



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

공 학 석 사 학 위 논 문

발전량 데이터 분석을 통한 신재생에너지  
주택지원사업지의 태양광발전 시스템 이상 감지

2022년 2월

부경대학교 대학원

에너지자원공학과

김 다 원

공 학 석 사 학 위 논 문

발전량 데이터 분석을 통한 신재생에너지  
주택지원사업지의 태양광발전 시스템 이상 감지

지도교수 최 요 순

이 논문을 공학석사 학위논문으로 제출함

2022년 2월

부 경 대 학 교 대 학 원

에 너 지 자 원 공 학 과

김 다 원

# 김다원의 공학석사 학위논문을 인준함

2022년 2월 25일



위 원 장 공학박사 여병철 (인)

위 원 공학박사 최요순 (인)

위 원 공학박사 정호영 (인)

# 목 차

표 목차 .....	ii
그림 목차 .....	iii
ABSTRACT .....	iv
<b>1. 서론 .....</b>	<b>1</b>
<b>2. 연구지역 .....</b>	<b>3</b>
<b>3. 연구방법 .....</b>	<b>5</b>
3.1 태양광발전 효율에 영향을 미치는 설계 및 환경 인자/ 선택변수 .....	6
3.2 태양광발전 시스템의 클러스터링 .....	7
3.3 이상 감지기준 설정 .....	9
<b>4. 연구결과 .....</b>	<b>11</b>
4.1 주변 환경인자 분석결과 .....	11
4.2 최적화된 클러스터 개수선정 .....	13
4.3 태양광발전 시스템 클러스터링 결과 .....	14
4.4 그룹별 태양광발전 시스템 이상 감지결과 .....	18
<b>5. 토의 .....</b>	<b>20</b>
5.1 현장검증 .....	20
5.2 이상 감지기준 구간의 타당성 .....	22
5.3 클러스터링의 필요성 검증 .....	24
<b>6. 결론 .....</b>	<b>26</b>
참고 문헌 .....	28

# 표 목 차

〈Table 1〉 PSF and $R^2$ value according to the number of clusters for selecting the optimal number of clusters .....	13
〈Table 2〉 Selected variables and clustering results for PV systems in the study area .....	16
〈Table 3〉 Statistical values of daily power generation data from PV systems at each group .....	17



# 그림 목 차

〈Fig 1〉 Aerial view of the study area (Gakbuk-myeon, Cheongdo-gun, Gyeongsangbuk-do, Korea) .....	4
〈Fig 2〉 Criterion for detecting the abnormalities when assuming a normal distribution .....	10
〈Fig 3〉 Four types of the intersections between Sun map and View map .....	12
〈Fig 4〉 Location of PV systems installed in the study area. Different colors and symbols were used for each group .....	15
〈Fig 5〉 Results of anomaly detection from PV systems at each group during November 2020. (a) Group 2. (b) Group 3. (c) Group 4. (d) Group 5 .....	19
〈Fig 6〉 View of PV systems detected as abnormalities: (a) PV_16. (b) PV_51. (c) PV_20. (d) PV_26 .....	21
〈Fig 7〉 Reference range of daily power production from PV systems except for IDs of 20 and 26 at the group 6 during November 2020 .....	23
〈Fig 8〉 Time-series analysis of hourly power production data from PV systems at each group in the study area. (a) Spring equinox. (b) Summer solstice. (c) Autumnal equinox. (d) Winter solstice .....	25

Anomaly detection of photovoltaic systems installed  
in renewable energy housing support project sites by analyzing power generation data

Da Won Kim

Department of Energy Resource Engineering, The Graduate School,  
Pukyong National University

### **Abstract**

In this study, we proposed a new method to detect the abnormalities among photovoltaic (PV) systems installed in renewable energy housing support project sites by analyzing power generation data. The site at the north side of Gakbuk-myeon, Cheongdo-gun, Gyeongsangbuk-do, Korea was selected as a study area where 63 PV systems have been installed and operated. By considering the factors related to system design and surrounding environment, the 63 PV systems were clustered into 6 groups using the K-means clustering method which is an unsupervised machine learning algorithm. The power production data from the PV systems for each group was analyzed and set as an abnormal value if it deviates from the range of  $\pm 2.58$  times the standard deviation from the mean (assuming a normal distribution and 99% confidence level confidence interval). As a result, several abnormalities among PV systems were detected during November 2020. The cause of abnormalities was confirmed through a site investigation. It is expected that the proposed method can be effectively used for fast problem diagnosis of PV systems in renewable energy housing support project sites.

# 1. 서론

오늘날 전 지구적인 기후변화에 대응하기 위해 전통적인 화석연료에서 신재생에너지로 에너지 시스템의 패러다임이 빠르게 변화하고 있다. 다양한 신재생에너지 기술 중에서도 태양광발전 기술은 차세대 성장에너지 원으로 잠재력이 높아 큰 주목을 받고 있다(Kwon et al, 2012). 국내에서도 2050 탄소중립 목표를 달성하기 위해 전체 에너지 시스템에서 태양광발전이 차지하는 비중을 지속해서 높이고 있어 태양광발전 시스템의 보급은 향후 더욱 확대될 전망이다(REMS, 2021).

국내에 보급된 태양광발전 시스템의 누적 설치량 증가율에 비해 시스템 운영 및 사후관리 문제가 꾸준히 지적되고 있으며, 이를 위한 효과적인 관리체계가 필요한 실정이다(Jung, 2019). 태양광발전 시스템의 모니터링과 문제진단을 위한 다양한 연구가 최근까지 수행되었다. Benjamin(2011)은 유럽 중앙부에 설치된 태양광발전소를 대상으로 운영관리가 시행된 태양광발전소와 그렇지 않은 태양광발전소의 PR(Performance Ratio, 성능)을 비교하였다. 운영관리가 시행된 태양광발전소는 81.3% PR을 나타내지만, 운영관리가 시행되지 않은 발전소의 경우 76.9%로 비교적 낮은 수준이었다. Benkercha and Moulahoum(2018)은 의사결정트리 알고리즘을 이용해 시스템 결함 검출 및 진단 방식을 제안하였다<sup>5</sup>. 5일간의 평가 결과 각각 99.86%, 99.80%의 진단 정확도를 보였다. 그 외에도 태양광발전 시스템 모니터링 기술들이 활발히 개발되고 있다(Anwari et al, 2011, Forero et al, 2006, EPRI, 2010, Woyte et al, 2013). 더불어 국제전기기술위원회(International Electrotechnical Commission, IEC)에서도 대규모 태양광발전 시스템의 유지관리를 위한 표준에 대해 논의하고 있다(Woyte,

2013).

본 연구에서는 발전량 데이터 분석을 통해 신재생에너지 주택지원사업지에 설치된 태양광발전 시스템의 이상 징후를 빠르게 감지하는 방법을 제안하고, 경상북도 청도군 각북면을 연구지역으로 선정하여 제안한 방법을 적용하였다. 그 결과 연구지역에 설치된 태양광발전 시스템을 설계 인자와 주변 환경인자를 고려해 6개의 그룹으로 분류할 수 있었고, 그룹별로 발전량 데이터의 통계적 분석을 통해 시스템 이상 발생을 빠르게 감지할 수 있었다.



## 2. 연구지역

연구지역은 경상북도 청도군 각북면으로 선정하였다(Fig. 1). 경상북도 청도군은 2017년부터 태양광, 태양열, 지열 등 2종 이상의 에너지원을 주택, 공공, 산업건물 등에 설치해 주민참여형 에너지 자립을 확대하는 신재생에너지 융복합지원사업지로 선정되었다. 청도군에는 9개의 읍·면이 있으며 그 중 연구지역으로 선정된 각북면은 청도군의 북서쪽에 있다. 각북면은 명대리, 우산리, 삼평리, 남산리, 덕촌리 등 8개의 리를 담당하며 각북면 대부분은 산지로 구성되어있다. 동쪽에는 가지산(1,241m), 운문산(1,195m)이 솟아있고 서쪽에는 비슬산(1,083m), 북쪽에는 선의산(756m), 구룡산(675m) 그리고 남쪽에는 화악산(932m), 천왕산(619m) 등으로 둘러싸여 있는 분지 지형이다. 내륙에 위치해 한서의 차가 크며 연평균기온은 13.2℃, 연 강수량은 1,243mm이다.

