Temporal Spatial FDA

with Korean Weather Data

202STG18 이재빈
202STG19 이채영

CONTENTS

1. INTRO

1) 분석 목적

2) EDA

2. PROCESS

1) Temporal
Spatial
FDA

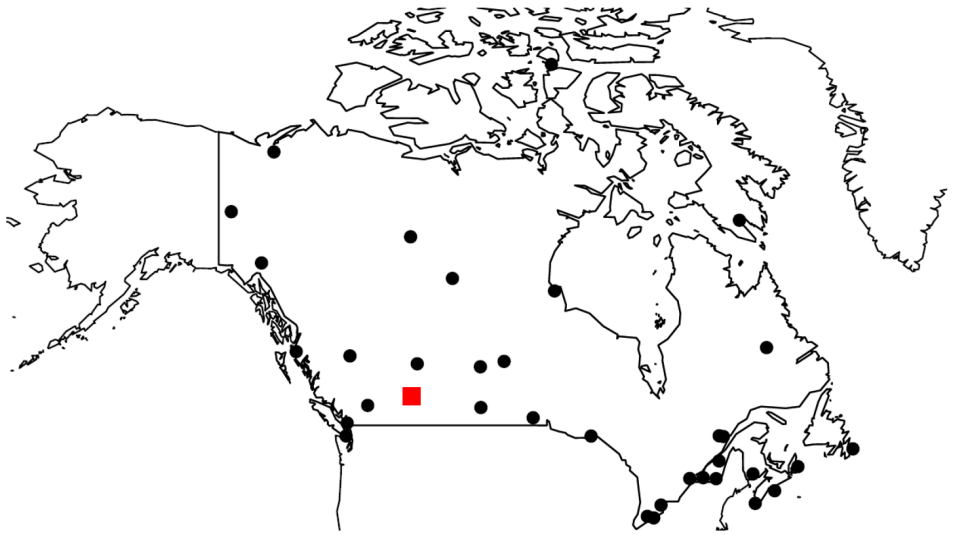
2) Multivariate
FDA

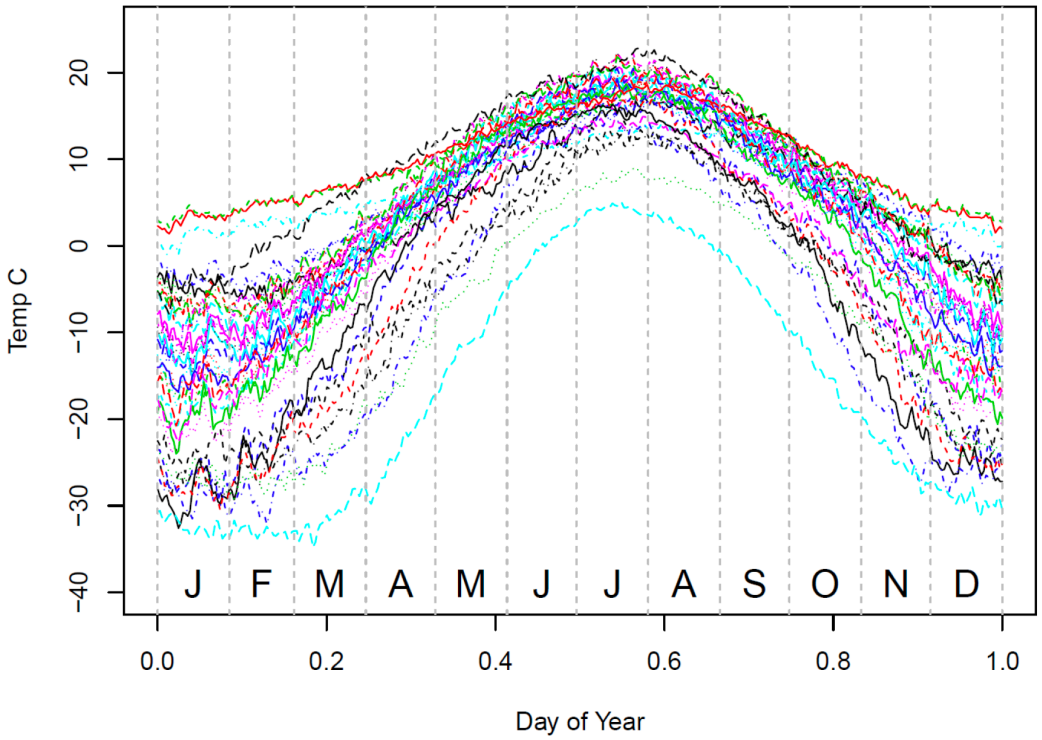
3. CONCLUSION

1 INTRO

Canadian Weather Data

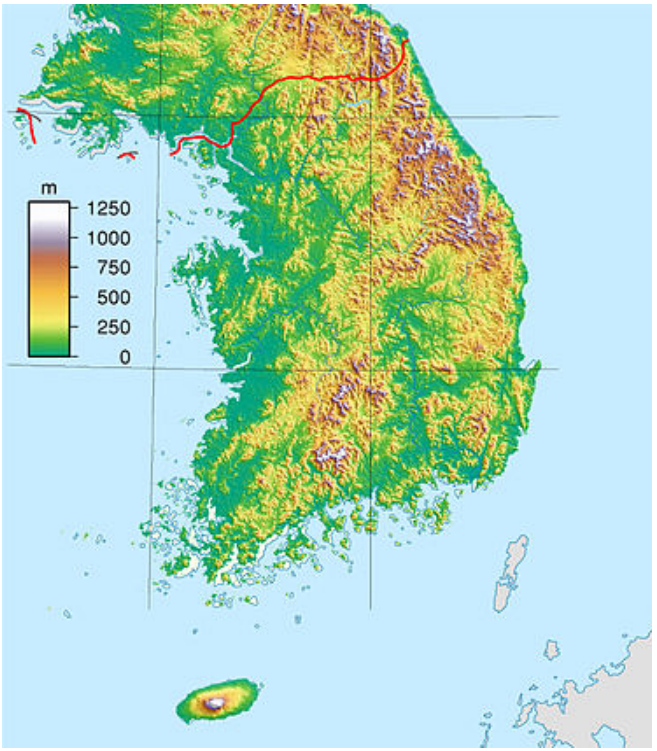
- 관측소가 Sparse 하게 위치해 있으며, 아래에 몰려 있는 양상
- 연교차가 30도에 달하는 극한 기후

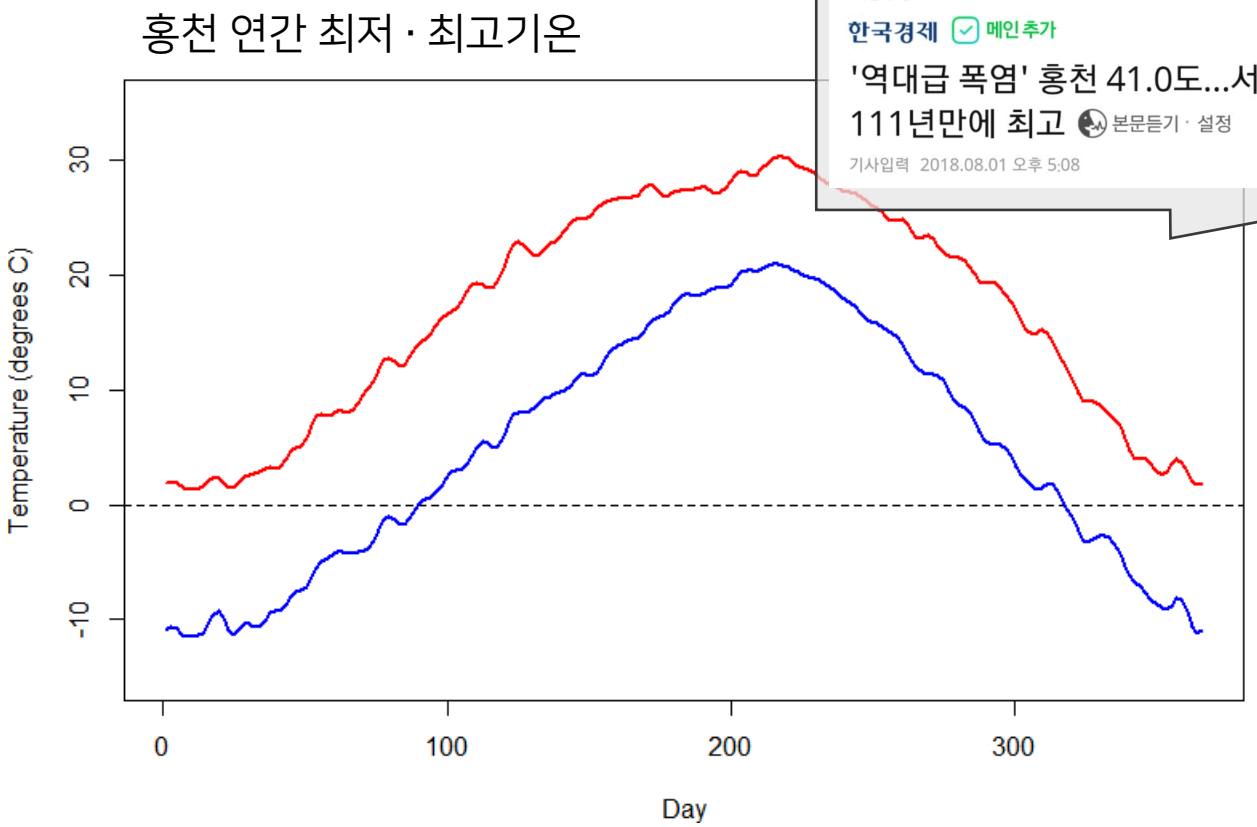




한국 기온 데이터

- 국토의 70% 이상이 산으로 이루어져 있는 특이한 지형
- 연교차가 큰 특성을 보임





news1 KOREA 메인 추가
[날씨]강원 전역 한파 이어져...홍천 영하 27.1 본문듣기 · 설정
기사입력 2018.01.25 오전 7:42
최종수정 2018.01.25 오전 7:43

