

KyungHee Kim

kkh97122647@gmail.com

Kyunghee Kim
김경희

Data Scientist

김경희 Kyunghee Kim

010.4127.4281
kkh97122647@gmail.com

Experience

- 2021 Joined Research Institute of Applied statistics
- 2021 Corporate Project Effectiveness Analysis Statistical Advisory (ArtsCompany)
- 2022 Published 'Time Series Outlier Detection and Clustering of Seoul citizen card User Behaviours
- 2022 Developed Outlier detection for multivariate Long Memory Processes
- 2022 Joined Drone Multivariate Data-Based Forecasting and Anomaly Detection Research (National Institute of Security Technology)

Awards

- 2017 Smart Life ICT Convergence APP contest (Seoul, Korea)
A Proposal for Sharing the Surplus Space of Universities
- 2020 The science and technology academic seminar, Academic Award
Analysis of movie box office factors and suggestion of utilization
- 2021 Weather bigdata contest: Develop a model to predict the occurrence of landslide by machine learning model
Minister of Environment Award
[Main work: Visualization, preporcessing, Drive insight from data]
- 2022 Research Matters 연구성과 경진대회 시계열 이상탐지
모델 관련 사회과학분야 우수상 (성균관대학교)

CONTENTS

1

Project1

Drone Sensor data Visualization and Anomaly detection

2

Project2

Time Series Outlier Detection and Clustering of Seoul Citizen Card User Behaviours

3

Project3

2021 Weather Big data Contest

Project1

Drone Sensor data Visualization and Anomaly detection

Keyword	Drone, Sensor Data, Time Series, Visualization, Anomaly Detection
Tool	R, Python
OS	Window, Linux
Role	Simulator-based data generation Visualize and analyze data and apply statistical anomaly detection models



Project1

National Institute of Security Technology Consignment Tasks [EDA]

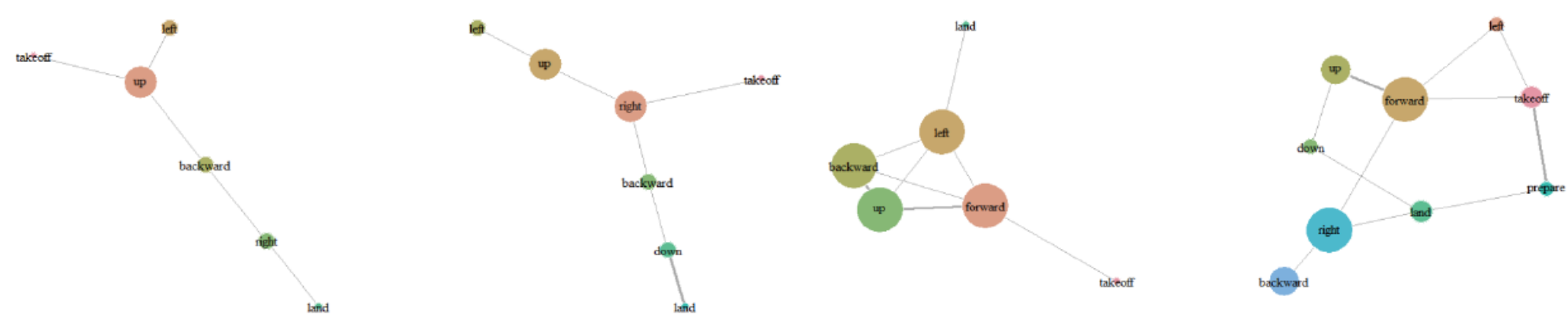


Fig1. network graph (Flight Behaviour)

