain year-round power generation at Shiroro dam. It is therefore recommended that dam operators maintain water optimization policies, ensure continuous monitoring of the various hydro climatic variables to provide early warning systems for effective performance of the dam and to protect downstream environment from hydrologic mishaps. In addition, there should be consideration for commencement and development of a pump water storage systems such that the tail water can be re-use particularly during periods of low inflow. One of the biggest changes for this method was:

- Hydrologic Variability which is jointly influenced by precipitation, evaporation, snowfall, soil moisture, groundwater and reservoir operation was found to be a significant factor in the annual hydropower generation. However it is hard to track all the conditions.

I came up with the conclusion after going over all of four articles:

Given the complexity of energy water facilities and limited information of existing hydropower data, it is unlikely to address all issues across multiple power plants and scales using ***a single model***.

Depending on the scale (e.g., national, regional, or plant-wise) and the nature of the problem (e.g., reservoir operation, risk assessment, or resource planning), one would likely need to develop different kinds of models for different challenges.

These literature reviews were so enriching and they led me into the research.

After the literatures review, I had short time to manage the data. I downloaded all the files from 2006 and 2015 from the website: http://app.cfe.gob.mx/Aplicaciones/OTROS/costostotales/ConsultaArchivoCostosyCapacidades.aspx. After I downloaded all of them, I cleaned those files and extract columns A, B, and F that correspond to "day," "plant name," and "output." Some of the data do not have enough data, so I could not include them in my cleaned excel, below:

2006 Mar, Apr, Nov, Dec

2007 Jan, Feb, Mar, Aug

2008 Aug,

2009 Feb, Mar, Jul, Sep, Oct, Dec

2010 Feb, Mar, Apr,

2011 Feb, Mar, May, Jun, Jul, Ago, Dec

2012 Feb, May, Sep, Oct, Nov, Dic,

2013 May, Ago, Sep, Oct, Nov

2014 Ene, Feb, Mar, Abr, May, Jun, Jul, Ago, Sep, Oct, Nov, Dic,

2015 Ene, Feb, Mar, Jun, Ago, Nov, Dic

I used R language (since I was taking R class – stat 133 – and I want to try my R skills into research). All the copies of the codes are also highlighted as well. Here is the summary of all the data for each excel file I downloaded:

Task 11.14 code plan

```
<Plan>
2006 (01-01 ~ 01-31): <I FOUND THAT THE NUMBER OF DAMS INCREASE AS TIME GOES>
ACLP1, AGEP1, AGMP1, ALTP1, ALXP1, AMDP1, AMIP1, ANGP1, APRP1, ARNP1, ATCP1, AZFGE,

BAJP1, BBNP1, BBRP1, BCUP1, BOQP1, BOTP1, BRTP1, BTSP1,

CABP1, CADP1, CAHCC, CBDP1, CBNP1, CCLP1, CCMP1, CCMP2, CCTP1, CDUP1, CDUP2, CECEP1, CFMP1, CIPP1, CJTP1, CLDP1, CLNP1, CLTP1, CLLP1, CMRP1, CNCP1, CODP1, COZP1, CPCP1, CPIP1, CPJP1, CPMP1, CPQP1, CPTP1, CPXP1, CRCP1, CREP1, CRLP1, CSOP1, CUHP1, CZUP1, CHBP1, CHLP1, CHPP1,

DBCP1, DBXP1,

EAAP1, EATCC, ECTP1, EFUP1, ELQP1, ENOP1, ENPP1, ENSP1, ENTP1, ESPP1,

FAMP1, FEHP1, FENP1, FETP1, FLEP1, FUNP1, FVLP1,

GACP1, GAOP1, GCPP1, GCPP2, GCPP3, GCPP4, GIRP1, GPPP1, GPPP2, GREP1,

HLIP1, HMSP1, HTSP1, HUIP1, HUOCC, HUTP1, HYAP1, HZTP1,

ICAP1, IELP1, INFP1, INJP1, INRP1, ITAP1, ITZP1, IXTP1, IZOP1,

JNSP1, JOLP1, JUAP1, JUMP1,

KIMP1,

LAEP1, LAVP1, LECP1, LEDP1, LEOP1, LERP1, LGTP1, LRAP1, LRPP1, LVOP1,

MAMP1, MDAP1, MDPP1, MDTP1, MICP1, MISP1, MMTP1, MMTP2, MNDP1, MONP1, MPSP1, MRIP1, MTOP1, MXCP1, MXIP1, MZDP1, MZTP1,

NCMP1, NCTP1, NECP1, NIZP1, NNGP1, NONP1, NPXP1, NVLP1,

OVIP1,

PEAP1, PEOP1, PGAP1, PGDP1, PGNP1, PGUP1, PJCP1, PJTP1, PJZP1, PLDP1, PLTP1, PMCP1, POEP1, PPIP1, PPXP1, PQEP1, PQSP1, PRIP1, PRXP1, PTCP1, PTLP1, PUDP1, PUIP1, PUNP1, PUSP1,

QRPP1,

RBCP1, RBTP1, RCAP1, RECP1, RIBP1, RITP1, RMAP1, RMNP1, RSCP1, RSMP1,

SATP1, SAUP1, SCHP1, SIMP1, SLAP1, SLMP1, SMPP1, SPUP1, SROP1, STOP1, SYCP1, SYDP1, TCBP1, TCCP1, TCOP1, TECP1, TEGP1, TEPP1, TETP1, TEZP1, TIIP1, TIJP1, TINP1, TJSP1, TLIP1, TMUP1, TMZP1, TPEP1, TPOP1, TROP1, TULP1, TUTP1, TUVP1, TXLP1,

UNIP1,

VADP1, VAEP1, VAEP2, VAVP1, VDMP1, VDRP1, VEEP1, VENP1, VEPP1, VICP1, VILP1, VIOP1, VLDP1,

XULP1,

ZEPP1, ZICP1, ZMNP1, ZMPP1,

02-01 ~ 02-28 (2006, 2007, 2009, 2010, 2011, 2013, 2014, 2015), 02-01 ~ 02-29 (2008, 2012)
03-01 ~ 03-31, 04-01 ~ 04-30, 05-01 ~ 05-31, 06-01 ~ 06-30, 07-01 ~ 07-31, 08-01 ~ 08-31, 09-01 ~ 09-30, 10-01 ~ 10-31, 11-01 ~ 11-30, 12-01 ~ 12-31


<Jin>

```

Line 1, Column 1 Tab Size: 4 Plain Text

Here is the codes I used to clean the data:

```
52
53  <Jin>
54
55
56  open "Catalog for hydro ALL CSV" data frame
57  open "each 12 months 2005-2015" data frame #Change it for every month
58  <assume both data frame have the edges of first row/column be headers of row/column> -> [2,2] of original == [1,1] of df
59
60  lettercurrent = each 12 months 2005-2015[1,2]
61
62  #for (i=1; i <=nrow(each 12 months 2005-2015); i++) -not need it!!!
63  {
64    if (number of data entries / number of days of each month = positive integer)
65    {
66      for (i=1; i <=positive integer; i++)
67      {
68
69        for(j=1; j <= nrow(Catalog for hydro ALL CSV); j++)
70        {
71
72          if (lettercurrent == Catalog for hydro ALL CSV[j,1])
73          {
74            for(k=1; k<= number of days of each month; k++)
75            {
76            Catalog for hydro ALL CSV[j, 8 +(k-1)] = each 12 months 2005-2015[1+(k-1),3]
77            }
78            *Go straight to line 83!
79          }
80
81        }
82
83        lettercurrent = each 12 months 2005-2015[1+ (i * number of days of each month),2]
84      }
85      write.csv(Catalog for hydro ALL CSV)
86    }
87
88    else
89    {
90    print ("error")
91    }
92  }
93
94
```

```
setwd("E:/Extracurricular activities/Research/Task/TASK 11.14")
Allhydro <- read.csv("Catalog for hydro ALL CSV.csv")
setwd("E:/Extracurricular activities/Research/Task/Hydro power monthly 2006-2016 cleaned/2006/cleaned")
Jan05 <- read.csv("real-cleaned Ene06.csv")
lettercurrent = trimws(as.character(Jan05[1,2]))
posinteger = length(Jan05[,1])%%31
typesdata = 7285/31
if (posinteger == 0)
{
for (i in 1:typesdata)
{
for(j in 1:nrow(Allhydro))
{
if (lettercurrent == as.character(Allhydro[j,1]))
{
for(k in 1:31)
{
Allhydro[j, 8 +(k-1)] = Jan05[k+((i-1) * 31),3]
}
break
# *Go straight to line 60
}
# lettercurrent = Jan05[1+ (i * 31),2]
}
lettercurrent = trimws(as.character(Jan05[1+ (i * 31),2]))
}
# write.csv(Allhydro)
} else
{
print ("error")
}
View(Allhydro)
write.csv(Allhydro,"cleanedALL.csv")
write.csv(Allhydro,"cleanedALL.csv")
```