한국경제 메인 추가
'역대급 폭염' 홍천 41.0도...서울 39.6도 111년만에 최고 본문듣기 · 설정
기사입력 2018.08.01 오후 5:08

데이터 소개

- [기상청] 1991년 ~ 2020년, 최근 30년 간의 기온 기후평년값
- 최저기온 & 최고기온 데이터 사용

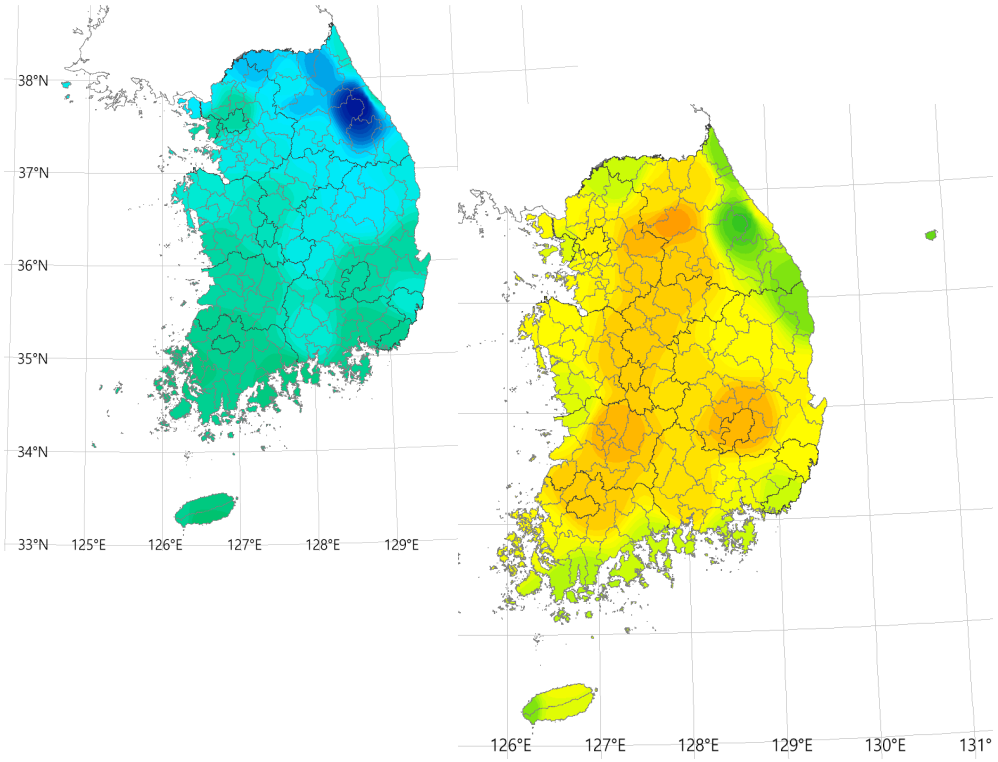
	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	date	가평조종	간성	간절곶	강남	강동	강릉	강북	강서	강진군	강화	거제	거창	경기광주
2	2020-01-01	0.7	4.3	8.3	2.4	2.1	6.2	2	2	6.3	2	8	6.1	2
3	2020-01-02	0.8	4.2	8.3	2.4	2.2	6	2.2	2	6.3	2	8	6.1	2.1
4	2020-01-03	1	4.3	8.4	2.6	2.4	5.9	2.4	2.2	6.6	2	8	6.2	2.3
5	2020-01-04	1.1	4.3	8.4	2.6	2.3	5.9	2.6	2.3	6.7	1.9	8	6.2	2.3
6	2020-01-05	1	4.3	8.2	2.5	2.2	5.8	2.6	2.1	6.5	1.8	7.9	5.9	2.1
7	2020-01-06	0.9	4	8.1	2.4	2	5.5	2.5	2	6.4	1.7	7.7	5.7	1.9
8	2020-01-07	0.6	4	7.8	2.1	1.8	5.5	2.2	1.9	6.2	1.6	7.5	5.4	1.6
9	2020-01-08	0.4	4	7.7	1.9	1.7	5.7	2.1	1.7	5.9	1.6	7.4	5.4	1.5
10	2020-01-09	0.4	3.8	7.5	1.8	1.7	5.5	2	1.5	5.6	1.5	7.3	5.3	1.4
11	2020-01-10	0.3	3.6	7.4	1.8	1.8	5.4	1.8	1.5	5.2	1.4	7.1	5.3	1.4
12	2020-01-11	0.2	3.5	7.4	1.6	1.7	5.2	1.6	1.3	4.9	1.2	7	5	1.3
13	2020-01-12	0.3	3.4	7.5	1.8	1.9	5	1.9	1.4	5	1.2	7	5.1	1.5
14	2020-01-13	0.5	3.4	7.6	1.9	2	4.9	2.2	1.5	5.4	1.3	7	5.2	1.7
15	2020-01-14	0.7	3.5	7.6	2.1	2.3	5	2.5	1.7	5.5	1.5	7	5.2	1.8
16	2020-01-15	1	3.8	7.8	2.4	2.5	5.1	2.8	2.1	5.8	1.6	7.1	5.4	2.1
17	2020-01-16	1.3	4.1	8.1	2.8	2.9	5.3	3.2	2.5	6.2	1.9	7.3	5.7	2.5
18	2020-01-17	1.6	4.4	8.2	3.1	3.1	5.4	3.5	2.8	6.6	2	7.4	5.8	2.8
19	2020-01-18	1.7	4.3	8.2	3.3	3.2	5.4	3.6	3	6.7	2.1	7.3	5.7	2.9
20	2020-01-19	1.6	4	8.1	3.2	3	5.2	3.5	2.9	6.8	2	7.3	5.5	2.8
21	2020-01-20	1.3	3.7	7.8	3	2.9	5	3.2	2.8	6.6	1.9	7.1	5.2	2.6
22	2020-01-21	1.1	3.5	7.5	2.7	2.6	4.8	2.8	2.5	6.2	1.7	6.8	4.9	2.4
23	2020-01-22	0.7	3.1	7.1	2.3	2.2	4.6	2.2	2	5.8	1.4	6.6	4.5	1.9

지역별 30년 연간 최저 · 최고기온 평균

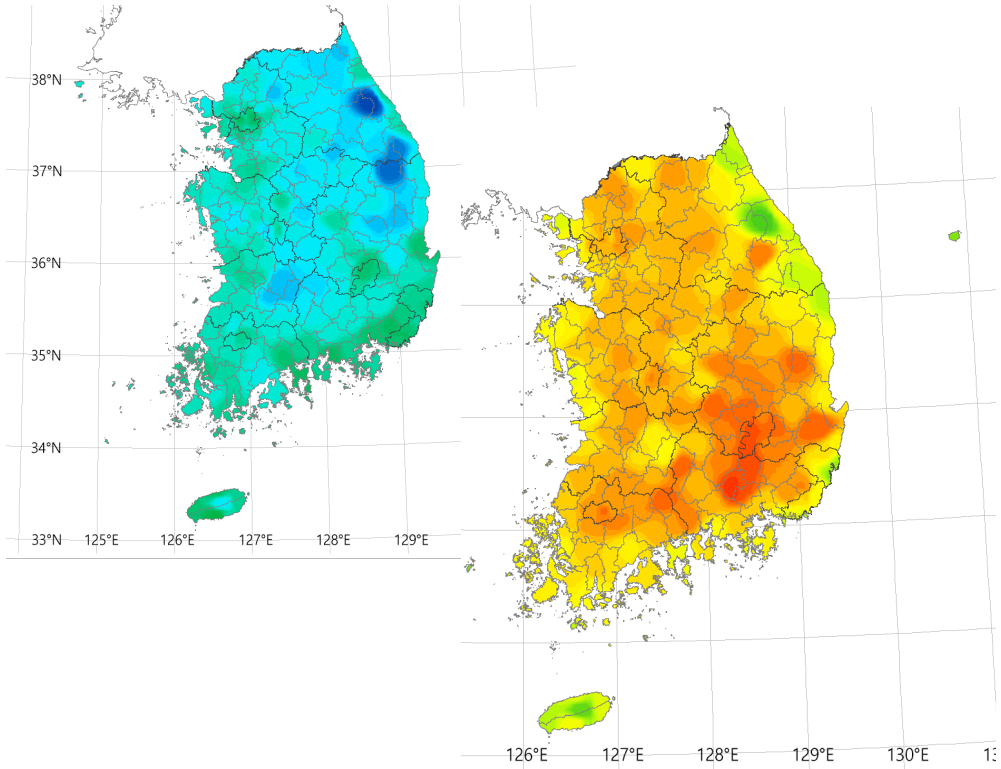
기상 관측소 (206개)