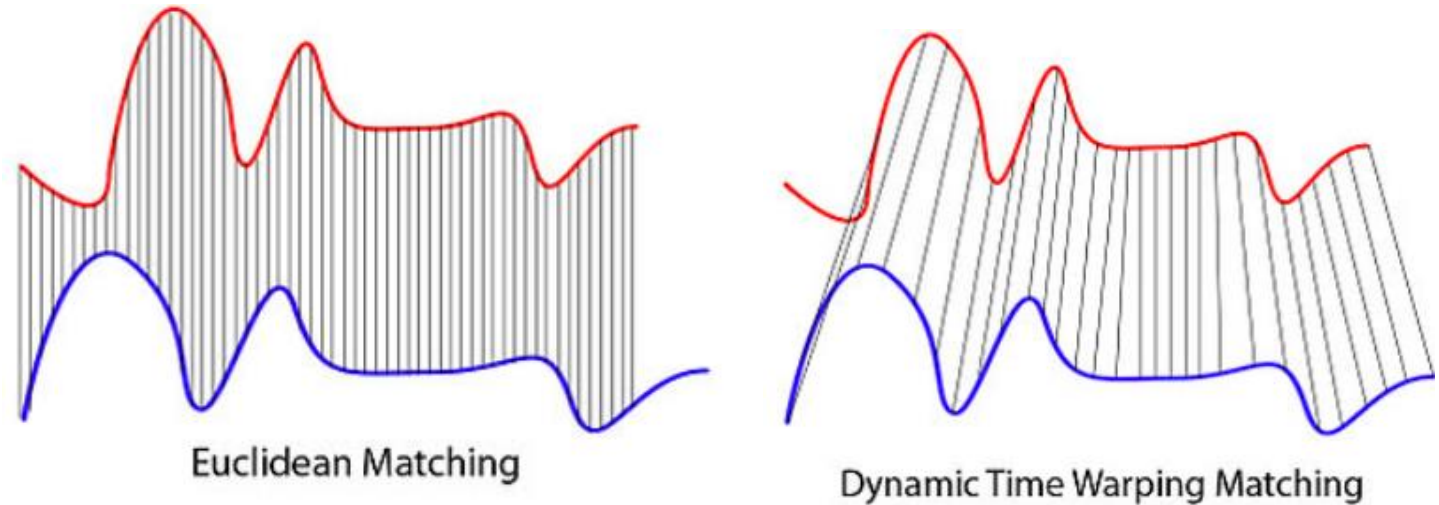


Fig2. Subsystem clustering based on DTW clustering

Project1

National Institute of Security Technology Consignment Tasks [Anomaly Detection]