연구지역에는 총 63개의 태양광발전(PV) 시스템이 설치되어 있다. 기존의 상업용 송전선에 물려 발전된 전기를 기존 계통선로에 송전하는 개통연계형으로 설계되었으며 고정형 시스템이다. 모두 동일 모델의 모듈과 인버터를 사용하였으며 모듈은 260W를 출력하고 인버터 정격용량은 3.1 kW다. 설치주택에 따라 그림자의 영향을 최대한 받지 않는 옥상, 경사 지붕, 옥외등 다른 곳에 설치되었고 최대 발전량을 얻을 수 있는 최적의 방위각과 경사각으로 조절되었다.



Fig 1. Aerial view of the study area (Gakbuk-myeon, Cheongdo-gun, Gyeongsangbuk-do, Korea)

### 3. 연구방법

본 연구에서 선택한 신재생에너지 주택지원사업지에 설치된 태양광발전 시스템은 모두 동일 모델의 모듈과 인버터를 사용하였지만, 모듈의 개수와 설치 환경에 차이가 있다. 따라서 연구지역에 설치된 전체 PV 시스템에 대하여 동일한 이상 감지 기준을 적용하기에는 어려움이 있다고 판단된다. 이에 본 연구에서는 태양광발전량에 영향을 미치는 설계 및 환경인자를 고려한 클러스터링(군집화) 알고리즘을 적용한 후, 그룹별로 이상 감지 기준을 설정해 이상값을 감지하는 기법을 개발하였다.



### 3.1 태양광발전 효율에 영향을 미치는 설계 및 환경인자/ 선택변수

정확한 태양광발전량을 파악하기 위해서는 설치 모듈의 출력량, 모듈 수량, 방위각, 경사각 등 설계변수와 기상조건, 그림자와 같은 환경변수 요소들이 고려되어야 한다(Kang, 2012). 신재생에너지 주택지원사업지에 설치된 태양광발전 시스템의 설계환경은 설치장소에 따라 차이가 있다. 모두 동일 제품의 모듈과 인버터를 사용하였지만, 설치 환경에 따라 옥상, 옥외, 경사 지붕 등 설치 위치의 차이가 있으며 설치된 모듈의 수량도 설치면적에 따라 차이가 있다. 모듈의 수량에 따라 총설치 용량이 달라지므로 총 설치용량을 설계 변수로 선정하였다.

태양광발전 시스템의 발전량에 가장 큰 영향을 미치는 요소는 일사량이며 연구지역의 위치별 각 하늘 영역에 따른 일사량을 밀도로 나타내는 Sun map을 사용하였다. 연구지역에서 정동(east), 정서(west) 방향을 바라봤을 때 운전자의 시야각에 들어오는 태양의 시간을 GIS 소프트웨어인 ArcGIS 프로그램을 이용하여 계산하였다. Rich 등이 개발한 반구형 Viewshed 알고리즘은 반사 일사량을 제외한 직달일사량과 산란일사량의 합인 전일사량을 계산한다(Fu, 2002). 지면과 주변 사물을 고려하였을 때, 태양광 패널에 도달하는 반사 일사량은 비교적 작은 값으로 본 연구에서는 그 영향이 미미하다고 가정하였다. Sun map을 통해 태양의 위치를 중심으로 일사량의 분포를 확인할 수 있다. 그림자는 태양광발전 시스템의 성능에 큰 영향을 미치는 요소이다(Kang, 2012). 연구지역의 지형적 특성을 고려하였을 때, 위치별 음영의 정량적인 수치를 산정하고자 하였다. 이에 수치표고 모델(Digital Elevation Model, DEM)을 활용하여 연구지역의 각 위치에 따른 View map을 제작하였다. View map은 Sun map과 반대로 각 지점의 하늘 영역에서 가린 정도를 나타낸다. 본 연구에서 Sun map 좌표상에 View map을 투과시켜 투과 여부에 따라 0과 1로 구분하여 각 위치의 일사량에 미치는 그림자의 범위(면적)를 계산하였다.

### 3.2 태양광발전 시스템의 클러스터링

PV 시스템이 설치된 주택의 각기 다른 환경을 고려해 클러스터링하고 각 그룹의 이상값 기준을 설정하기 위해, 본 연구에서는 머신러닝 비지도 학습에서 사용되는 대표적인 알고리즘인 K-means 클러스터링(MacQueen, 1967) 방법을 이용하였다. 클러스터링을 통한 설계·환경 유사 그룹 생성을 위해서는 선택 변수를 정의해야 하며 본 연구에서는 모듈의 용량, Viewshed와 일사량에 따른 픽셀 수를 선택 변수로 정의하였다. 또한, 유사도를 측정하는 방법으로 시계열 자료의 클러스터링에서 가장 일반적으로 활용되는 유클리드 거리(Euclidean distance)를 적용하였다(Lee, 2020).

K-means 클러스터링은 임의로 설정된 K개의 초기 중심점(centroid)과 각 객체(데이터)와의 직선거리 제곱의 합을 최소화하는 것을 목적으로 한다. 이는 관성 최소화(minimization of inertia)라고도 알려져 있다(Bonnin, 2017). 모든 객체는 임의로 설정된 K개의 중심점과 거리를 계산하여 거리가 가장 짧은 클러스터에 배정되며 중심점 변화가 충분히 작아질 때까지 반복된다. 이때 각 클러스터에 소속된 객체의 결과는 매번 다르게 나타난다. 본 연구에서 PV 시스템별 유형 분류를 위하여 고려한 변수는 각 PV 시스템의 모듈 출력량, 모듈의 수, 총발전량, 설치 높이, Sun map과 View map의 교점 면적이다. 이 중 변수 간 상관관계가 높은 변수를 제외하고 사용변수를 표준화한 다음 클러스터링을 반복 수행하였다.

K-means 클러스터링은 대용량 데이터를 이용에 편리하지만 임의로 설정된 초기중심점에 따라 수행할 때마다 다른 클러스터링 결과가 나온다. 그리고 사용자가 클러스터의 개수를 사전에 설정해야 하는 단점도 있다. 최적의 클러스터 개수를 정량적으로 설정하기 위한 척도로 Pseudo-F statistic(PSF) 와 결정계수( $R^2$ )를 적용하였다. PSF는 클러스터 내 동질성을 설명하며  $R^2$ 는 클러스터에 의해 설명되는 변이의 정도를 나타낸다. 아

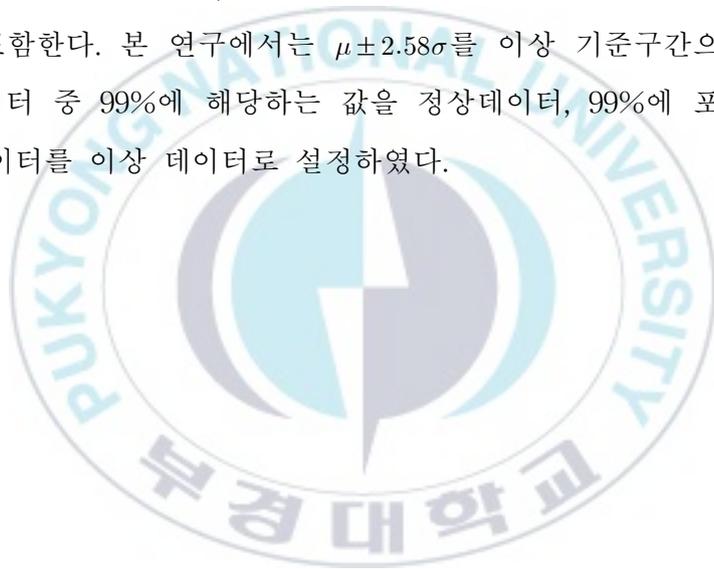
래 식의  $R^2$ 는 예측 모델의 추정 정도가 실험 결과를 표현하는 적합 정도를 나타내는 척도이다. 여기서  $x_i$ 는 객체(데이터),  $y_i$ 는 예측값,  $\bar{y}$ 는 예측값의 평균값,  $n$ 은 자료의 수이다.  $R^2$ 은 0과 1 사이의 값을 가지며 값이 1에 가까울수록 관측치가 회귀직선에 가깝다는 것을 의미한다. PSF 값이 최대치이고  $R^2$ 값이 최대한 클 때의 클러스터 수를 최적의 상태로 정하였다. 클러스터 개수(K)를 다양화한 후 반복적으로 클러스터링을  $n$ 번씩 수행하였다.