Basically, I checked whether each file has enough data (for example, we know that January has 31 days, and I composed the code to check whether each hydropower plant 31 outputs. Unless they do, I made the code not run.) If they do have enough data, I picked the selected columns and appended into the newly created excel files. One of the biggest problem was that my codes do not work when they do not have enough data. As I mentioned, some of the data was not collected by Mexico government, and I want to regard them as “NA” and ran them. However, I could not figure them out due to the lack of my coding skills.

3. Recommendations

Before conclude the report, I want to give some recommendations to the future researchers.

First, always try to do team-work, instead of doing all the works on your own. Each researcher is selected since they are good at different parts, and those skills can be used in different ways and fields. Sometimes, supervisors can assign you the tasks that you are not capable of finishing it, and I want to advise you not to depress it, but take them as good opportunities to upgrade yourself. Find other researchers to work with and feel free to ask supervisors to get more helps. They are usually open to help each out.

Second, it is always better to have good communication skills. There will be regular meetings (usually weekly) and you need to share your progress and any problem you face. Be prepared to talk in front of the team, and don't be afraid to share obstacles with the team. For every meeting, they are always giving me the solutions for each problem I had. Also, they could be critical to your ideas, but it is not aggressive, but it helps you develop ideas and leads you to be a better researcher.

Third, always choose the field you are passionate of. In your first meeting (or during the interview process), you will be asked to share what you are good at. Every member will be assigned different tasks based on their majors and what you have learnt before. Although supervisors assign you some missions, you can always ask supervisors to work on different fields. Persuade them if you really desire to work on different teams or tasks. It is always better for you to work for something that you love. Because I personally found that the research is not a short-term work, and it especially takes so much time for the researchers who had no such research experience before. Since it is a long-term work, don't be depressed just because you are not good enough to contribute for the team. I personally took almost the entire semester to learn what the research team is really doing. Just believe yourself, and read many related articles for your benefits. While researching, you can definitely tell that it will be smoothly worked for you if you love into.

Fourth, do cooperate with other co-researchers. Always believe in that you are not alone. All of them are friendly enough to work with you. Do not compete with any member in your team, but try to work together to achieve a goal.

Last but not least, I am so proud of myself, supervisors, and other co-researchers to contribute on the SWITCH team. This team is formed for friendly-environment purpose, and whoever work for this team, I want to advise them to be proud of themselves that they are actually contributing on the world issues.