1. 최근 30년 기온 데이터 사용

- 지구 온난화 영향으로 온도가 높아지는 경향을 보여, stationary 가정을 위해 최근의 데이터 사용



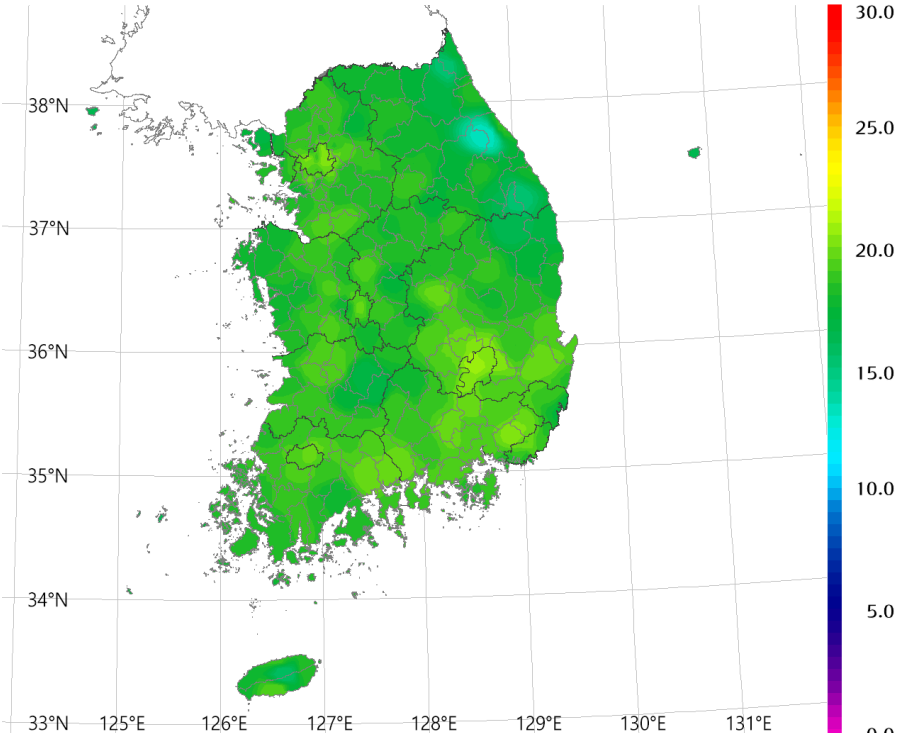
1961년 ~ 1990년 최저 · 최고기온



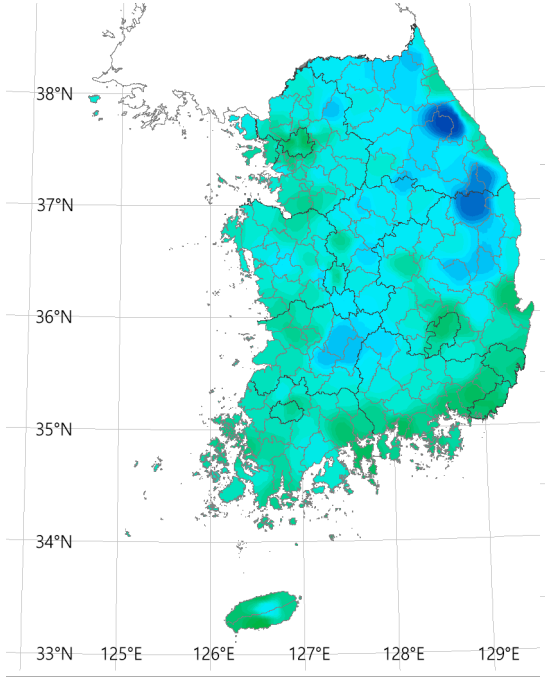
1991년 ~ 2020년 최저 · 최고기온

2. 최저/최고기온 데이터 선택

- 지역별 특성을 상세하게 반영하기 위해, 최저 · 최고기온 데이터 사용

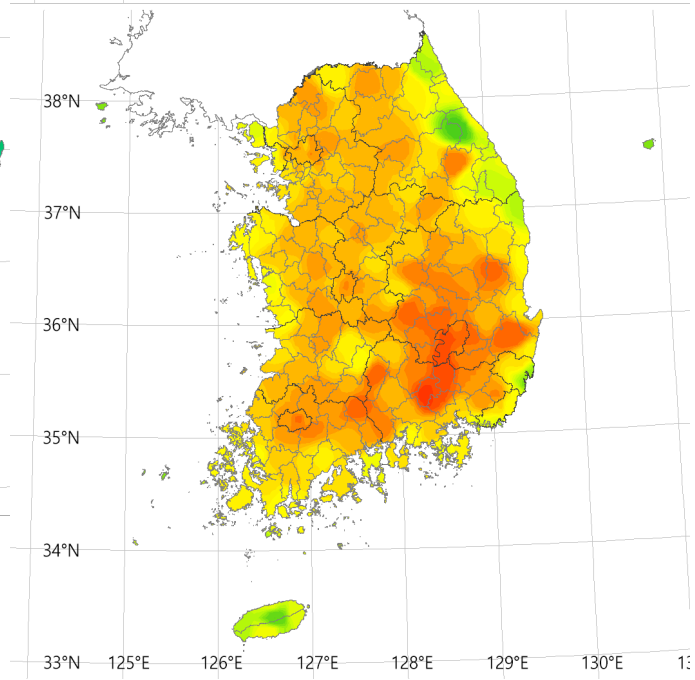


평균기온



▲ 최저기온

▼ 최고기온



3. 도서지역 관측소 제외

- 내륙 관측소에 비해 거리 차이가 많이 나기에, 도서지역 관측소 제외

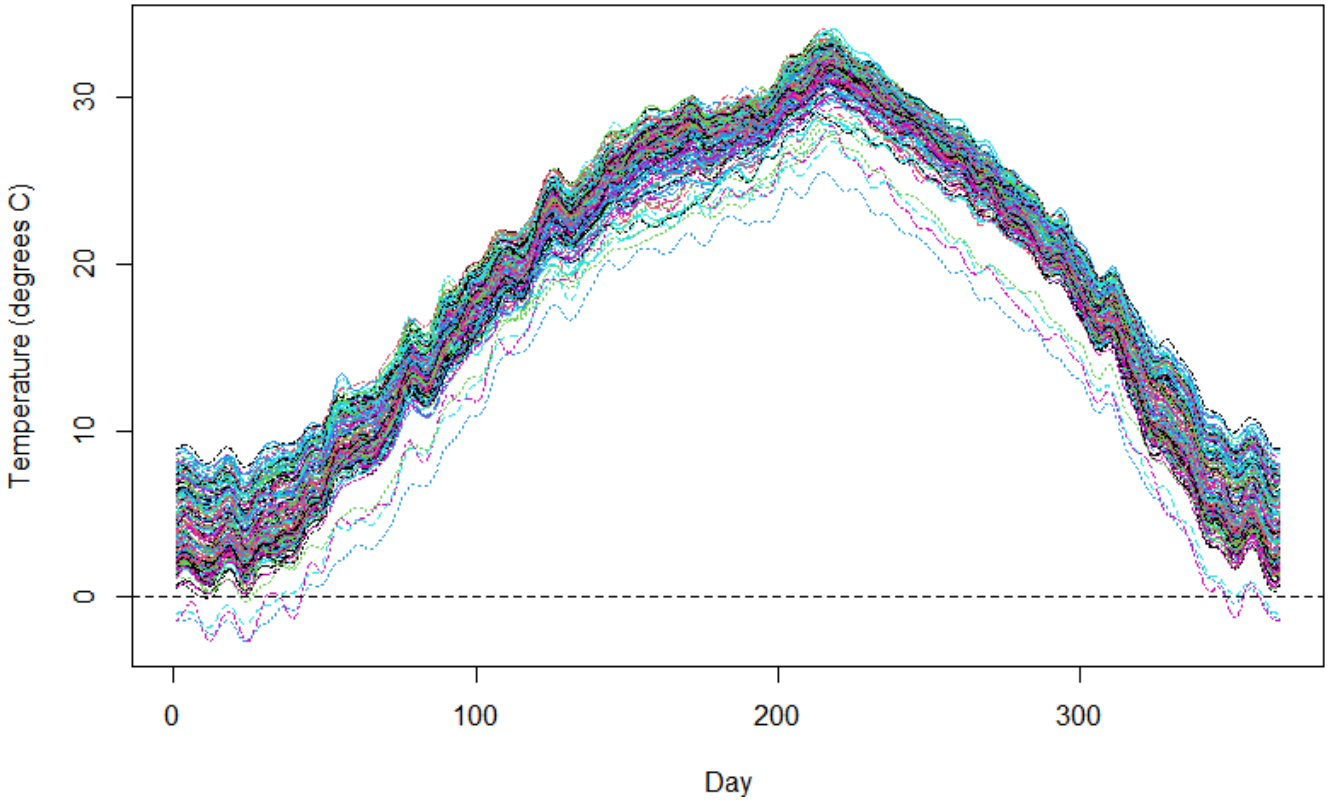
전국 관측소 (219개)

제외 후 관측소 (206개)