$$R^2 = 1 - \frac{\sum_{i=1}^n (x_i - y_i)^2}{\sum_{i=1}^n (x_i - \bar{y})^2}$$



### 3.3 이상 감지기준 설정

본 연구에서는 클러스터별 이상치 기준을 설정하고 발전량이 현저하게 낮은 경우 또는 평소와 다른 패턴을 나타낸 경우를 이상치 데이터로 구분하고 이상 감지 현상이 발생한 것으로 판단하였다. 정규분포를 가정할 때 이상치 기준(신뢰구간)은 Fig. 2과 같이 표현할 수 있다. 정규분포는 평균 ( $m$ )값에서 최대 확률값을 가지면 평균에서 멀어짐에 따라 확률은 0 값에 근접하게 된다. 분포곡선과 x축으로 둘러싸는 넓이는 전도 수를 나타낸다. 은 전체 데이터 중 68.3%, 의 범위는 95.5% 그리고 의 범위는 99.7%의 데이터를 포함한다. 본 연구에서는  $\mu \pm 2.58\sigma$ 를 이상 기준구간으로 설정하여 전체 데이터 중 99%에 해당하는 값을 정상데이터, 99%에 포함되지 않는 1%의 데이터를 이상 데이터로 설정하였다.



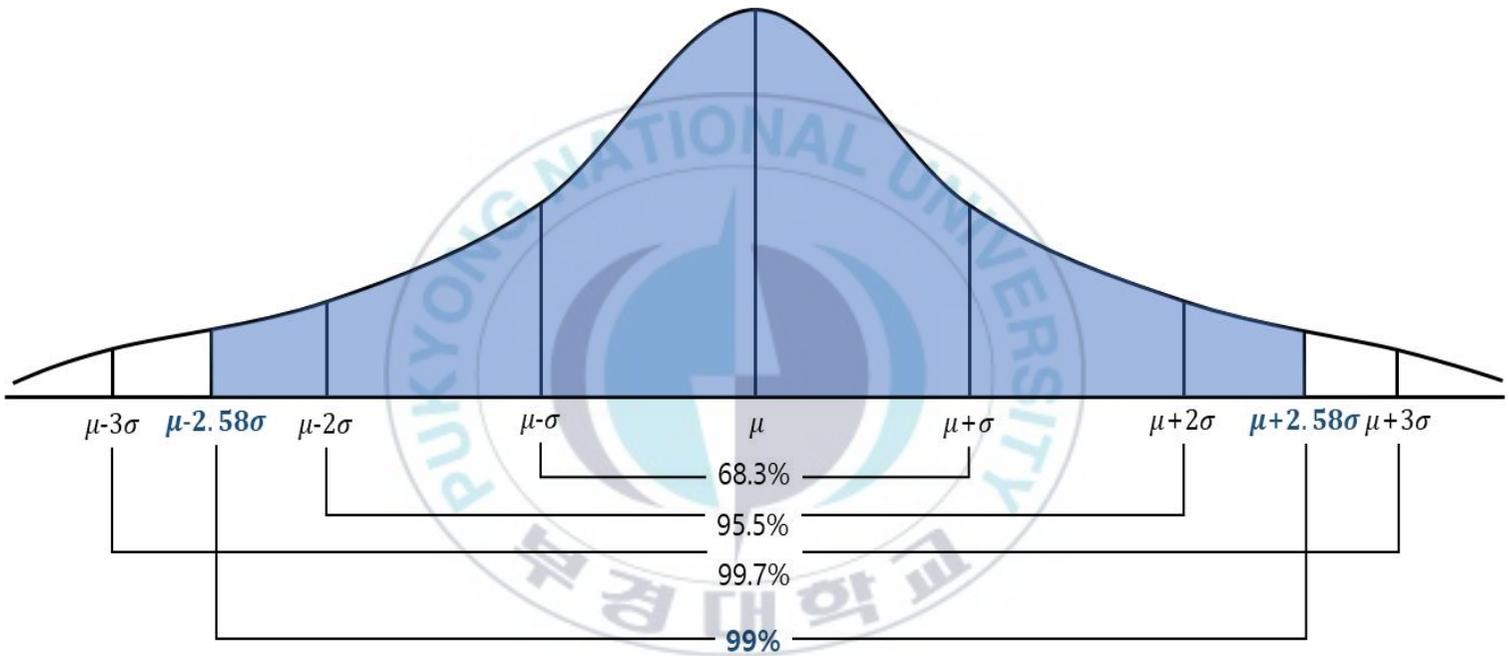


Fig 2. Criterion for detecting the abnormalities when assuming a normal distribution

## 4. 결과

### 4.1 주변 환경인자 분석결과

연구지역의 태양광발전 모델은 모두 같지만, 설치 환경에 따라 모듈의 개수, 설치 위치(설치 높이), 설치 방향 등의 차이가 있다. 수집된 데이터 중 클러스터링 선택 변수로 활용 가능한 정보는 방위각, 경사각, 설치 높이, 모듈 출력량, 모듈 수량, 총 설치용량, 동쪽과 서쪽 일사량에 미치는 그림자의 면적으로 총 8가지 항목이다. 이 중 공통적인 정보이거나 발전량에 미치는 영향이 비교적 미비한 항목은 분석에서 제외하였다.

ArcGIS 프로그램을 이용해 총 63개 PV 시스템에 대한 Sun map과 View map을 제작하고 두 map을 이용해 정동과 정서에서 일사량에 미치는 그림자의 면적을 각각 측정하였다. 좌우 면적에 따라 Fig 3과 같이 4개의 타입으로 분류하였고 이후 클러스터링 과정에서 선택 변수로 활용하였다.