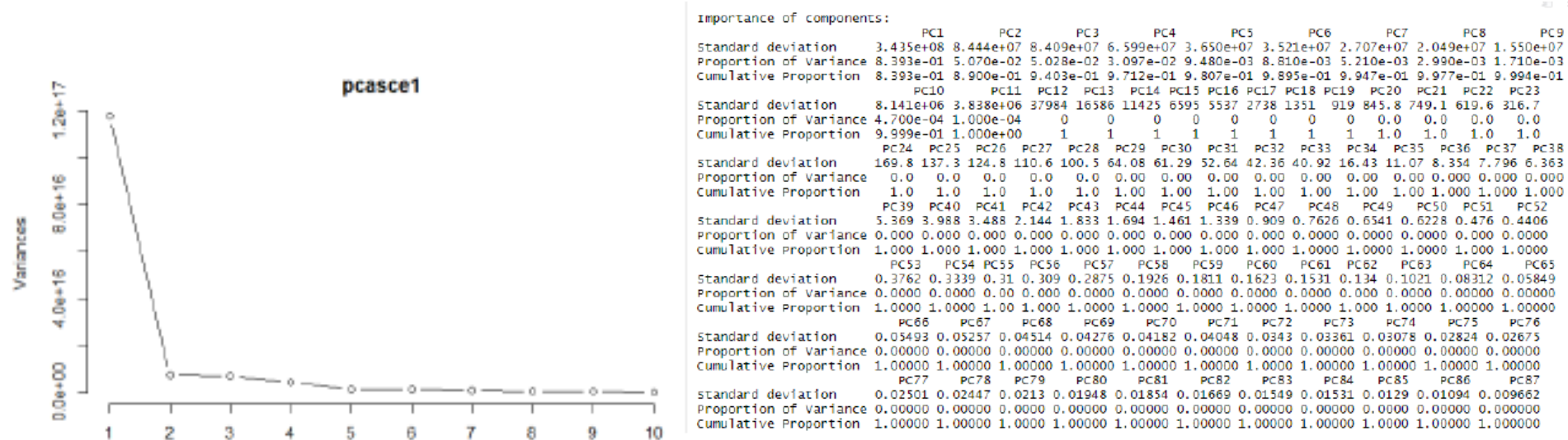


Fig3. PCA Result

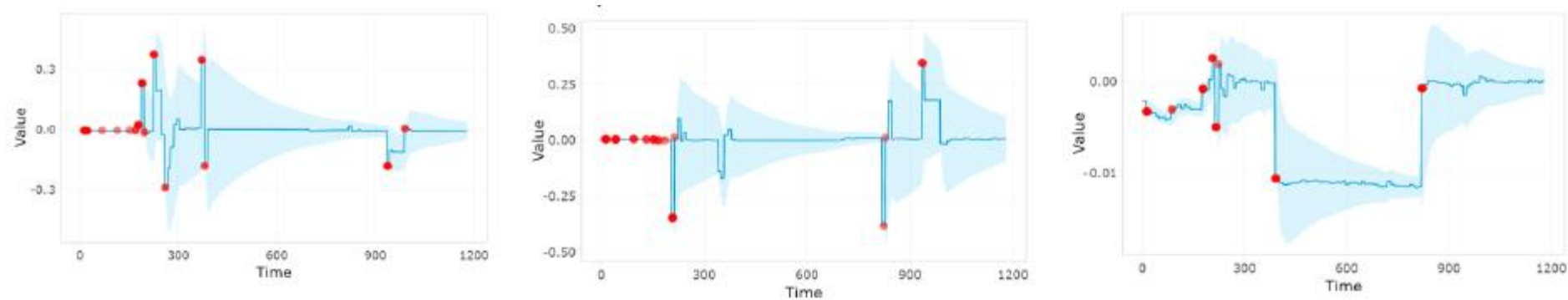


Fig4. Anomaly Detection Result

Time Series Outlier Detection and Clustering of Seoul Citizen Card User Behaviours

Keyword	Seoul citizen Card, ARMA model, Outlier Detection, DTW clustering
Tool	R
Author	1 st Author
Reference	김경희. "서울시민카드 사용자 유입과 이탈에 대한 시계열 이상치 탐지와 DTW 클러스터링." <i>통계연구</i> 22 (2022): 63-76.



Project2

Goal

It is intended to help achieve the goals of the Seoul Citizens Card public app by using it to increase the continuous inflow and frequency of use of Seoul Citizens Card users

* Public App Goals: Making a wide range of public services available to citizens conveniently

Data

Daily time series Seoul Citizen Card Usage Information Data

How to?

The outlier detection technique identifies when there is a rapid change in the frequency of app access by gender/age, and DTW clustering is used to cluster users with similar time series patterns and then to find out cluster characteristics

Result

Short-term events result in an immediate influx of users, but it was a one-off.

In addition, groups responding to short-term events easily leave after the event ends

Mid- to long-term events (Seoul Tourism Discount Pass) continue to lead to inflow

