creative commons

COMMONS DEED

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

dCollection

공학전문석사 학위 연구보고서

배전계통 실제 정전데이터를 활용한 배전설비 정전횟수와 기상요인 간 연관관계 모델링에 관한 연구

- 부산·울산지역을 중심으로 -

2022년 2월

서울대학교 공학전문대학원

응용공학과 응용공학전공

최 문 주

배전계통 실제 정전데이터를 활용한 배전설비 정전횟수와 기상요인 간 연관관계 모델링에 관한 연구

- 부산·울산지역을 중심으로 -

지도교수 문 승 일

이 프로젝트 연구보고서를 공학전문석사 학위 연구보고서로 제출함

2022년 2월

서울대학교 공학전문대학원

응용공학과 응용공학전공

최 문 주

최문주의 공학전문석사 학위 연구보고서를 인준함

2022년 2월

위 원 장 윤 용 태 (인)

위 원 문 승 일 (인)

위 원 김 국 헌 (인)

국문초록

본 연구는 부산, 울산지역을 중심으로 실제 정전데이터를 활용하여 배전설비 정전횟수와 기상요인 간 연관관계를 모델링하는 연구이다. 한국전력공사 부산울산지역본부의 소속된 13개 지사의 2015~2019년 5개년 간 실제 정전횟수를 종속변수로 설정하였고, 선행연구를 참고하여 기상청의 종관기상관측(ASOS)자료 중 풍속, 강수, 적설 3개를 고 영향 기상요인으로 설정하였다. 그러나 부산, 울산지역 특성상 적설빈도가 2015~2019년 5개년 간 단 2회 발생(2018. 1. 10, 2019 1. 31)하여, 통계 관측수의 부족으로 유의확률이 산출되지 않아 통계적인 해석이 불가하였다. 그래서 적설을 제외한 풍속과 강수를 고 영향 기상요인으로 설정하였다.

선행연구에 근거하여 2개의 고 영향 기상요인을 설정하였고, Pearson 상관분석을 통하여 풍속과 강수 중 상관관계가 높은 기상요인인 최대순간풍속과 일 강수량을 각각 최종 독립변수로 설정하였다. 선행연구처럼 특정 배전설비의 정전확률 예측모형이 아닌 전체 배전설비 정전횟수 예측모형을 모델링하기 위하여 IBM SPSS Statistics 26을 활용하여 다중 선형 회귀분석을 수행하였고, 다중 선형 회귀분석으로는 과소적합(Under-fitting) 모형이 도출되어, MATLAB R2021a의 Curve Fitting Toolbox를 활용한 다항 비선형 회귀분석을 시행하여 최적적합(Optimal-fitting) 형태의 정전횟수 예측모형을 도출할 수 있었다.

도출된 예측모형은 1차로 원본 데이터인 2015~2019년 5개년 데이터를 기반으로 예측한 정전횟수와 실제 정전횟수를 비교하여 검증하였고, 2차로 2020년의 데이터를 기반으로 예측한 정전횟수와 실제 정전횟수를 비교하여 정전모형을 검증하였다.

2020년의 데이터로 검증결과, (실제 정전횟수 - 예측 정전횟수) 기준 ±4건 이내에 포함된 예측된 정전횟수는 366개 중 351개로 95.62%를 차지하여 정전 예측모형의 성능이 상당히 높은 것을 확인할 수 있었다. (실제 정전횟수 - 예측 정전횟수) 의 오차가 -4건을 초과하는 부분은 366개 중 11개이며 3.00%로, 실제 발생한 정전횟수 대비 예측한 정전횟수가 많은 경우라고 볼 수 있다. 이는 전기사업자가 태풍 등 기상예보를 반영하여 재난상황에 선제적으로 대비 한 결과이며, 전기사업자 측면에서 인력과 장비가 충분하게 확보되어 실제 정전을 원활하게 대비할 수 있으므로 큰 문제가 되지 않는다.