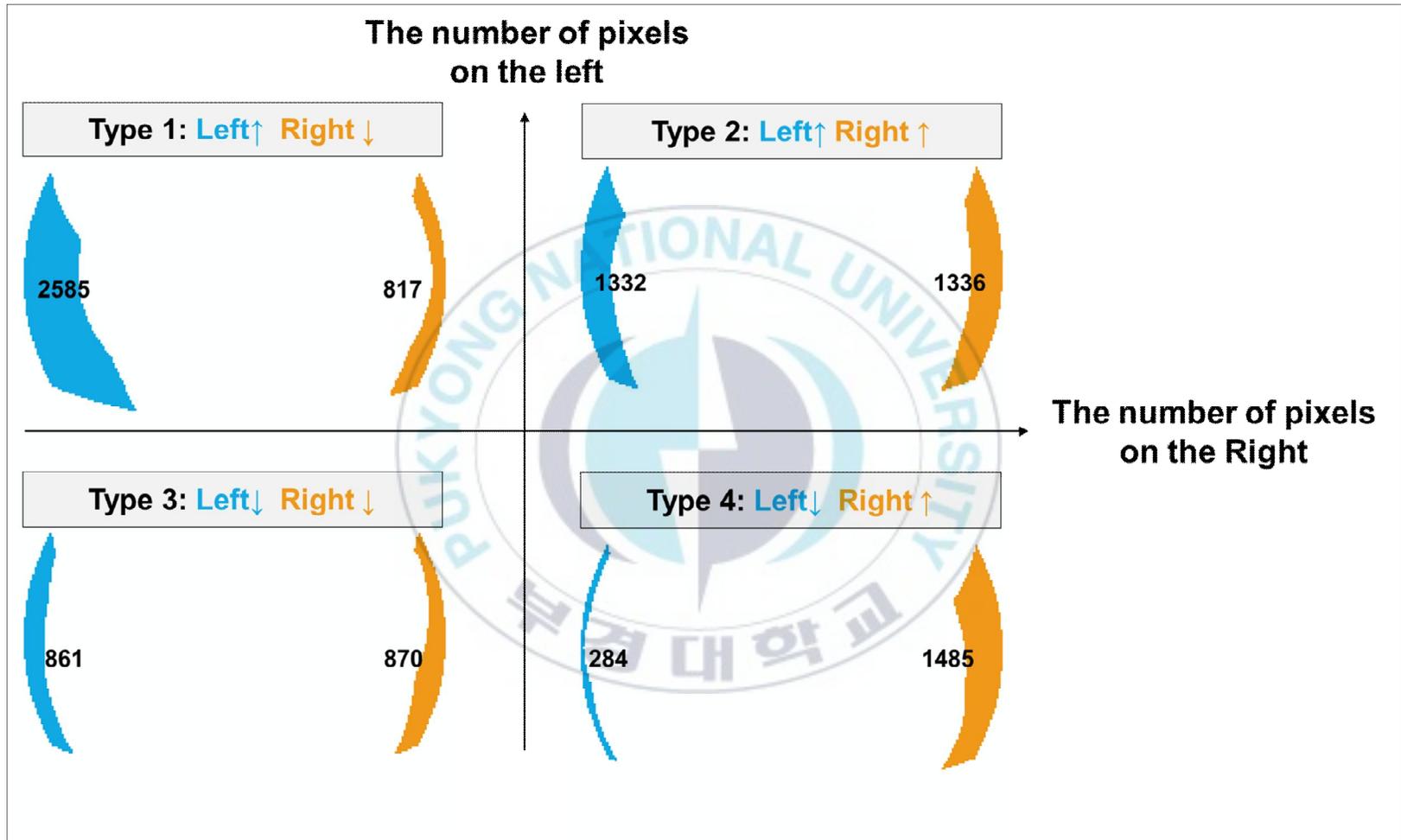


Fig 3. Four types of the intersections between Sun map and View map

## 4.2 최적화된 클러스터 개수 선정

본 연구에서는 효과적인 이상값 검출을 위해 대규모 태양광발전시스템을 설계와 환경 요소에 따라 클러스터링하였다. K-means 클러스터링은 비지도 학습 모델로 학습 과정이 없이 사용할 수 있지만, 사전에 클러스터 개수를 설정해야 한다. 최적의 클러스터 개수를 설정하기 위한 척도로 PSF와  $R^2$ 를 계산하였다. 클러스터링 결과의 타당성을 정량적으로 판단하기 위해 PSF 통계량이 일반적으로 사용된다(Milligan, 1958, Statistical research institute, 2008). PSF는 클러스터 수의 변화에 따라 국부적 최고점을 보이는 곳에서 최적의 클러스터 수를 결정하는 방법으로 클러스터 간의 분리 정도를 측정하여 값이 크면 더는 클러스터를 합치는 것이 의미가 없다고 본다(Yi, 2006).

클러스터 개수를 2에서부터 하나씩 늘려가며 PSF와 값을 계산하였다. 값은 최소 0.75를 넘는 것을 기준으로 설정하였다. 계산 결과 클러스터 개수가 많을수록 PSF 값과 값은 지속적으로 높게 계산되었다. 하지만 클러스터 개수가 7개가 넘어간 이후부터 각 클러스터에 해당하는 객체의 수가 5개 이하로 분류되었고 그 결과 본 연구에서 제시한 이상 감지 구간 계산 방식에 적용하기에 어려움이 있다 판단하였다. 따라서 K는 클러스터별 객체 수가 이상 감지 구간 계산방식에 적용 가능하며 PSF와 가 최대 수치로 계산되는 6으로 선택되었다(Table 1).

Table 1. PSF and  $R^2$  value according to the number of clusters for selecting the optimal number of clusters

Number of clusters	PSF	$R^2$
6	42.7	0.81
5	31.8	0.76
4	31.6	0.68
3	25.1	0.50
2	21.3	0.28

#### 4.3 태양광발전 시스템 클러스터링 결과

선택 변수는 총 설치용량과 일사량에 미치는 그림자 면적을 선택 변수로 정의하였다. 각 PV 시스템의 선택 변수에 해당하는 값과 클러스터링 결과는 Table 2에 제시하였다. 63개의 PV 시스템은 2개, 9개, 16개, 8개, 23개, 5개씩 각각 6개의 그룹으로 클러스터링 되었다. 클러스터링 결과 총 설치용량, 설치 높이 그리고 정동, 정서 면적에 대해 타당하게 구분되었음을 확인하였다. Table 3은 6개의 그룹별로 포함된 PV 시스템의 수와 일 발전량 데이터의 주요 통계값을 보여준다. Fig 4는 연구지역에 설치된 PV 시스템들의 위치를 클러스터링 결과에 따라 도시한 것이다.