Project2

Outlier detection:
By using auto.arima model

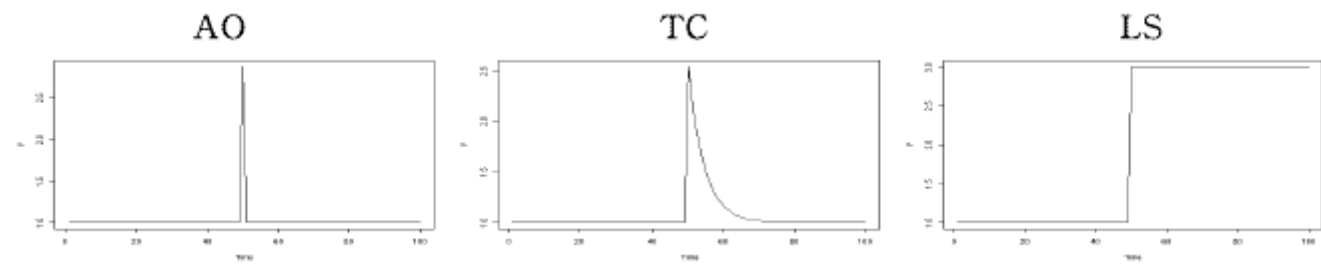


Fig1. Define Outlier

Time	20남	20여	30남	30여	40남	40여	50남	50여
4					AO			
5	AO	AO						AO
6			AO		AO		AO	
10	AO	AO	AO	AO	AO			
15					AO			
19							AO	
20			TC		TC	TC		TC
29			AO		AO	AO	AO	AO
30	LS							
34			TC	TC	TC	TC	TC	TC
39			LS					
44			LS		LS			
54		LS		LS		LS	LS	
91	TC	TC	AO	AO	TC	TC	TC	TC