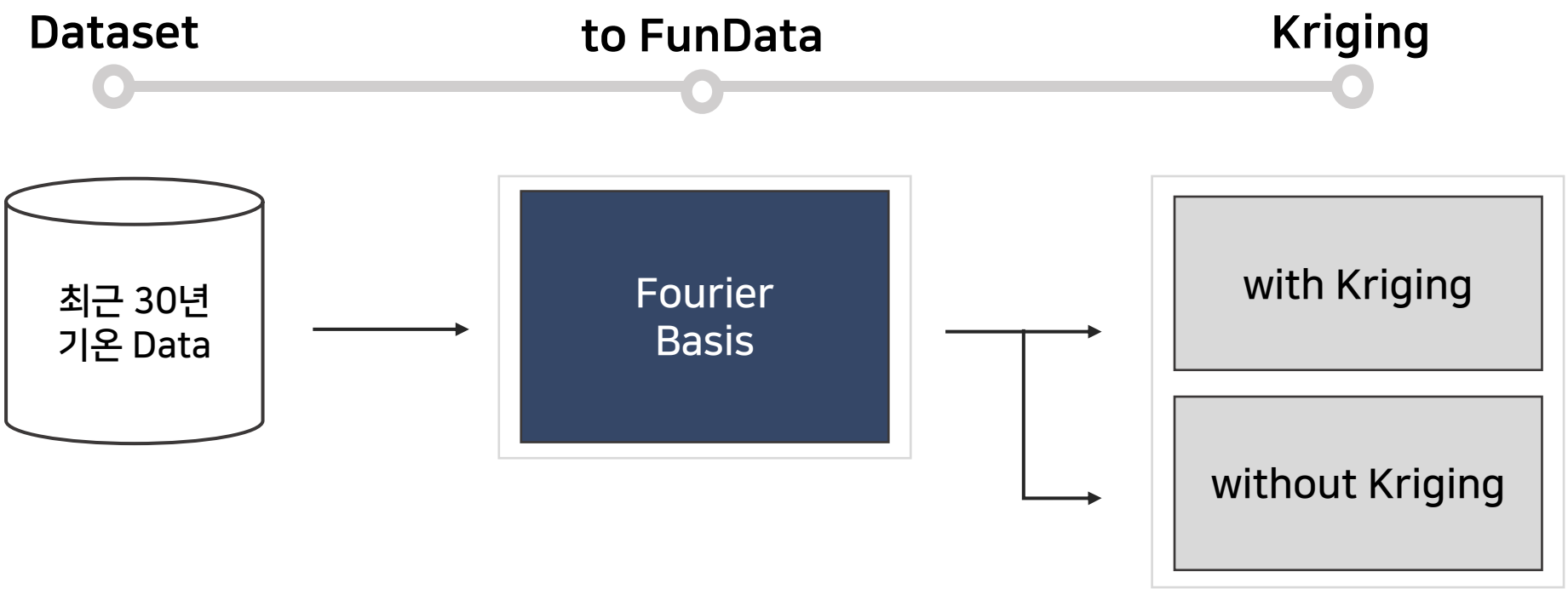
2 PROCESS

Temporal Spatial FDA

- 365일의 기온 패턴 & 관측소 간의 거리를 고려하여, 특정 지역의 기온 추세 예측

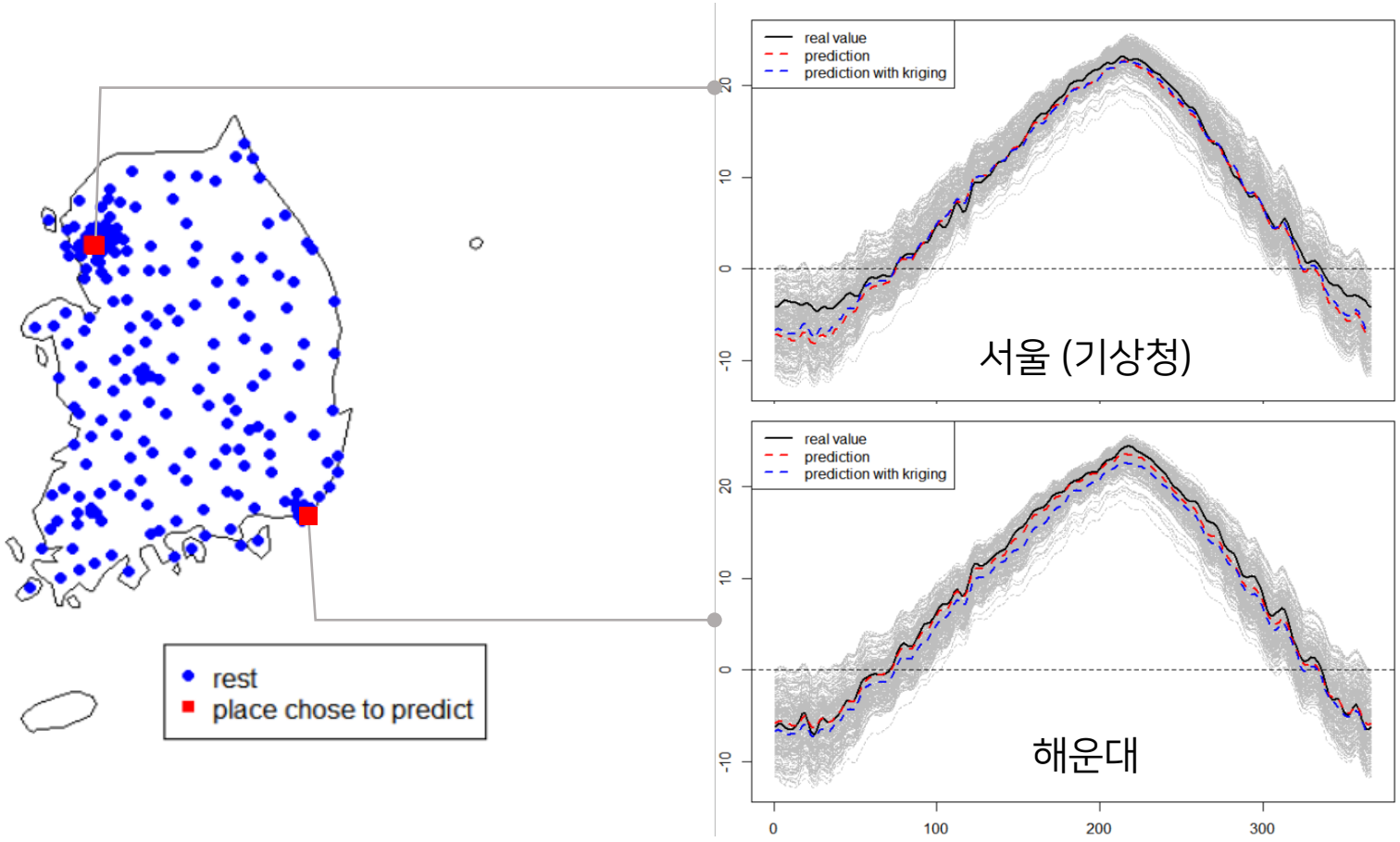


MODEL FLOW


Dataset
to FunData
Kriging
최근 30년
기온 Data
Fourier
Basis
with Kriging
without Kriging