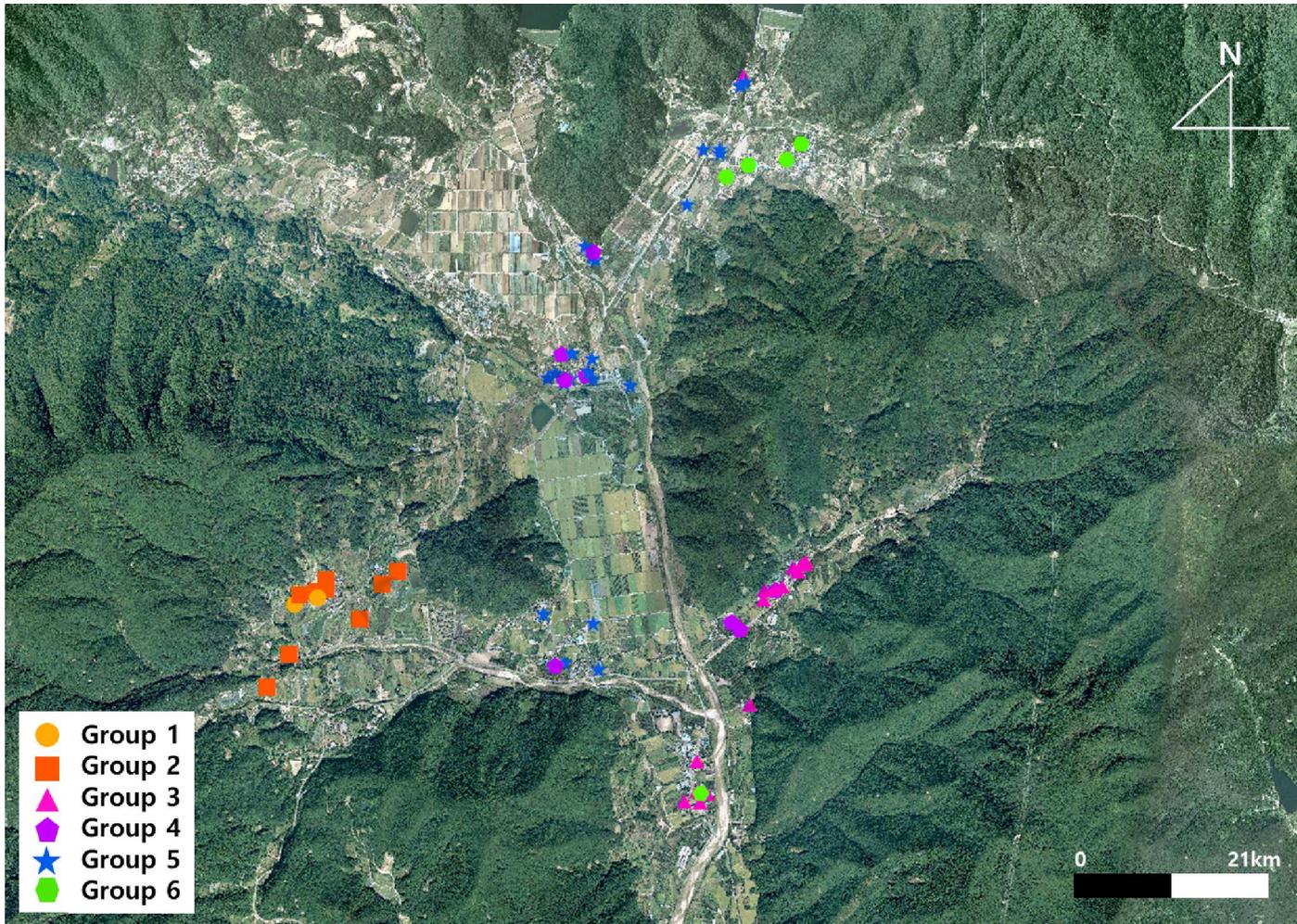


Fig 4. Location of PV systems installed in the study area. Different colors and symbols were used for each group

Table 2 Selected variables and clustering results for PV systems in the study area

ID	Total installed Production (kW)	Height Offset (m)	Left Pixels	Right Pixels	Group	ID	Total installed Production (kW)	Height Offset (m)	Left Pixels	Right Pixels	Group
PV_01	3.12	3	710	914	5	PV_34	3.12	5	952	720	5
PV_02	3.12	4	746	868	5	PV_35	3.12	2	962	730	5
PV_03	2.08	4	284	1485	1	PV_36	3.12	2	1093	988	5
PV_04	3.12	4	297	1272	2	PV_37	3.12	4	903	745	5
PV_05	3.12	2	303	1294	2	PV_38	2.6	3	946	730	4
PV_06	3.12	4	295	1351	2	PV_39	3.12	6	1171	455	6
PV_07	3.12	3	319	1286	2	PV_40	3.12	3	1060	540	6
PV_08	3.12	4	440	1131	2	PV_41	3.12	3	1037	727	5
PV_09	3.12	3	633	1158	2	PV_42	3.12	3	1163	468	6
PV_10	3.12	6	1421	1041	3	PV_43	2.08	4	298	1316	1
PV_11	2.08	4	1227	764	4	PV_44	2.08	5	956	838	4
PV_12	2.6	4	1296	756	4	PV_45	2.6	5	687	902	4
PV_13	2.6	4	1327	1040	4	PV_46	3.12	4	291	1451	2
PV_14	3.12	4	1235	1350	3	PV_47	3.12	4	1314	1316	3
PV_15	3.12	4	1327	1040	3	PV_48	3.12	4	1332	1336	3
PV_16	3.12	6	2585	817	3	PV_49	3.12	4	803	920	5
PV_17	3.12	4	1419	1040	3	PV_50	3.12	6	1300	1224	3
PV_18	3.12	4	1544	728	5	PV_51	3.12	2	1338	1031	3
PV_19	2.08	4	1110	791	4	PV_52	3.12	4	1210	1448	3
PV_20	3.12	4	1117	793	5	PV_53	3.12	4	1399	1253	3
PV_21	3.12	5	842	900	5	PV_54	3.12	6	1076	532	6
PV_22A	3.12	4	861	870	5	PV_55	3.12	4	1042	709	5
PV_22B	3.12	4	861	870	5	PV_56	3.12	2	625	977	5
PV_23	2.6	4	912	849	4	PV_57	3.12	4	1521	1002	3
PV_24	3.12	4	955	860	5	PV_58	3.12	4	1427	1258	3
PV_25	3.12	4	954	809	5	PV_59	3.12	4	384	1149	2
PV_26	3.12	4	964	840	5	PV_60	3.12	4	1165	690	5
PV_27	3.12	4	1153	788	5	PV_61	3.12	4	459	68	6
PV_28	3.12	3	1110	793	5	PV_62	3.12	4	1435	1034	3
PV_31	3.12	6	1140	1414	3	PV_63	3.12	4	1402	1039	3
PV_32	3.12	4	1106	867	5	PV_64	3.12	4	332	1234	2
PV_33	3.12	4	1003	778	5						

Table 3 Statistical values of daily power generation data from PV systems at each group

	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
number of PV system	2	9	16	8	23	5
min	0	0	0	0	0	0
Q1	5.80	7.78	8.65	6.45	7.92	7.54
median	9.15	12.88	13.46	10.10	12.94	11.75
mean	8.39	11.96	12.50	9.38	12.25	11.13
Q3	11.37	16.38	16.75	12.61	16.56	15.08
max	14.60	21.98	22.11	16.51	139.86	20.2
Standard Deviation	3.72	5.64	5.53	4.15	9.05	5.07

#### 4.4 그룹별 태양광발전 시스템 이상 감지결과

클러스터링 된 4개 그룹의 2020년 11월 일 발전량 데이터를 이용해 일별 이상 기준을 설정하고 발전량이 현저하게 낮은 경우 또는 평소와 다른 패턴을 나타낸 경우의 데이터 검증을 수행하였다. 그룹 2와 그룹 4에서는 기준구간을 벗어나는 PV 시스템이 발견되지 않았고 그룹 3과 그룹 5에서 기준구간을 벗어나는 PV 시스템이 각각 2개씩 검출되었다(Fig. 5). 그룹 3의 대부분 데이터가 구간을 벗어나지 않지만, 평균보다 지속해서 발전량이 적거나 그룹의 발전량 패턴과 일치하지 않는 데이터 값이 발생한 경우가 검출되었다. 그룹 5에서는 기준구간을 벗어난 PV 시스템 3개가 검출되었다. 이 중 하나는 일시적인 감지로 추이를 지켜보았으나 기준구간을 벗어나는 경우가 없었다. 나머지 두 개의 PV 시스템은 지속적으로 발전량이 기록되지 않다가 급격하게 높은 발전량을 기록하는 패턴을 나타내었고, 기준값을 지속해서 벗어나는 것으로 확인되었다.