4. Code

4.1. Data Extraction

```
open "Catalog for hydro ALL CSV" data frame

open "each 12 months 2005-2015" data frame #Change it for every month

<assume both data frame have the edges of first row/column be headers of row/column> -> [2,2]

of original == [1,1] of df


lettercurrent = each 12 months 2005-2015[1,2]


#for (i=1; i <=nrow(each 12 months 2005-2015); i++) -not need it!!!

{

 if (number of data entries / number of days of each month = positive integer)

  {

   for (i=1; i <=positive integer; i++)

   {


    for(j=1; j <= nrow(Catalog for hydro ALL CSV); j++)

    {


     if (lettercurrent == Catalog for hydro ALL CSV[j,1])

     {

      for(k=1; k<= number of days of each month; k++)

      {
```

```
          Catalog for hydro ALL CSV[j, 8 +(k-1)] = each 12 months 2005-2015[1+(k-1),3]
          }
          *Go straight to line 83!
        }

      }

    lettercurrent = each 12 months 2005-2015[1+ (i * number of days of each month),2]
    }
   write.csv(Catalog for hydro ALL CSV)
  }

  else
  {
  print ("error")
  }
}
```

4.2. Data Cleaning

```
setwd("E:/Extracurricular activities/Research/Task/TASK 11.14")
Allhydro <- read.csv("Catalog for hydro ALL CSV.csv")
setwd("E:/Extracurricular activities/Research/Task/Hydro power monthly 2006-2016
cleaned/2006/cleaned")
Jan05 <- read.csv("real-cleaned Ene06.csv")
```

```
lettercurrent = trimws(as.character(Jan05[1,2]))

posinteger = length(Jan05[,1])%%31

typesdata = 7285/31

if (posinteger == 0)

{

for (i in 1:typesdata)

{

for(j in 1:nrow(Allhydro))

{

if (lettercurrent == as.character(Allhydro[j,1]))

{

for(k in 1:31)

{

Allhydro[j, 8 +(k-1)] = Jan05[k+((i-1) * 31),3]

}

break

# *Go straight to line 60

}

# lettercurrent = Jan05[1+ (i * 31),2]

}

lettercurrent = trimws(as.character(Jan05[1+ (i * 31),2]))

}

# write.csv(Allhydro)
```

```
} else
{
print ("error")
}
View(Allhydro)
write.csv(Allhydro,"cleanedALL.csv")
```


```
write.csv(Allhydro,"cleanedALL.csv")
```

5. References

1. Li, Gang, Chen-Xi Liu, Sheng-Li Liao, and Chun-Tian Cheng. "Applying a Correlation Analysis Method to Long-Term Forecasting of Power Production at Small Hydropower Plants." Applying a Correlation Analysis Method to Long-Term Forecasting of Power Production at Small Hydropower Plants (2015): 1-15. Web. 19 Dec. 2016.

2. Tellez, Alejandra Monroy. "Internship_Report." Internship_Report (2015): 1-17. Web. 19 Dec. 2016.

3. Kao, Shih-Chieh, Michael J. Sale, Moetasim Ashfaq, Rocio Uria Martinez, Dale P. Kaiser, Yaxing Wei, and Noah S. Diffenbaugh. "Energy." Projecting Changes in Annual Hydropower Generation Using Regional Runoff Data: An Assessment of the United States Federal Hydropower Plants (2014): 1-12. Web. 19 Dec. 2016. <http://hydropower.ornl.gov/docs/pubs/Projecting_Changes_In_Annual_Hydropower_Generation_Using_Regional_Runoff_Data_An_Assessment_Of%20_The_United_States_Federal_Hydropower_Plants.pdf>.

4. M, Suleiman Y., Ifabiyi I. P, and Department of Geography and Environmental Management, University of Ilorin, Nigeria. "Momona Ethiopian Journal of Science (MEJS), V7(1):55-63, 2015 ©CNCS, Mekelle University, ISSN:2220-184X The Role of Rainfall Variability in Reservoir Storage Management at Shiroro Hydropower Dam, Nigeria." Ed. Department of Geography, Federal University of Technology, Minna, Nigeria. Momona Ethiopian Journal of Science (MEJS), V7(1):55-63, 2015 ©CNCS, Mekelle University, ISSN:2220-184X The

Role of Rainfall Variability in Reservoir Storage Management at Shiroro Hydropower Dam, Nigeria (2015): 1-9. Web. 19 Dec. 2016.