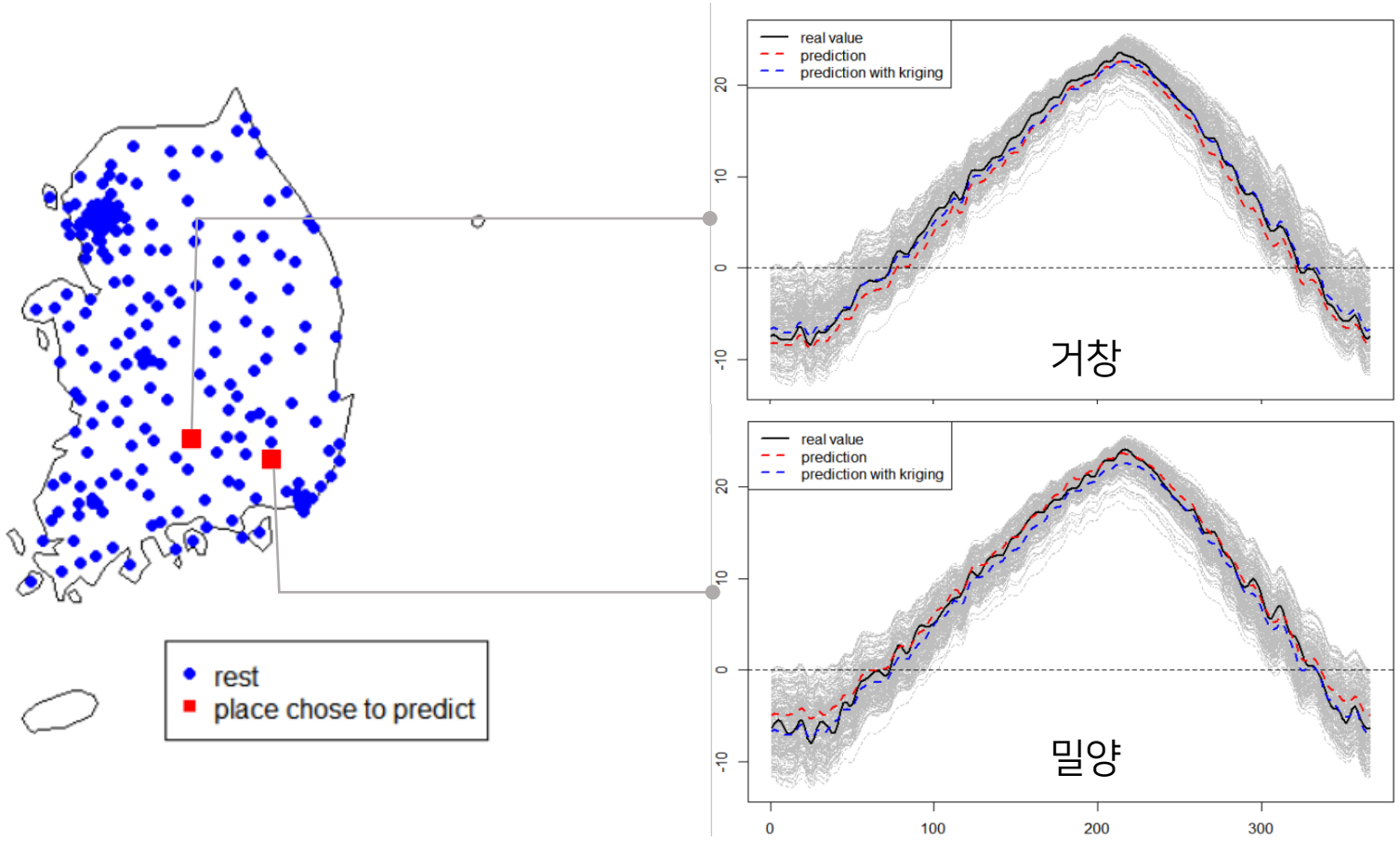
1. 예측이 잘 된 지역

- [서울 (기상청), 해운대] 주변에 관측소가 많은 지역



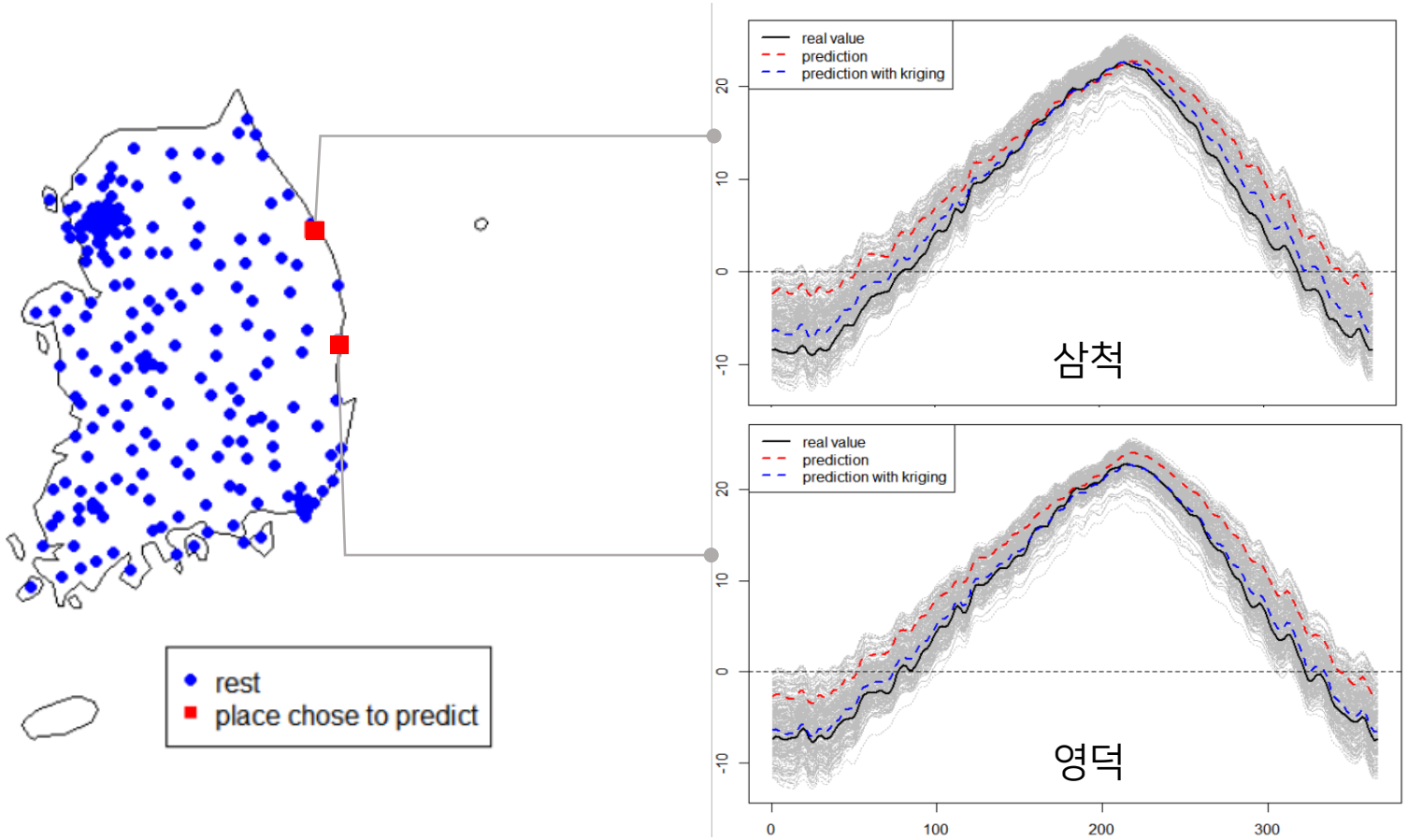
1. 예측이 잘 된 지역

- [거창, 밀양] 내륙 지방



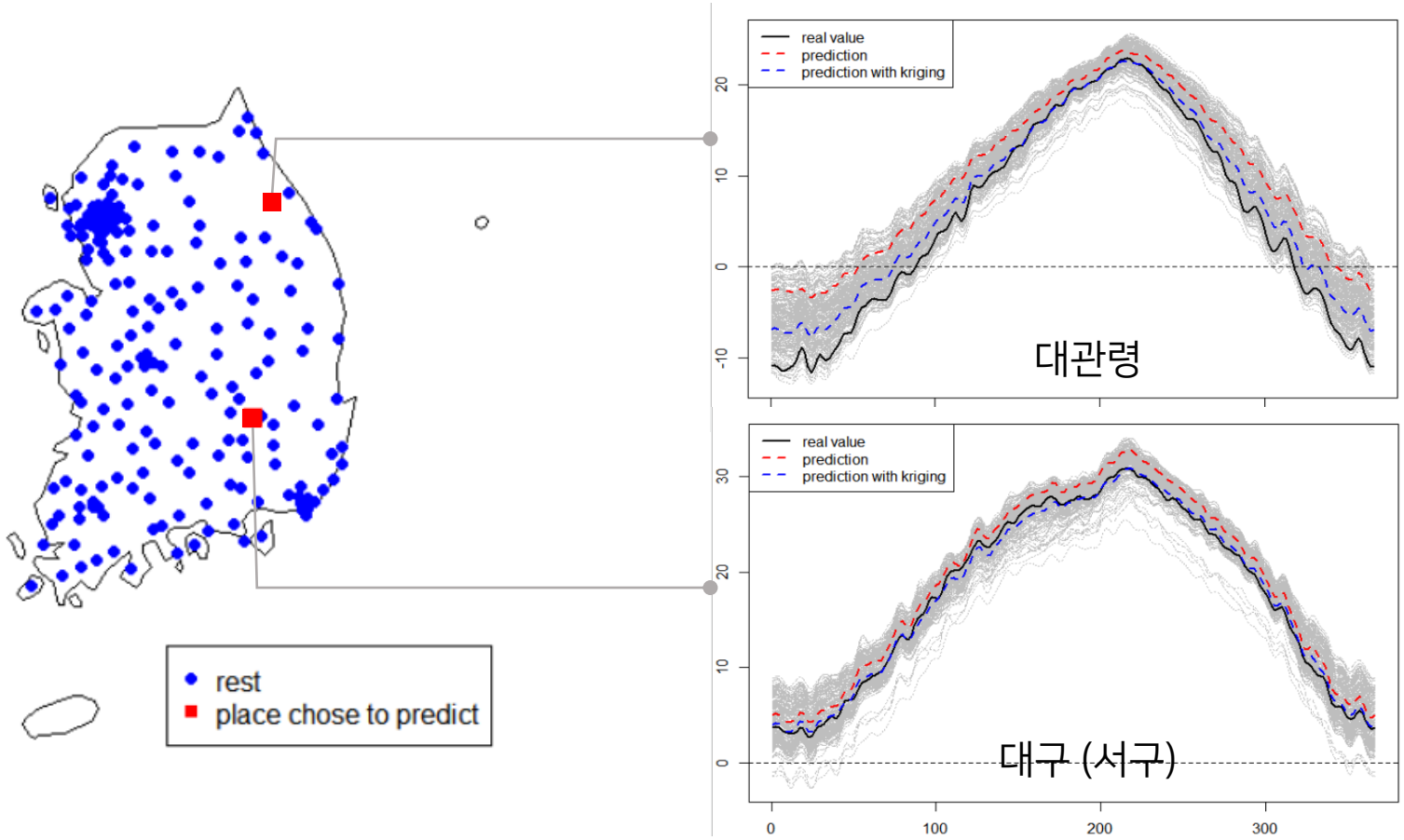
1. 예측이 잘 된 지역

- Kriging 시 예측력이 좋아지는 지역
- [삼척, 영덕]



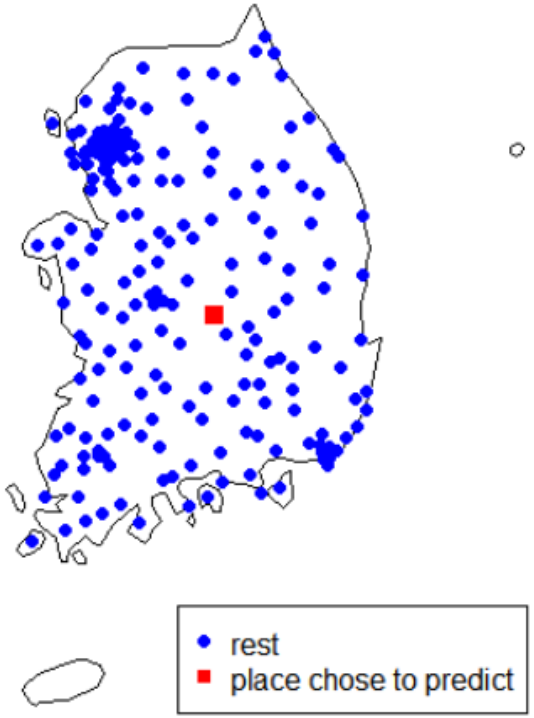
1. 예측이 잘 된 지역

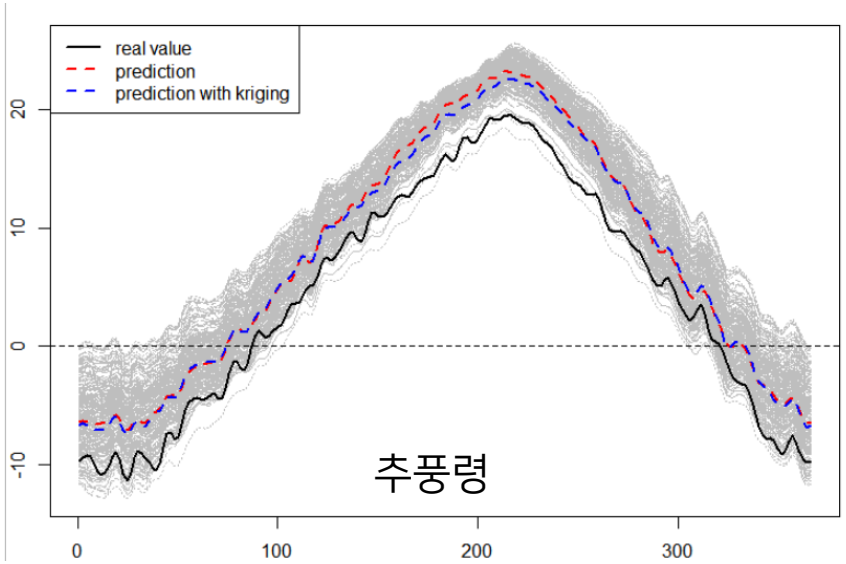
- Kriging 시 예측력이 좋아지는 지역
- [대관령, 대구(서구)]