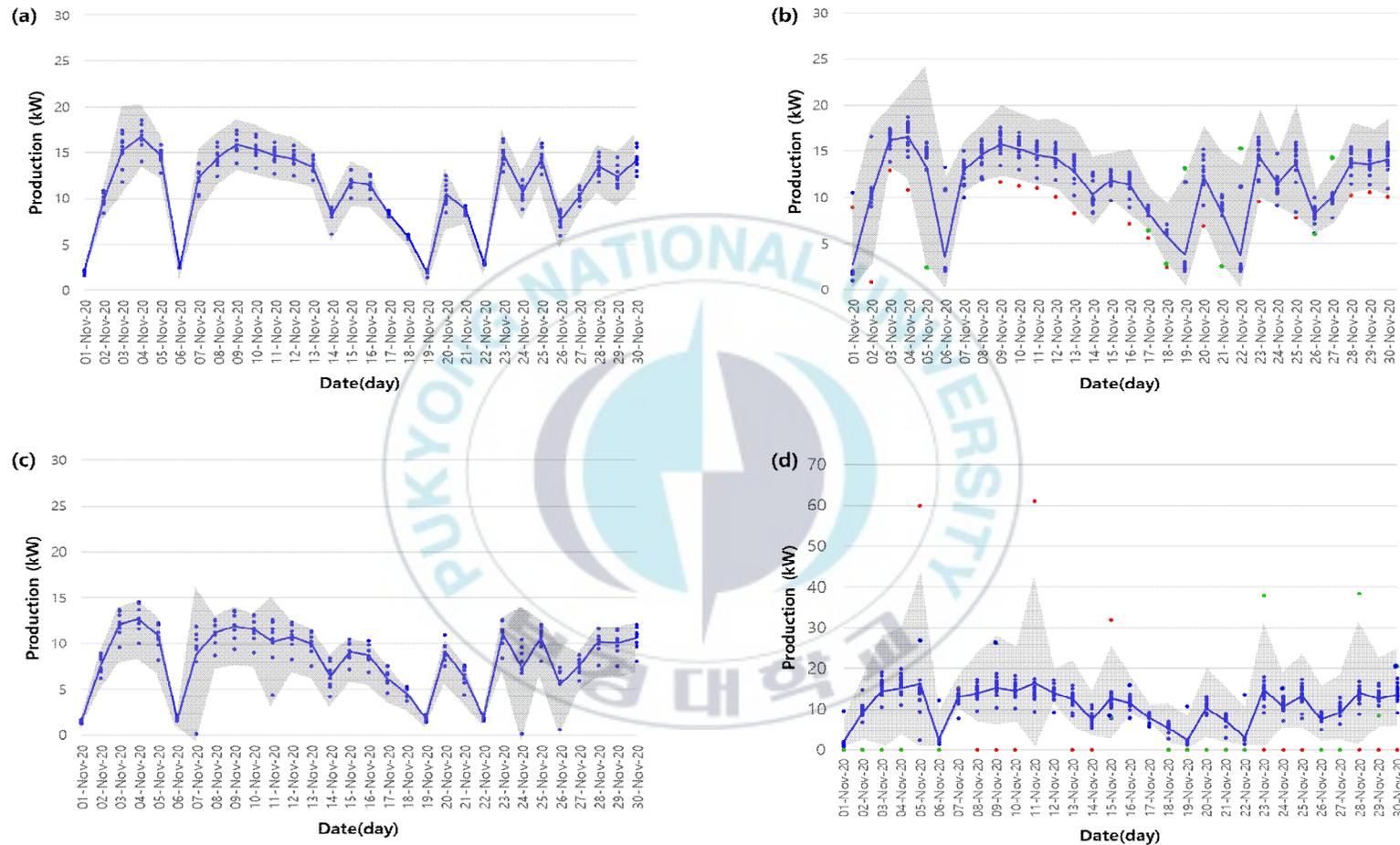


Fig 5. Results of anomaly detection from PV systems at each group during November 2020. (a) Group 2. (b) Group 3. (c) Group 4. (d) Group 5

## 5. 토의

### 5.1 현장검증

기준구간을 벗어나는 4개의 PV 시스템을 대상으로 현장 조사를 통해 이상값 발생 원인을 파악하였다(Fig. 6). 빈번하게 낮은 발전량이 기록된 PV\_16은 현장 방문 결과 산 바로 아래 위치하여 그림자의 영향을 높게 받고 있음을 확인하였다. 또한, PV\_51은 높은 건물이 2m 이내에 있어 같은 그룹 내 다른 PV 시스템들보다 비교적 낮은 발전량이 기록된 것으로 보인다. 급격하게 높은 발전량이 기록되어 있는 PV\_20과 PV\_26은 발전량 데이터를 전송하는 기기의 문제로 이상값이 발생한 것으로 판단된다.

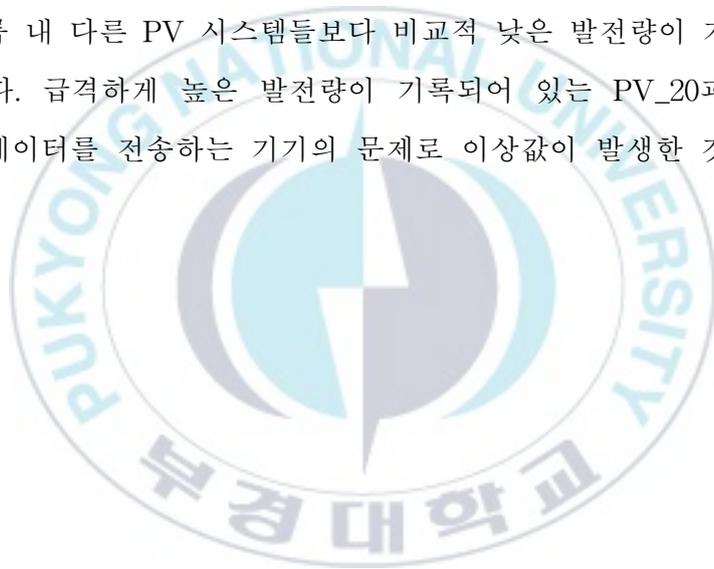




Fig 6. View of PV systems detected as abnormalities: (a) PV\_16. (b) PV\_51. (c) PV\_20. (d) PV\_26

## 5.2 이상 감지기준 구간의 타당성

본 연구에서는 모든 데이터를 이용해 이상 기준구간을 설정하고 태양광 주택보급 사업지에 적용하였다. 이상 기준구간에는 이상값을 가지는 PV 시스템의 데이터 또한 포함되어 있으며 이상값이 이상 기준구간에 높은 영향을 미치는 것을 Fig 7을 통해 알 수 있다. 이상값을 포함한 이상 기준구간이 적절한 기준인지 판단하기 위해 이상값을 가진 PV 시스템 데이터를 제외한 후 이상 기준구간을 제외하지 않은 이상 기준구간과 비교해 보았다. 이상 기준구간이 이상값의 영향을 가장 많이 받은 그룹 5의 2020년 11월 일발전량 데이터를 이용하였다.

이상값을 가진 PV 시스템의 데이터를 제외하지 않은 이상 기준구간(회색)과 제외한 이상 기준구간(파란색) 이상 기준구간이 안정된(고른) 것을 확인할 수 있다. 회색 구간에 속했던 데이터 대부분이 구간 안에 속해있으며, 일부 벗어난 PV 시스템이 있으나 지속적이지 않고 일시적으로 발생한 것이며 기존의 기준에서 이상값을 검출된 PV 시스템은 새로운 기준에서 또한 검출되었음을 알 수 있다. 또한, 기존의 이상 기준구간에서 이상값으로 검출되었던 PV 시스템은 파란색 이상 기준구간에서도 검출되었다. 이상값 데이터를 고려했음에도 불구하고 동일한 결과가 나타난 것으로 보아 본 연구에서 개발된 이상 기준구간은 이상값 검출을 위해 적절한 방법인 것으로 판단된다.