Fig2. Outlier Detection Result table

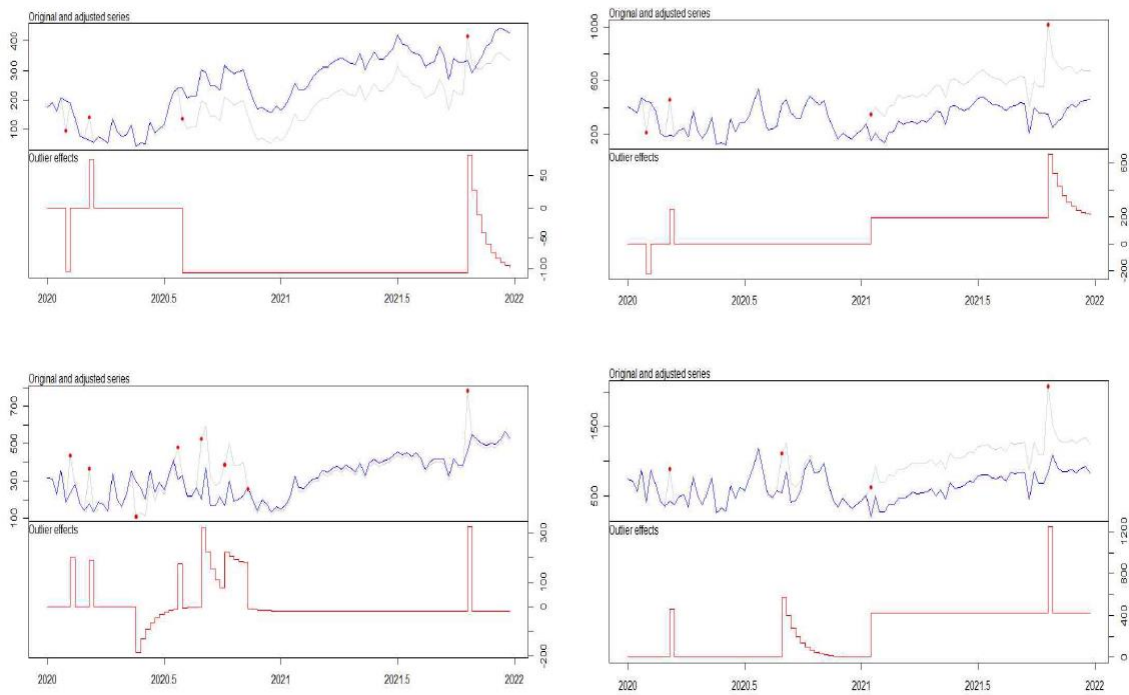


Fig3. Outlier Detection Visualization

Project2

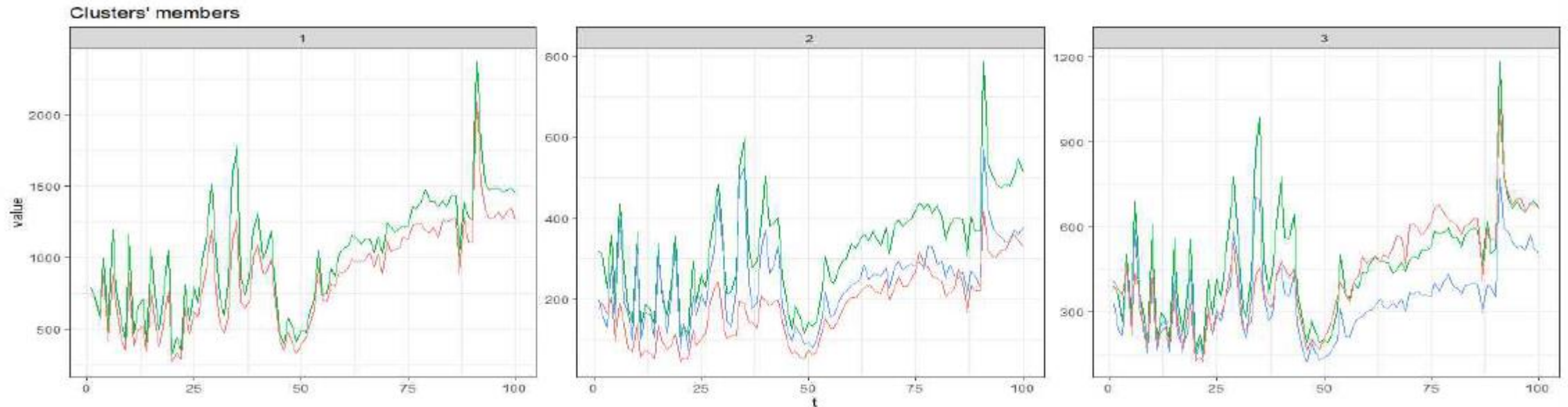


Fig4. Time plot for each Clusters

Cluster 1: 30s female, 40s female

Cluster 2: 20s male, 30s male, 50s male

Cluster 3: 20s female, 40s male, 50s female

Consider cluster result together
with outlier detection result



This cluster has steady access.

There is a chance to leave

Clusters responding to events or updates.

Project3

Keyword

landslide, weather data(Structured), Map data (Unstructured)

Tool

R, Python, SQL

Role

Data preprocessing and visulalization, Derive variables, Derive insight from data.

2021 Weather Bigdata contest



Project3

Improve predictive performance through data preprocessing, visualization, and variable generation(90%→95%)

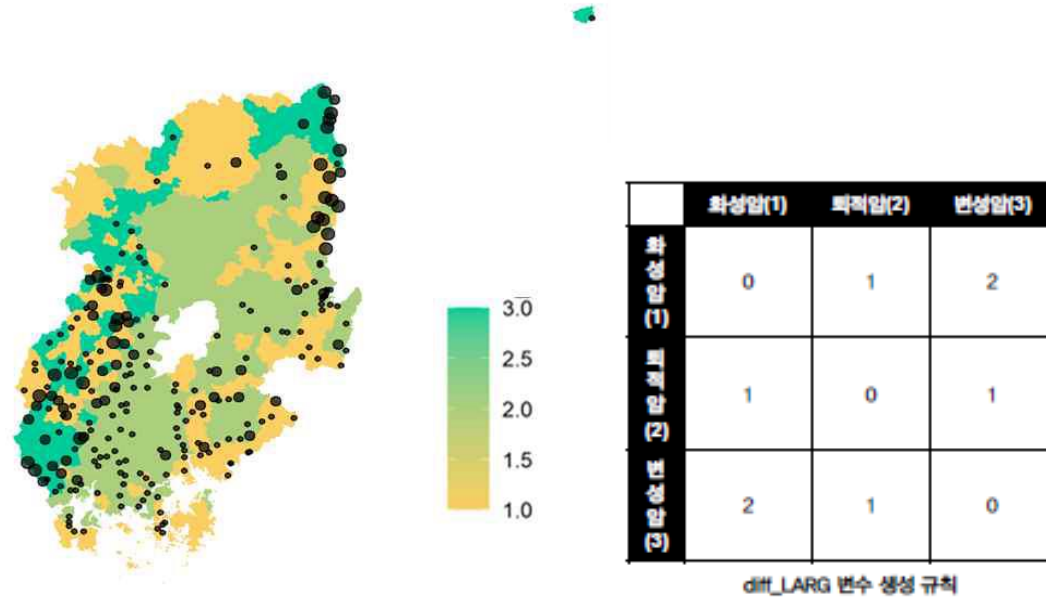


Fig5. Make new diff_LARG

I matched PRRCK_LARG code variable to map,
There were a lot of occurrence of landslide at
boundaries especially when the code changed
Therefore, I proposed a new variable as following fig5.