그러나. (실제 정전횟수 - 예측 정전횟수)가 +4건을 초과하는 부분은 366개 중 4개이며 1.09%로, 예측 정전횟수 대비 실제 정전횟수가 많은 결과이다. 이는 예측한 정전횟수를 기반으로 인력과 장비를 확보할 경우, 인력과 장비가 충분히 확보되지 않아 현장에서 큰 혼란을 야기할 수 있다. 즉, (실제 정전횟수 - 예측 정전횟수)가 +4건을 초과하는 경우에는 전기사업자가 실제 정전에 적절하게 대응하기 어려울 수 있으며, 도출된 정전횟수 예측모형의 한계점을 엿볼 수 있는 경우이다.

향후 연구 과제로는 부산, 울산지역에 한정하지 않고, 전국적으로 권역별 기상요인을 파악하여 각 권역의 기상특징에 적합한 정전횟수 예측모형의 도출이 필요하다고 생각한다.

주요어 : 배전계통, 정전데이터, 정전횟수, 기상요인, 회귀분석
학 번 : 2020-26551

목 차

제 1 장 서론 ····· 1

제 1 절 연구의 배경 및 필요성 ····· 1

제 2 절 연구의 방법 및 범위 ····· 3

제 3 절 연구의 구성 및 개요 ····· 6

제 2 장 이론적 배경과 선행연구 고찰 ····· 8

제 1 절 현황 및 배전계통 분석 ····· 8

1. 배전설비 현황 ····· 8

2. 배전계통 전기품질 ····· 10

3. 배전계통 공급신뢰도 ····· 12

제 2 절 정전의 종류 및 분석 ····· 14

제 3 절 선행연구 분석 ····· 16

제 3 장 연구의 설계 및 방법 ····· 19

제 1 절 데이터 수집 및 분석 ····· 19

제 2 절 정전예측모형 설계 ····· 23

1. Pearson 상관분석 ····· 23

2. 다중 선형 회귀분석 ····· 30

2. 다항 비선형 회귀분석 ····· 34

제 4 장 연구결과 ····· 41
제 1 절 정전예측 모형 결과 도출 ····· 41
제 2 절 정전예측 모형 검증 ····· 42
1. 2015~2019년 데이터 기반 정전모형 검증 ····· 42
2. 2020년 데이터 기반 정전모형 검증 ····· 48

제 5 장 결론 ····· 53
제 1 절 연구의 결과 ····· 53
제 2 절 한계점 및 향후 연구방안 ····· 55

참고문헌 ····· 56
Abstract ····· 58

표 목 차

[표 1-1] 한국전력공사 본부별 배전설비 회선수 현황 ····· 4
[표 2-1] 부산울산지역본부 배전설비 현황 ····· 9
[표 2-2] 전기품질(Power Quality) 관리항목 ····· 11
[표 2-3] 공급신뢰도(Reliability) 관리지표 ····· 13
[표 2-4] 배전계통 정전(Interruption) 분류체계 ····· 15
[표 2-5] 선행연구와 본연구의 비교표 ····· 18
[표 3-1] 부산울산지역본부 지사별 정전내역 ····· 20
[표 3-2] 부산울산지역본부 정전원인별 정전내역 ····· 21
[표 3-3] 부산울산지역본부 정전종류별 정전내역 ····· 22
[표 3-4] “강수” 관련 기술통계분석 결과 ····· 23
[표 3-5] “풍속” 관련 기술통계분석 결과 ····· 24
[표 3-6] 강풍주의보 와 강풍경보 기준 ····· 24
[표 3-7] “정전횟수” 관련 기술통계분석 결과 ····· 25
[표 3-8] “강수” 관련 Pearson 상관분석 결과 ····· 26
[표 3-9] “풍속” 관련 Pearson 상관분석 결과 ····· 27
[표 3-10] 21개 기상요인 Pearson 상관분석 결과 ····· 28
[표 3-11] 다중 선형회귀 모형의 설명력 ····· 32
[표 3-12] 다중 선형회귀 모형의 ANOVA ····· 32
[표 3-13] 다중 선형회귀 모형의 계수 ····· 33