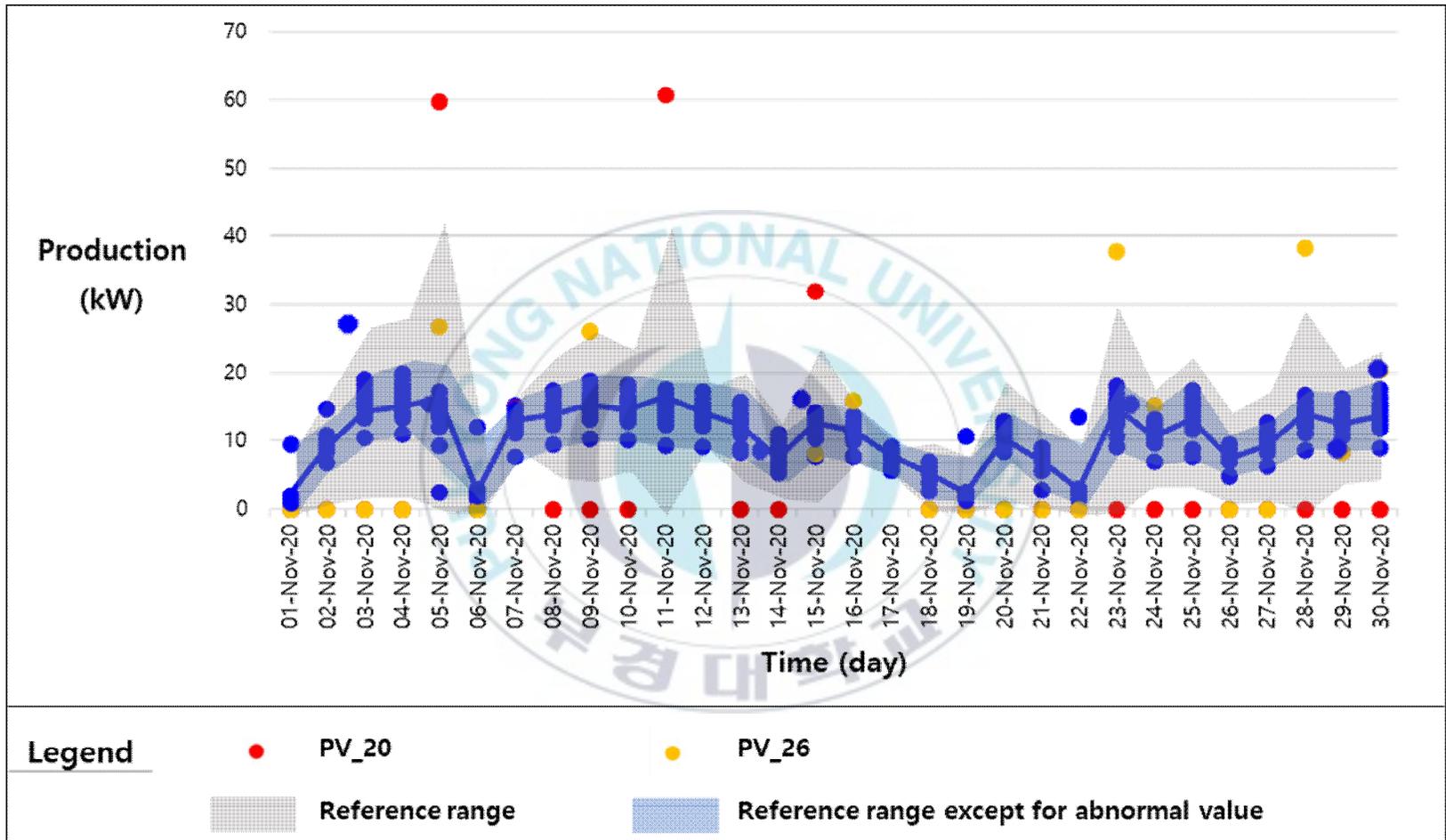


Fig 7. Reference range of daily power production from PV systems except for IDs of 20 and 26 at the group 6 during November 2020

### 5.3 클러스터링의 필요성

본 연구에서 선정된 지역은 주택 간의 거리가 멀지 않아 설치 환경의 차이가 크지 않다고 생각될 수 있다. 환경적 차이가 크지 않은 지역에 대해 클러스터링 작업의 필요성에 의문을 가질 수 있다. 따라서 클러스터링의 필요성을 확인하고자 추가적인 분석을 진행하였다. 태양이 움직이는 궤적을 15° 간격으로 24등분 한 24개의 절기 중 각 계절의 중앙에 해당하는 춘분, 하지, 추분, 동지의 시간별 발전량 데이터를 이용하여 분류된 클러스터의 절기별 발전량 패턴 분석을 수행하였다.

연구지역의 시간에 따른 절기별 발전량 그래프 결과는 Fig 8과 같다. 시간에 따른 각 그룹의 그래프는 절기별 비슷한 패턴을 보이거나 그룹별 발전량의 차이를 확인할 수 있다. 춘분의 11시에서 14시 사이, 하지의 15시, 추분의 14시에서 16시 그리고 동지의 10시에서 11시, 13시에서 15시 사이 최대 1kWh 이상의 그룹별 발전량 차이가 발생하였다. 이는 이상 진단을 판단하는 기준을 클러스터링을 통해 구분하지 않고 하나의 공통된 기준으로 설정할 경우 오류를 범할 수 있다. 주택지원사업의 특성에 따라 유사한 설계환경과 기후환경을 가지고 있지만, 태양광발전 시스템이 설치된 주택의 지형 및 환경에 차이가 존재한다. 이는 정확한 기준을 설정하고 올바른 판단을 위해 지역적 차별성이 없더라도 설치 환경을 고려한 클러스터링 과정이 필요하다는 점을 보여준다.

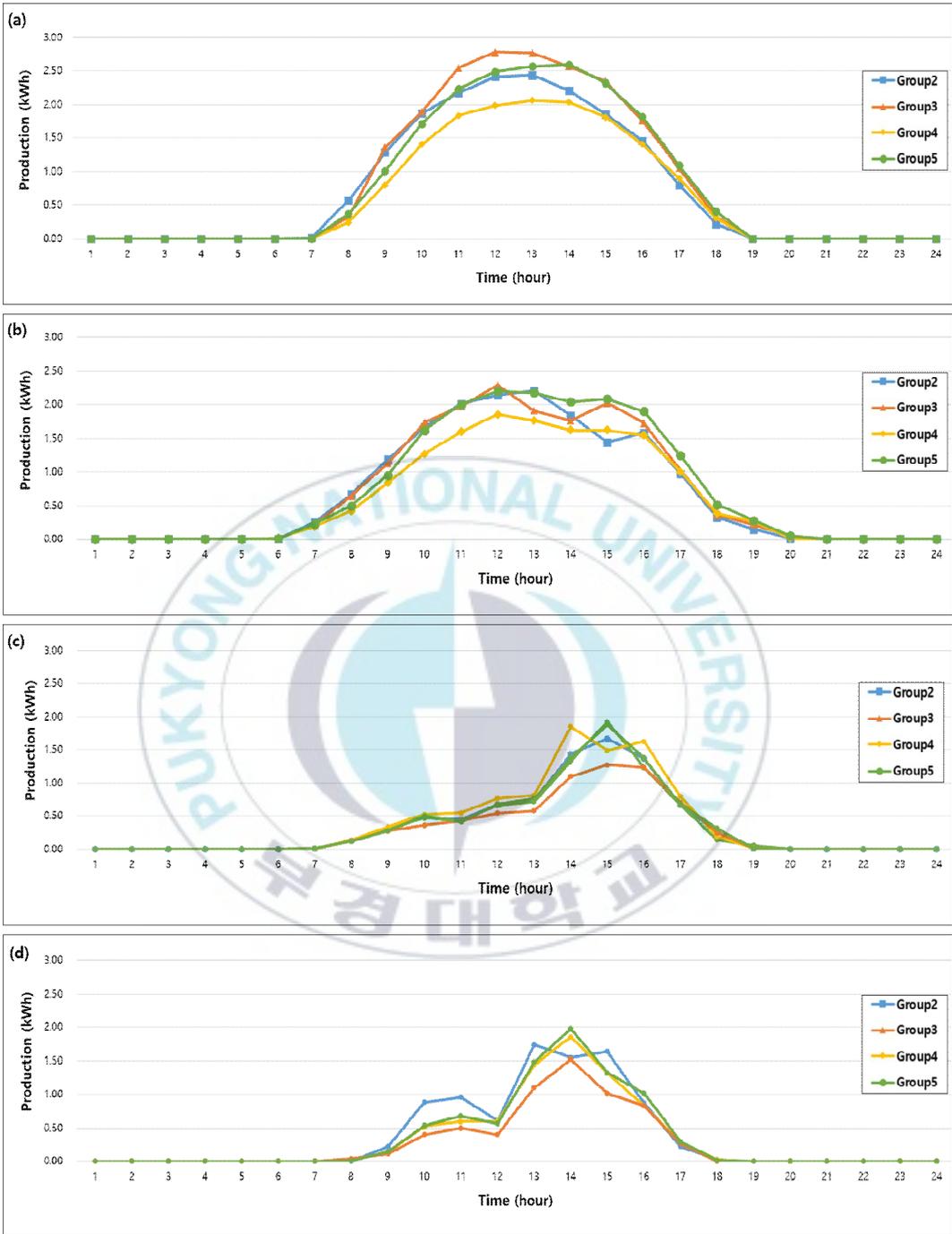


Fig 8. Time-series analysis of hourly power production data from PV systems at each group in the study area. (a) Spring equinox. (b) Summer solstice. (c) Autumnal equinox. (d) Winter solstice

## 6. 결 론

본 연구에서는 일 발전량 데이터 분석을 통해 신재생에너지 주택지원사업지에 설치된 태양광발전 시스템의 이상 발생을 조기에 감지할 수 있는 기법을 개발함으로써 효과적 모니터링을 가능하게 하고자 하였다. 본 연구의 결론은 다음과 같다.