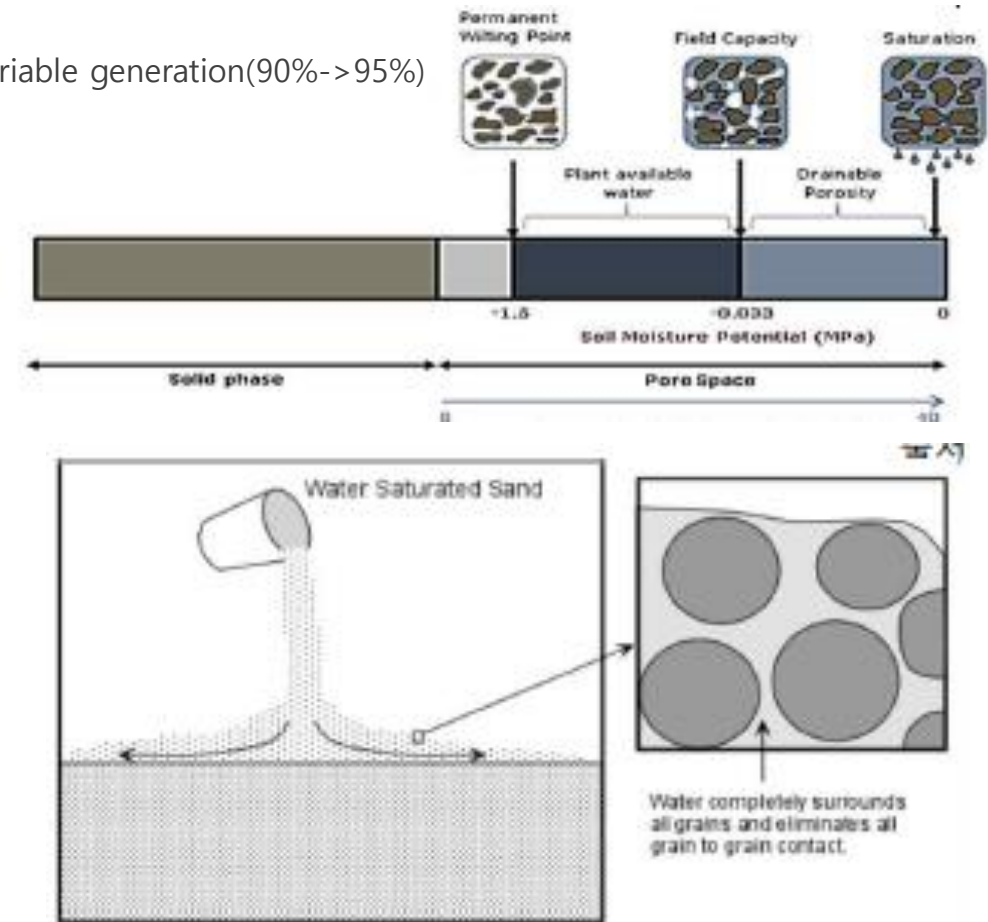


Fig6. Particle structure according to wet condition of soil

Merging additional data, which can be
affected to landslide.

Data Scientist

김경희 Kyunghee Kim

Tools

R	Can Reproduce a paper
Python	Visualization, preprocessing and do modeling
SQL	Data extraction from DB, can do some basic queries.
Cooperation	
tool(Notion)	Utilized for team projects and study arrangements

Teaching

Basis and application of AI (SKKU)

Teaching Machine and Deep learning using Python (linear regression, RNN, CNN, GNN)

R seminar (SKKU)

Advanced R

Language

English OPIC – IH (220917)

Certification

Data Analysis Semi- Specialist (ADsP) 0150874

THANK YOU