[표 3-14] 다항 비선형 회귀분석 결과 ····· 37
[표 4-1] 다중 선형회귀 및 다항 비선형 회귀 결과 ····· 41
[표 4-2] 2015~2019년 정전모형 검증결과 ····· 44
[표 4-3] 2020년 정전모형 검증결과 ····· 49

그 림 목 차

[그림 1-1] 2020년 제10호 태풍「하이선」경로 ····· 3
[그림 2-1] 한국전력공사 부산울산지역본부 관할구역 ····· 8
[그림 3-1] 과소적합, 최적적합, 과대적합 비교 ····· 35
[그림 3-2] MATLAB 다항 비선형 회귀 모형의 시각화 ·· 38
[그림 3-3] MATLAB 다항 비선형 회귀모형의 잔차 ····· 39
[그림 3-4] MATLAB 다항 비선형 회귀모형의 등고선 ···· 39
[그림 4-1] 2015~2019년 정전모형 검증그래프 ····· 42
[그림 4-2] 2019년 제17호 태풍「타파」개요 ····· 46
[그림 4-3] 2020년 정전모형 검증그래프 ····· 48
[그림 4-4] 2020년 제10호 태풍「하이선」개요 ····· 50
[그림 4-5] 태풍 “하이선” 대비 공문 ····· 51

제 1 장 서론

제 1 절 연구의 배경 및 필요성

현대의 도시들은 다양한 사회 시스템과 서로 긴밀하게 연결되어 있다. 따라서 현대의 도시에 재난이 발생한다면, 그 피해는 각 사회 시스템에 연쇄적으로 나타나게 될 것이다. 또한, 도시에 발생하는 여러 재난들은 발생 시기 빛 범위를 정확하게 예측하기 어렵기 때문에 재난에 완벽하게 대응하는 것은 불가능하다. 이러한 특징을 가지고 있는 재난에 대응하기 위해 시스템을 유지, 관리하는 것은 매우 중요하다. [1]

특히 이러한 재난 상황에서 에너지 시스템인 전력시스템이 아주 중요하다. 지구온난화와 극한 기후 일수의 증가는 냉난방기기 사용의 증가를 야기하고, 이는 전력사용량의 증가를 초래한다. 또한 자연재해로 인한 송전, 배전설비 등의 파괴는 전력공급의 중단을 초래한다. 그리고 이러한 전력시스템의 수요와 공급에 대한 불균형은 정전으로 표출된다. [1]

이렇듯 기후변화는 전력시스템에 막대한 피해를 주는 재난으로 분류될 수 있고, 기후변화가 가속화됨에 따라 그 피해는 더욱 커질 것으로 예상된다. 또한 이러한 전력시스템에서 발생하는 피해는 전력시스템 뿐만 아니라 이와 연결되어 있는 타 에너지 시스템과 도시시스템 등 또 다른 시스템에 악영향을 줄 수도 있다. 따라서 기후변화로부터 발생한 전력시스템의 피해를 해결할 수 있는 방안 마련이 필요한 시점이다. 또한 폭우, 태풍, 폭염, 한파와 같은 기후변화의 경우 다른 기타 재난과 마찬가지로 발생시기와 범위를 예측하기 힘들기 때문에 더욱 정교한 예측모델을 필요로 한다. 하지만 현재까지 이러한 전력시스템 피해에 적극적으로 대응하고, 피해를 최소화시킬 수 있는 정교한 예측모델에 대한 연구는 미흡

한 상태이다. [1]