- (1) 신재생에너지 주택지원사업지의 태양광발전 시스템 이상 감지를 위해 태양광 시스템의 설계인자(모듈의 발전량, 모듈의 수량, 방위각, 설치 높이, 총 설치용량)와 환경인자(일사량, 그림자)를 고려하여 이상 발생을 감지할 수 있는 기법을 개발하였다.
- (2) 일 발전량 데이터를 이용해 일별 평균과 표준편차를 계산하고 평균으로부터 표준편차의 2.58배 범위를 벗어나는 경우 즉, 데이터 중 99%에 포함되지 않는 1% 경우를 이상 감지 기준으로 설정하였다.
- (3) 다른 조건의 PV 시스템의 이상 진단을 판단하는 기준 하나로 정의하기에는 무리가 있다고 판단되어 머신러닝 비지도 학습 모델인 k-means 알고리즘을 이용하여 63개의 PV 시스템을 유사한 특징을 가진 6개의 군집으로 클러스터링하였다. 이 중 4개의 클러스터에 이상 발생을 감지 기준을 적용해본 결과 각각 2개씩 이상 기준구간을 벗어나는 PV 시스템이 감지되었다. 감지된 4개의 PV 시스템 중 PV\_16, PV\_51은 이상 감지 기준보다 낮은 발전량을 기록하였고 PV\_20, PV\_26은 이상 감지 기준보다 높은 발전량을 기록하였다.
- (4) 현장 검증을 통해 감지된 이상 감지 기준보다 낮은 발전량이 기록된 PV\_16과 PV\_51은 산 바로 아래 위치하거나 설치 높이보다 높은 건물이 근처에 있어 그림자의 영향을 받아 발전량이 지속해서 낮게 나타난 것을 확인하였다. 발전량이 높은 PV\_20과 PV\_26은 주변에 그림자의 영향을 주는 장애물이 없었고 패널의 오염 또한 확인되지 않았다. 따라서 PV\_26과 PV\_20은 발전량 데이터 전송 과정에 문제가 있는 것으로 판단된다.
- (5) 이상감지 구간이 이상값의 영향을 받아 변형되는 것을 확인하였고, 이에 이상감지구간의 타당성 확인을 위해 이상값을 제외한 정상데이터만을 이용해 이상

감지구간을 다시 계산하였다. (동일한 데이터를 이용하였으며, 기존 구간에서 검출된 이상값만을 제외하였다.) 기존의 이상 감지구간보다 안정된(고른) 형태가 되었지만 새로운 구간에서 검출된 이상값의 결과가 기존의 결과와 동일함을 확인하였다. 따라서 이상값이 이상 감지구간에 영향을 미치는 것은 맞으나 이상값 검출결과의 차이가 없으므로 본 연구에서 제시된 이상 기준구간은 이상값 검출을 위해 적절한 방법임으로 판단된다.

- (6) 환경적 차이가 크지 않은 지역에 대한 클러스터링 작업의 필요성 확인을 위해 각 그룹의 절기별 시간발전량을 비교하였다. 비교결과 절기별 각 그룹의 시간별 발전량이 유사한 패턴을 나타내고 있으나 일부 구간들에서 발전량이 최대 1kWh 이상 차이가 발생한 것이 확인되었다. 이상 진단을 판단하는 기준을 클러스터링을 통해 구분하지 않고 하나의 공통된 기준으로 설정할 경우 오류를 범할 수 있으므로 지역적 차별성이 없더라도 설치 환경을 고려한 클러스터링 과정이 필요하다는 점을 확인하였다.

## 참고 문헌

1. Kwon. S. J., Min. Y.B., Choi. J.S., Yoon. C.Y., Performance Analysis of Photovoltaic System for Greenhouse, Journal of Agriculture & Life Science, Vol. 46, No. 5, pp. 143-152, 2012.
2. REMS, Korea energy agency, <https://rems.energy.or.kr/pub/view/bbsNotice> (Accessed on 1 October 2021)
3. Jung. Y. H., A Study on Consumers Problems and Improvement Methods in Solar Energy, Korean Consumer Agency, 2019.
4. Benjamin. A. C., Technical certification of PV power plants: The key to bankability for commercial and utility scale solar, Solar Energy, 2011.
5. Moulahoum. S., Benkercha. R., Fault detection and diagnosis based on C4.5 decision tree algorithm for grid connected PV system, Solar Energy, Vol. 173, pp. 610-634, 2018.
6. Anwari. M., Dom. M. M, Rashid. M. I. M., Small Scale PV Monitoring System Software Design, Energy Procedia, Vol. 12, pp.586-592, 2011.
7. Forero, N., Hernandez, J., Gordillo, G., Development of a Monitoring System for a PV Solar Plant, Energy Conversion and Management, Vol. 47, No.15, pp.2329-2336, 2006.
8. EPRI, Addressing Solar Photovoltaic Operations and Maintenance Challenges &#8211; A Survey of Current Knowledge and Practices, USA, pp. 4-7, 2010.
9. Woyte. A., Richter. M., Moser. D., Mau. S., Reich. N, Jahn., Monitoring of Photovoltaic Systems: Good Practices and Systematic Analysis, European Photovoltaic Solar Energy Conference, p. 3686-3694, 2013.
10. Klise. G. T., Balfour, J. R., Keating. T. J., Solar PV O&M Standards and Best Practices &#8211; Existing Gaps and Improvement Efforts, Sandia National Laboratories, USA, p. 3686-3694, 2013.
11. Kang. S. H., Lee. Y. H., Hwang. J. H., Cho. Y. H., The analysis of the

- renewable energy supply ratio for the school building applied PV system, Journal of the Korean Solar Energy Society, Vol. 32, No. 2, pp. 50-57, 2012.
12. Fu. P., Rich. P. M., A Geometric Solar Radiation Model with Application in Agriculture and Forestry, Computers and Electronics in Agriculture, Vol. 27, No.1, pp. 25-35. 2002.
  13. MacQueen, J., Some methods for classification and analysis of multivariate observations, Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, Vol. 1, No. 14, pp. 281-297, 1967.
  14. Lee. I. M., Min. J. H., Kim. K. T., Kho. S. H., Generating Travel Patterns of Public Transportation Users Using a k-means Clustering Based on SmartCard Data, The Korean Society For Railway, Vol. 32, No. 3, pp. 204-215, 2020.
  15. Bonnin. R., Machine Learning for Developers: Uplift your regular applications with the power of statistics, analytics, and machine learning, Packt Publishing Ltd, UK, p. 60-72, 2017.
  16. Milligan. G. W., Cooper. M. C., An examination of procedures for determining the number of clusters in a data set, Psychometrika, Vol. 50, pp. 159-179, 1958.
  17. Statistical research institute, Segmentation of rural areas based on the attributes of agricultural and fishing villages, Korea rural economic institute, Technical report, pp. 40, 2008.
  18. Yi. S. K., Hong. S. E., Park. S. H., A Similar Price Zone Determination of Public Land Price Using a Hybrid Clustering Technique, The Korean Geographical Society, Vol. 41, No. 1, pp. 121-135, 2006.