2. 예측이 어려운 지역

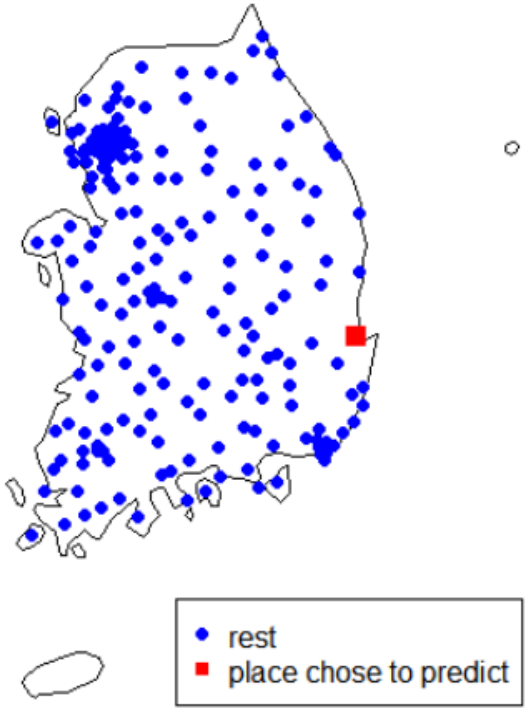
(1) [추풍령] 산맥

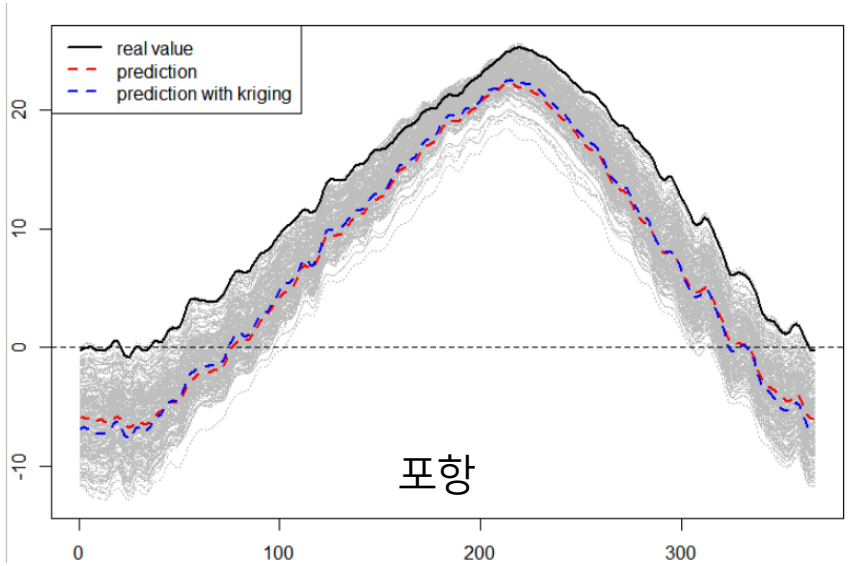




2. 예측이 어려운 지역

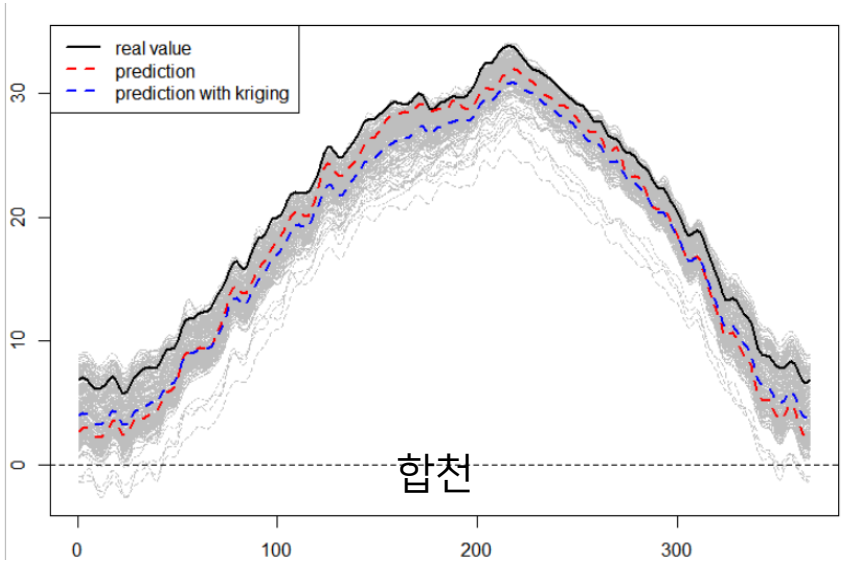
(2) [포항] 해안가





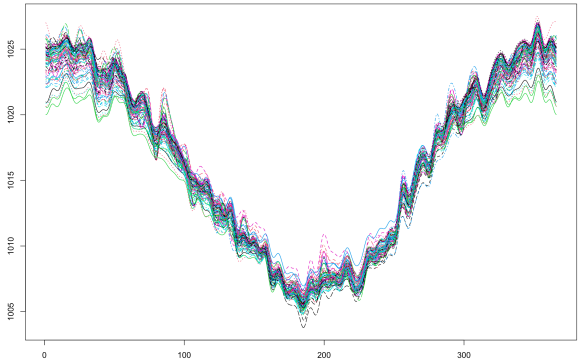
2. 예측이 어려운 지역

(3) [합천] Kriging 시 예측력이 떨어지는 지역

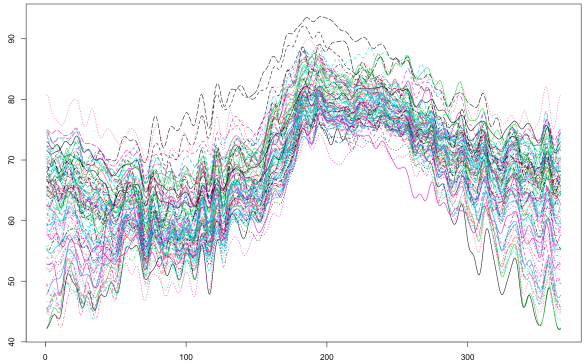


기상 상황 반영

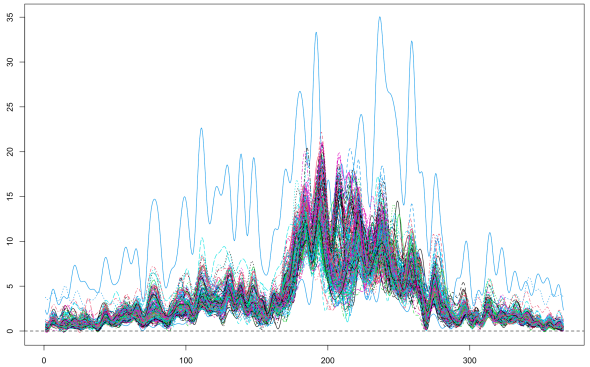
- 예측이 어려웠던 지역에 대해, Clustering 하여 Temporal Spatial FDA 진행하고자 함
- 기온 외 다양한 기상 요소를 반영하여, Multivariate FDA



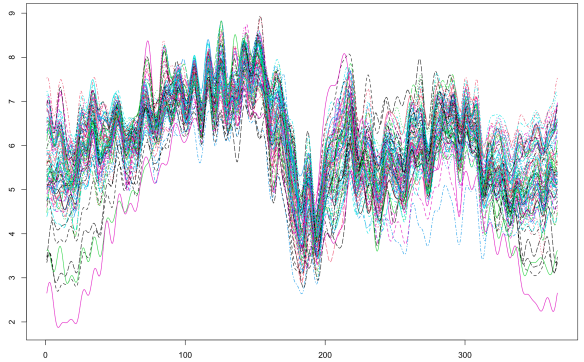
기압



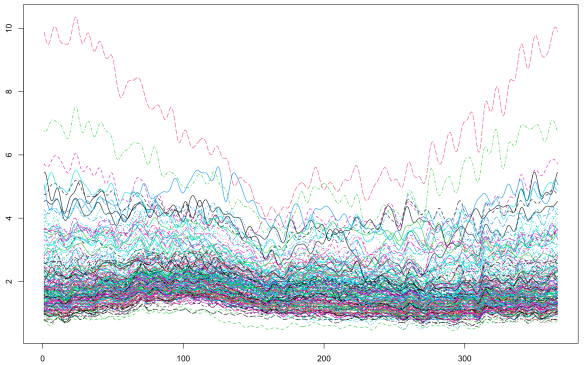
상대습도



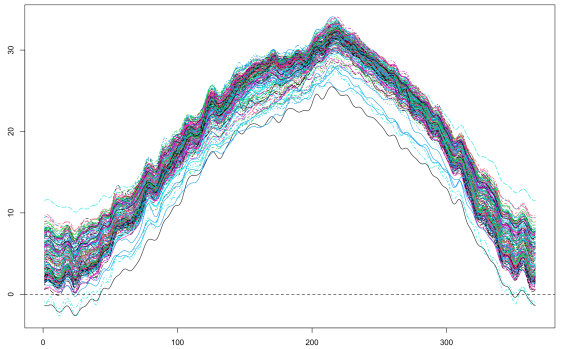
강수량



일조시간



풍속



기온

Multivariate FDA

- [Trial] 다양한 변수들 사이의 상관관계를 반영하고자 함

1

Package ‘funHDDC’

March 17, 2021

Type Package
Title Univariate and Multivariate Model-Based Clustering in Group-Specific Functional Subspaces
Version 2.3.1
Date 2021-03-17
Author A Schmutz, J. Jacques & C. Bouveyron
Maintainer Charles Bouveyron <charles.bouveyron@gmail.com>
Depends MASS,fda, R (>= 3.1.0)
Description The funHDDC algorithm allows to cluster functional univariate (Bouveyron and Jacques, 2011, <doi:10.1007/s11634-011-0095-6>) or multivariate data (Schmutz et al., 2018) by modeling each group within a specific functional subspace.
License GPL-2
Encoding UTF-8
LazyData false
RoxygenNote 6.0.1
Suggests knitr, rmarkdown
VignetteBuilder knitr
NeedsCompilation no
Repository CRAN
Date/Publication 2021-03-17 10:40:02 UTC