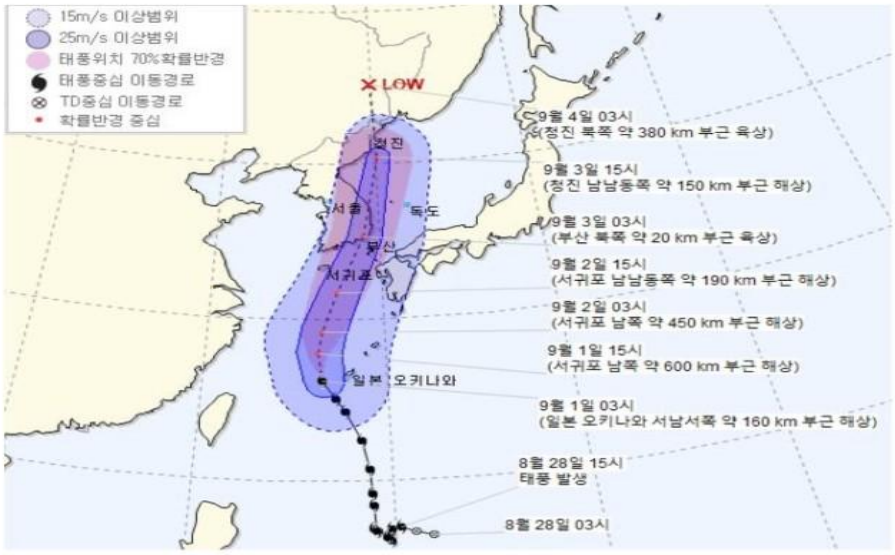
본 연구는 Decision tree 분석 방법과 Logistic 회귀분석 방법을 통하여 정전 예측모형을 도출한 선행연구와는 달리, 다중 선형 회귀분석과 다항 비선형 회귀분석을 통하여 정전예측 모형을 모델링 하였다. 그리고 20개 이상의 다수의 독립변수를 사용하여 다중공선성 등의 발생을 야기할 수 있는 선행연구와는 달리, 고 영향군의 2개 기상요인을 독립변수로 선정하여 모형의 신뢰도를 높였다.

또한, 배전설비의 정전확률을 예측하는 기존의 선행연구와는 달리, 배전설비의 정전횟수를 예측하는 모델을 도출하여 전기사업자 측면에서 실제 전력산업 현장에 즉시 활용이 가능하도록 했으며, 마지막으로 일부 배전설비에 국한하지 않고, 정전이 발생할 수 있는 모든 배전설비에 대한 정전횟수 예측모델을 설계 및 검증하였다.

이러한 맥락하에 본 연구의 목적은 전력시스템의 피해결과물인 정전에 영향을 미치는 고 영향 기상요인을 파악하고, 이를 통하여 배전설비의 정전횟수를 예측하는 최적의 정전예측 모형을 도출하는 것이다. 배전설비 정전횟수 예측모형은 전기사업자 측면에서는 정전대비 인력 및 장비 확보에 반영할 수 있으며, 대국민 상대로는 재난문자 등과 연동하여 사회적 재난에 대비할 수 있다.

제 2 절 연구의 방법 및 범위

본 연구의 시간적 범위는 정전예측모형 도출을 위하여 2015년~2019년 5개년으로 설정하였고, 정전예측모형을 검증하기 위하여 2020년도를 설정하였다. 또한 연구의 공간적 범위는 전국이 아닌 부산, 울산지역으로 한정하였다. 그 이유는 아래의 [그림1-1] 과 같이 부산, 울산지역은 태풍의 경로에 위치하여 기상요인과의 연관성이 뚜렷할 것으로 예상되기 때문이다.



[그림1-1] 2020년 제10호 태풍 “하이선” 경로 [2]

또한, 본 연구를 부산, 울산으로 한정한 이유는 선행연구의 고 영향 변수인 3개 기상요인 강수, 풍속, 적설 중에서 적설이 5년간 단 2회 발생하여 낮은 적설 빈도로 인해 강수와 풍속에 집중하여 배전설비 정전횟수와 연관관계를 파악할 수 있으며, [표1-1] 과 같이 한국전력공사의 15개 본부 중 서울, 경기 등 수도권 다음으로 배전설비 회선수가 많아, 타 본부 대비 배전설비의 정전이 많이 발생하기 때문이다.