- Multivariate Functional Data **Clustering**
- K-means 기반, EM-algorithm 최적화
- [ISSUE] Functional Data 최대 2개 반영

2

Package ‘MFPCA’

September 4, 2020

Type Package
Title Multivariate Functional Principal Component Analysis for Data Observed on Different Dimensional Domains
Version 1.3-6
Date 2020-08-31
Maintainer Clara Happ-Kurz <clara.happ@stat.uni-muenchen.de>
Description Calculate a multivariate functional principal component analysis for data observed on different dimensional domains. The estimation algorithm relies on univariate basis expansions for each element of the multivariate functional data (Happ & Greven, 2018) <doi:10.1080/01621459.2016.1273115>. Multivariate and univariate functional data objects are represented by S4 classes for this type of data implemented in the package 'funData'. For more details on the general concepts of both packages and a case study, see Happ-Kurz (2020) <doi:10.18637/jss.v093.i05>.
URL https://github.com/ClaraHapp/MFPCA
License GPL-2
Imports abind, foreach, irlba, Matrix, methods, mgcv (>= 1.8-33), plyr, stats
Depends R (>= 3.2.0), funData (>= 1.3-4)

- Multivariate Functional Data **PCA**
- Different Domain 에서 PC 찾아 차원 축소 $\hat{x}_i = \sum_{m=1}^{M} \hat{\rho}_{im} \hat{\psi}_m$
- [ISSUE] 각 Functional Data 에서 차원 축소

Multivariate Clustering

- [Trial] 모든 조합에 대해 Multivariate Clustering 을 진행한 후, majority Voting

3

```
> res.multi1<-funHDDC(list(fd.temp1991_high1.1, fd.pressure1991),
+                     K=6,init="kmeans",threshold=0.2)
      model K threshold complexity        BIC
1 AKJBKQKDK 6       0.2      3,764 -45,097.00

SELECTED: model  AKJBKQKDK  with  6  clusters.
Selection Criterion: BIC.
```

기온 ~ 기압

```
> res.multi2<-funHDDC(list(fd.temp1991_high1.1, fd.humidity1991),
+                     K=6,init="kmeans",threshold=0.2)
      model K threshold complexity        BIC
1 AKJBKQKDK 6       0.2      3,371 -87,108.74

SELECTED: model  AKJBKQKDK  with  6  clusters.
Selection Criterion: BIC.
```

기온 ~ 습도

```
> res.multi3<-funHDDC(list(fd.temp1991_high1.1, fd.rain1991.1),
+                     K=6,init="kmeans",threshold=0.2)
      model K threshold complexity        BIC
1 AKJBKQKDK 6       0.2      5,709 -61,065.00

SELECTED: model  AKJBKQKDK  with  6  clusters.
Selection Criterion: BIC.
```

기온 ~ 강수량

```
> res.multi4<-funHDDC(list(fd.temp1991_high1.1, fd.sun1991),
+                     K=6,init="kmeans",threshold=0.2)
      model K threshold complexity        BIC
1 AKJBKQKDK 6       0.2      4,543 -40,526.96

SELECTED: model  AKJBKQKDK  with  6  clusters.
Selection Criterion: BIC.
```

기온 ~ 일조시간

```
> res.multi5<-funHDDC(list(fd.temp1991_high1.1, fd.wind1991.1),
+                     K=6,init="kmeans",threshold=0.2)
      model K threshold complexity        BIC
1 AKJBKQKDK 6       0.2      3,763 -33,469.10

SELECTED: model  AKJBKQKDK  with  6  clusters.
Selection Criterion: BIC.
```

기온 ~ 풍속

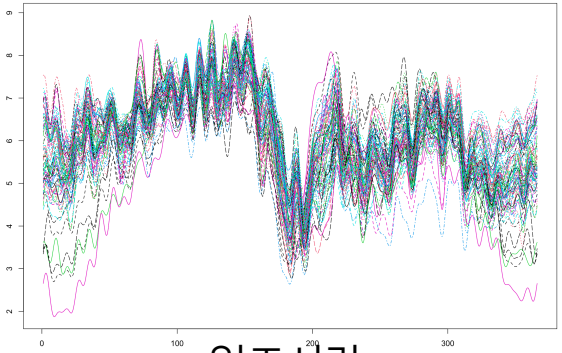
```
> table(res.multi1$class)

 1  2  3  4  5  6
 5 15 10  8 35 12
> table(res.multi2$class)

 1  2  3  4  5  6
15  6  3  8 22 31
> table(res.multi3$class)

 1  2  3  4  5  6
 3 24 25 13 11  9
> table(res.multi4$class)

 1  2  3  4  5  6
 3 22 13 17  6 24
> table(res.multi5$class)

 1  2  3  4  5  6
 9 12  2 17 35 10
```