[표1-1] 한국전력공사 본부별 배전설비 회선수 현황('21.6월 기준) [3]

본부명	회선수		
	가공	지중	가공+지중
서울본부	485	272	757
남서울본부	465	449	914
인천본부	596	300	896
경기북부본부	485	164	649
경기본부	1,256	413	1,669
강원본부	393	38	431
충북본부	405	74	479
대전세종충남본부	752	274	1,026
전북본부	398	118	516
광주전남본부	583	208	791
대구본부	879	116	995
경북본부	267	17	284
부산울산본부	907	243	1,150
경남본부	537	118	655
제주본부	121	25	146
합계 (15개)	8,529	2,829	11,358

데이터의 출처는 기상 데이터의 경우, 기상청의 기상자료개방포털에서 종관기상관측(ASOS)의 2015~2020년 6개년 간 자료를 추출하여 기상요인 자료로 활용했으며, 정전데이터의 경우는 한국전력공사 부산울산지역본부에 속하는 13개 지사의 2015~2020년 6개년간의 배전설비 일일 정전횟수 데이터를 사용하였다.

먼저 관련문헌 연구를 통해 배전계통의 전기품질과 공급신뢰도에 대하여 정의하였고, 기상청 자료를 활용하여 약 30여개 기상요인에 대한 데이터를 구축하였다. 그리고 선행연구를 고려하여 배전설비 정전에 큰 영향을 미치는 풍속, 강수, 적설을 기상요인으로 선별하였고, 부산, 울산지역의 지역적 특성을 고려하여 3개 기상요인 중에 적설을 제외한 풍속, 강수를 최종 기상요인을 확정하였다. 확정된 2개 기상요인 중 Pearson 상관관계 분석을 통하여 상관관계가 높은 최대순간풍속과 일 강수량을 최종 독립변수로 확정하였다. 또한, 종속변수를 정전횟수로 설정하고 다중 선형 회귀분석과 다항 비선형 회귀분석을 통하여 모델링을 진행하였다.

그리하여 본 연구는 기상요인에 따른 정전확률을 예측하는 기존의 선행연구와는 달리 기상요인에 따른 정전횟수를 예측하는 모델을 도출하여 전기사업자 측면에서 실제 전력산업 현장에 즉시 활용이 가능하도록 했으며, 일부 배전설비에 국한하지 않고, 정전이 발생할 수 있는 모든 배전설비에 대해 정전횟수 예측모델을 설계 및 검증을 하였다. 또한, 다수의 독립변수를 활용한 선행연구와는 다르게 2개의 독립변수에 한정하여 모델링을 하여 예측모형의 신뢰도를 높였다.

제 3 절 연구의 구성과 개요

본 연구의 제 1장에서는 전력계통의 정전에 적극적으로 대응하고, 정전 피해를 최소화시킬 수 있는 정교한 예측모델의 필요성과 본 논문의 전체적인 구성에 대해 살펴보았다.

제 2장은 부산, 울산지역의 일반현황 및 배전설비 현황에 대해 소개를 했으며, 배전계통의 전기품질과 공급신뢰도에 관한 이론적 배경과 “오도은·김현수(2017)은 기상기후 기반 배전 개폐기 고장예측에 관한 연구” [9], “임청원·한상국(2018)은 빅데이터 기반 전력계통 고장예측시스템 구축을 위한 기상요인 및 배전설비 고장에 관한 연구”[10] 2건의 선행연구 분석을 통하여 본 연구와의 차이점과 비교하였다.

제 3장은 연구의 설계 및 방법으로 기상 데이터는 기상청의 기상자료개방포털에서 종관기상관측(ASOS)자료를 사용하였고, 정전데이터는 한국전력공사 부산울산지역본부의 3,969건의 정전데이터를 1,826개(365일*5개년) 날짜에 매칭하여 사용하였다. 또한, 데이터 분석에 앞서 Excel 기술통계분석을 수행하여 각각의 변수의 특성에 대하여 간략하게 파악하였고, IBM SPSS Statistics 26을 통하여 Pearson 상관분석, 다중 선형 회귀분석을 수행하였다. 그러나 다중 선형 회귀분석으로는 과소적합 모형이 도출되어, MATLAB R2021a의 Curve Fitting Toolbox 기능으로 다항 비선형 회귀분석을 수행하여 일 강수량과 최대순간풍속에 따라 정전횟수를 예측하는 최적의 모형을 설계하였다.