- 군집 index 불일치
- 군집의 중심 (centroid) 찾을 수 없음

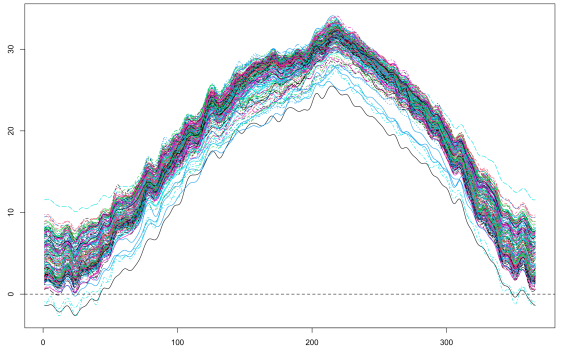
Multivariate Clustering

- [Trial] 특정 조합에 대해 Multivariate Clustering 을 진행한 후, 기온 예측

4



일조시간



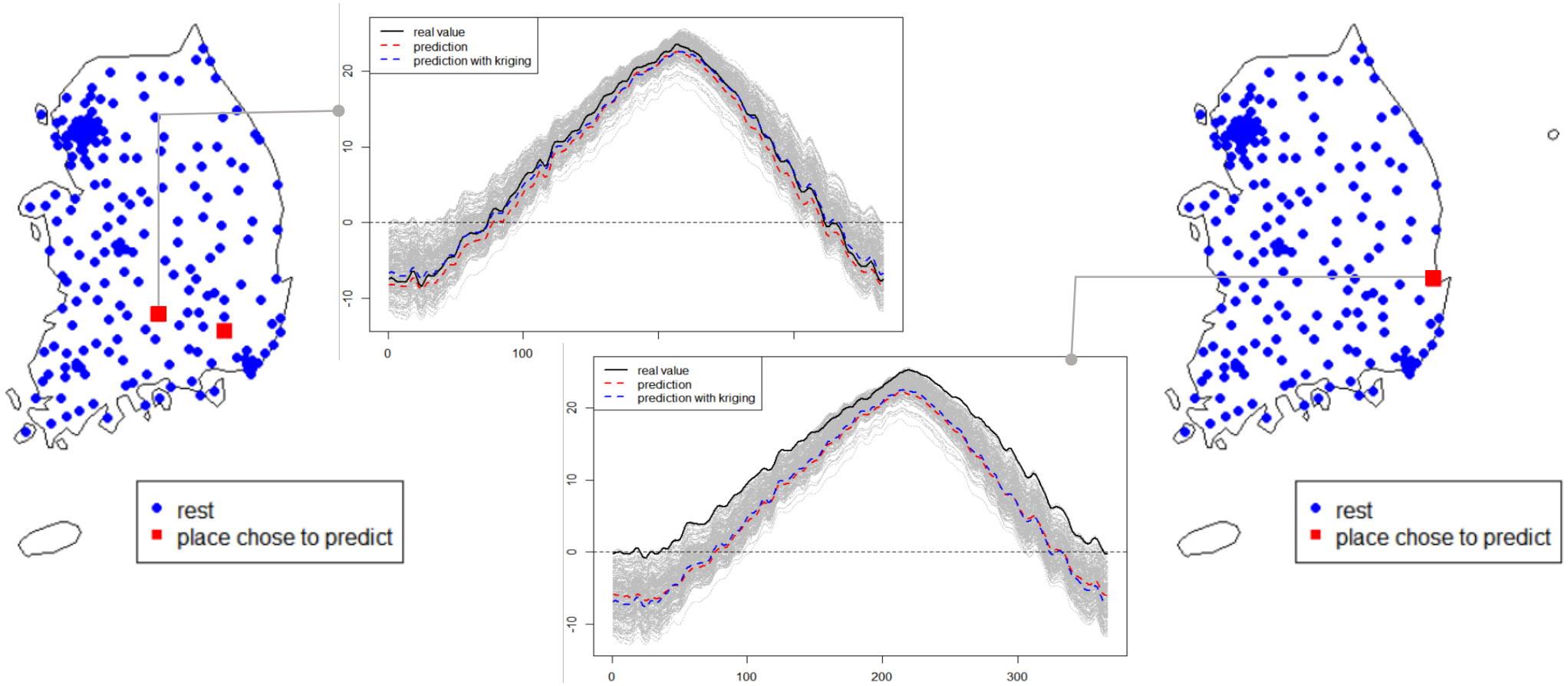
기온

FunHDDC Cluster	1	2	3	4	5	6
	3	22	13	17	6	24

3 CONCLUSION

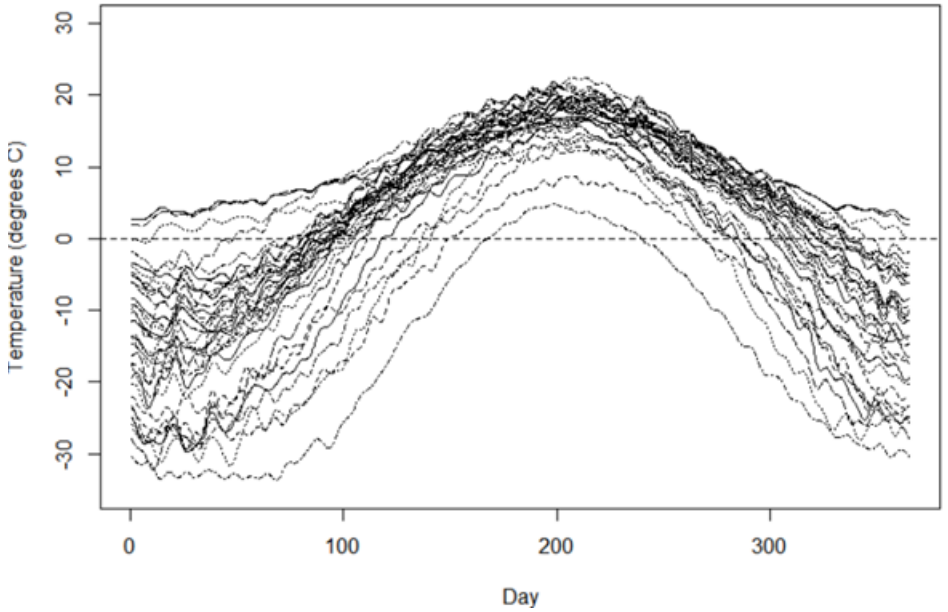
결론

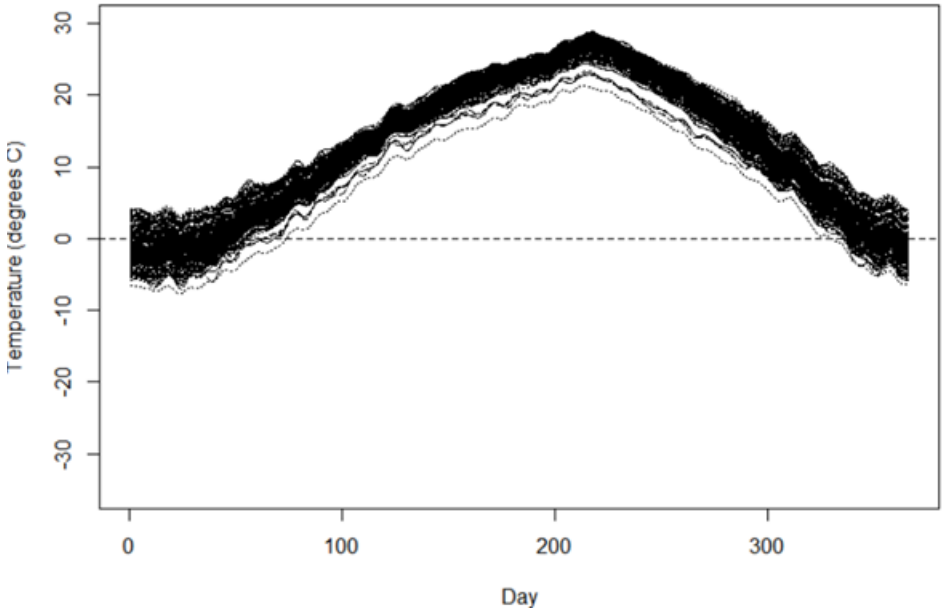
- 예측이 잘 된 지역 & 예측이 어려운 지역



캐나다 & 한국 기온 비교

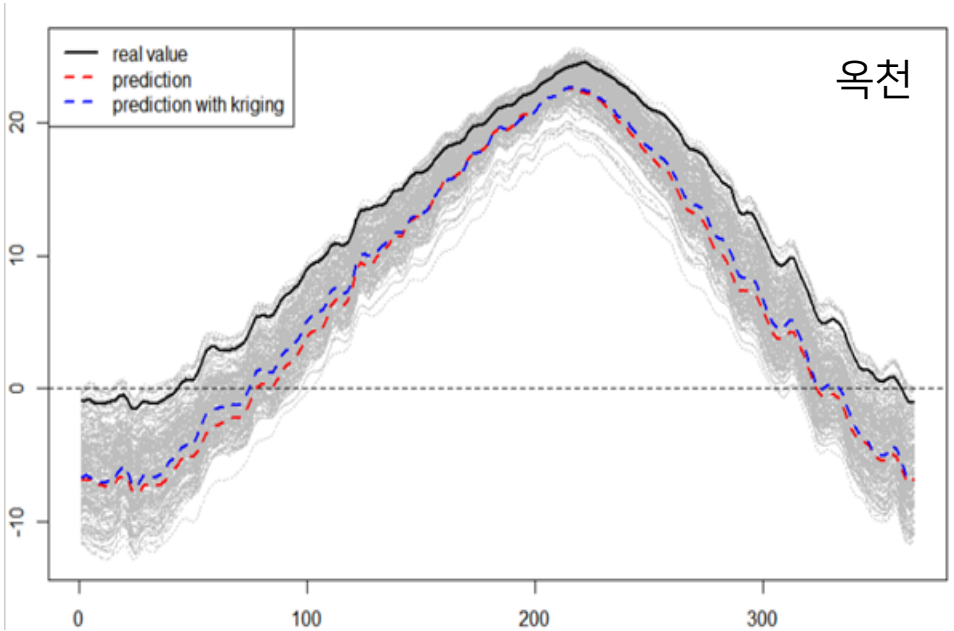
- Range 를 맞춘 후 비교해 보았을 때, 한국의 예측력이 높다는 사실을 확인할 수 있음



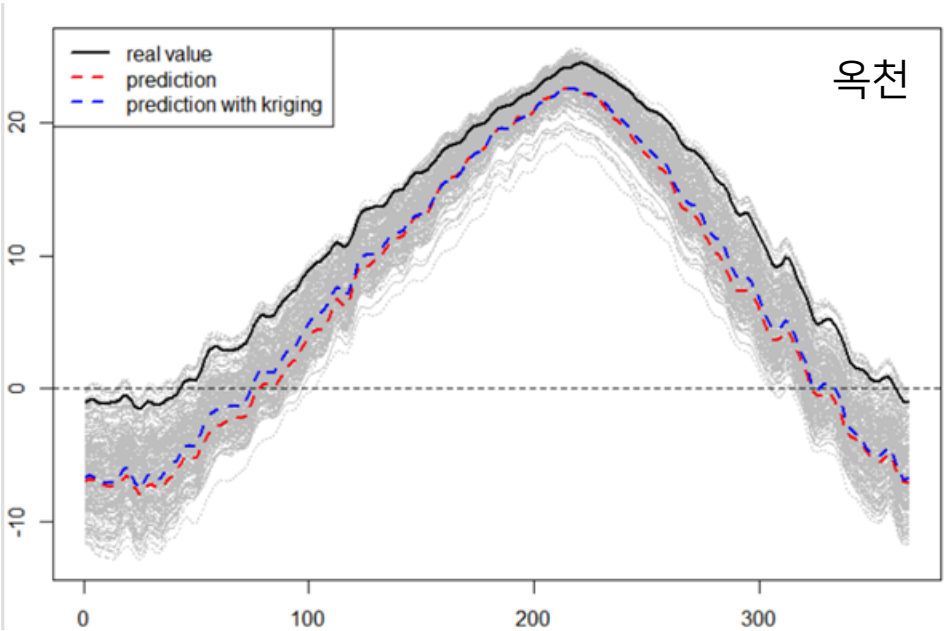


Appendix

- B-spline 과 Fourier Basis 비교



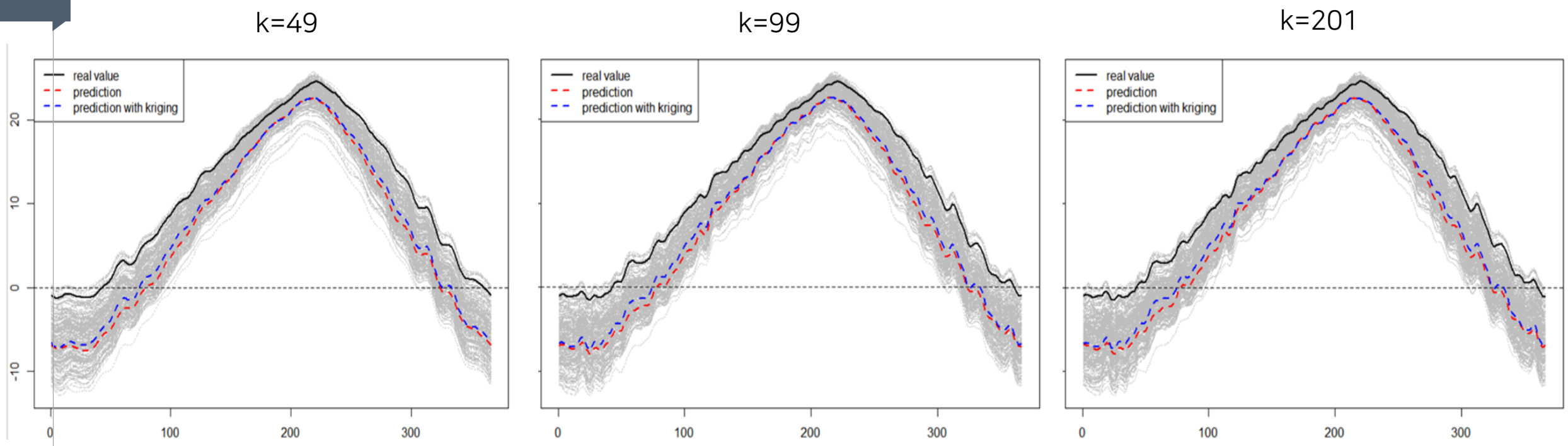
B-spline Basis



Fourier Basis

Appendix

- K의 개수에 따른 Fourier Basis 비교



감사합니다