제 4장은 제 3장에서 도출된 정전횟수 예측모형 검증을 위하여 1차로 원본 데이터인 2015~2019년 5개년 기상 데이터를 기반으로 예측한 정전횟수와 실제 정전횟수를 비교하여 검증했으며, 2차로 2020년의 기상 데이터를 기반으로 예측한 정전횟수와 실제 정전횟수를 비교하여 정전모

형 검증을 수행하였다.

그리고 마지막으로 제5장에서는 본 연구의 결과와 도출된 정전횟수 예측모형의 한계점 및 향후 연구 방향에 대하여 언급하였다.

제 2 장 이론적 배경과 선행연구 고찰

제 1 절 현황 및 배전계통 분석

1. 배전설비 현황

본 연구는 부산, 울산지역에 한정하여 연구를 수행하였다. 그 이유는 첫째, 부산, 울산지역은 태풍의 경로에 위치하여 날씨와 연관성이 뚜렷할 것으로 예상되며, 둘째, 선행연구의 고 영향 변수인 3개 기상요인 강수, 풍속, 적설 중에서 적설이 5년간 단 2회 발생하여 낮은 적설빈도로 인해 강수와 풍속에 집중하여 배전설비 정전횟수와 연관관계 파악할 수 있으며, 셋째, [표1-1]의 한국전력공사의 15개 본부 중 수도권 다음으로 배전설비 회선수가 많아, 타 지역본부 대비 배전설비의 정전이 많이 발생하기 때문이다.

한국전력공사 부산울산지역본부는 행정구역상으로는 부산광역시(15개구, 1개군), 울산광역시(4개구, 1개군), 경남(김해시, 양산시) 를 관할하며 인구는 부산 344만명, 울산 115만명, 김해 56만명, 양산 36만명, 총 551만명로 전국대비 10.5% 수준이며, 총 면적은 2,779㎢으로 전국대비 2.8%를 차지한다. 관할 구역은 [그림2-1]과 같다. [4]

[그림2-1] 한국전력공사 부산울산지역본부 관할구역 [4]

또한, 고객호수는 209만 3천호(2019. 12. 31 기준)로 전사대비 8.7%이며, 판매량은 62,481Gwh으로 전사대비 12.3% 수준이다. 배전설비는 회선수 1,150개, 전선긍장 12,493C-km으로 전사대비 5.1% 수준이며, 부산울산지역본부에 속하는 13개 지사의 주요 배전설비 현황은 [표2-1]과 같다. [4]

[표2-1] 부산울산지역본부 배전설비 현황 [4]

지사명	회선수	전선긍장(c-m)			지지물(기)	변압기	
		고압	저압	소 계		대수	용량(kVA)
직할	65	448,789	302,465	751,254	13,651	9,561	745,585
울산지사	165	1,863,880	1,468,326	3,332,206	59,916	22,246	1,361,727
김해지사	167	2,222,330	1,490,672	3,713,002	70,637	22,963	1,207,290
동래지사	99	753,027	460,254	1,213,281	22,077	14,324	996,610
남부산지사	128	1,012,279	385,722	1,398,001	17,566	12,049	936,378
양산지사	96	1,296,014	1,069,764	2,365,778	41,603	12,875	719,901
중부산지사	102	642,357	359,563	1,001,920	16,497	11,256	842,072
북부산지사	99	610,763	300,591	911,354	17,110	11,194	770,435
동울산지사	50	757,246	550,754	1,308,000	23,690	8,175	512,574
서부산지사	93	1,201,097	642,270	1,843,367	28,888	9,255	565,045
기장지사	34	722,608	614,986	1,337,594	24,424	6,906	347,717
서울산지사	36	808,732	730,449	1,539,181	29,494	7,345	324,458
영도지사	16	154,455	106,044	260,499	4,981	2,810	204,145
총계(13개)	1,150	12,493,577	8,481,860	20,975,437	370,534	150,959	9,533,937

2. 배전계통 전기품질

배전계통의 성능에 대한 지표, 즉 공급서비스의 평가기준은 크게 공급신뢰도(Reliability) 및 전기품질(Power Quality)로 구분되며, 일반적으로 공급신뢰도는 1분 이상의 정전(Interruption) 현상을 규정하고, 전기품질(Power Quality)은 1분 미만의 정전 및 미세 전압변동에 대하여 규정한다.

IEC 기준에 의하면 전기품질을 '통상의 운전 상태에서 고객에게 공급되는 전력의 특성을 정의하는 파라미터의 집합으로 공급의 연속성과 전압의 특성에 관한 것' 이라고 정의하고 있으며, [5] IEEE 기준에 의하면 '전력설비의 동작과 구내 설비계통 및 다른 접속기기에 적합하도록 하는 민감한 설비의 전력공급과 접지의 개념'으로 정의하고 있다. [6] 또한, 한국전력공사의 배전계통 공급신뢰도 관리기준에서는 전기품질이란 '전력설비에 영향을 주는 시스템과 공급전압의 특성을 말하며 전압, 전류, 주파수 이외에 순간 전압변동, 고조파, 플리커 등 전력시스템에서 발생하는 광범위한 전자기적 외란현상'을 의미한다. [7]

세계적으로 대부분의 전력회사는 전압, 정전시간, 주파수를 전기품질의 3대 관리요소로 설정하여 운영하고 있으며, 한국의 경우 정전시간 및 규정전압 유지율은 한전의 배전계통에서 관리하고 있고, 주파수는 한국전력거래소에서 일괄 관리하고 있다. 전기품질에 영향을 미치는 외란은 크게 '전압' 에 관련된 것과 '파형(Waveform)의 왜곡'에 관련된 것으로 구분할 수 있다. 이중 전압에 관련된 것은 그 크기와 지속시간에 따라 순간전압강하, 순간전압상승, 순간정전 등으로 나누어진다. IEEE 1159에서는 전력품질 요소를 과도특성(Transient characteristics), 단주기 변동(Short duration variations), 장주기 변동(Long duration variations) 및

파형왜곡(Waveform distortion) 등의 4가지 주요 특성과 전압 불평형(Voltage imbalance), 전압 변동(Voltage fluctuations), 전력 주파수 변동(Power frequency variations) 등으로 표준화하여 [표2-2] 와 같이 분류하고 있다. [6]

[표2-2] 전기품질(Power Quality) 관리항목 [6]

용어(Terms)	정의(Definition)	지속시간	크기(P.U.)
단시간 전압변동 (ShortDuration Variation)	Instantaneous:0.5～30cycles Momentary:30cycles～3초 Temporary : 3초~1분 ※ Sustained : 1분 이상	0.5Cycle～1분	0.1 미만
순간전압강하 (VoltageSag)	교류전압 실효치의 순간적 감소 현상	0.5cycle～1분	0.1～0.9
순간전압상승 (VoltageSwell)	교류전압 실효치의 순간적 증가 현상	0.5cycle～1분	1.1～1.8
전압 불평형 (Imbalance)	3상전압, 전류의 평균에서 최대편차값을 평균값으로 나눈 것	정상상태	
고조파 (Harmonics)	정현파전압, 전류 기본주파수의 정수배의 주파수	정상상태	0～20%
플리커 (Flicker)	부하전류의 크기가 변동을 나타낼 때 전 등의 밝기가 전압동요로 깜박이는 것	간헐적임	0.1～7%

3. 배전계통 공급신뢰도

공급신뢰도란 '배전계통의 성능 및 공급서비스의 평가기준이며 계통의 시스템과 고객 Needs를 경제적, 안정적으로 관리하기 위한 지수로 예정된 기간 동안 예상되는 운전 상태에서 전기설비가 적절한 성능을 발휘할 수 있는 확률'을 의미한다. [7]

전력계통에 있어서 배전계통은 고객에게 직접 전력을 공급하는 특성을 가지므로 배전계통에서 발생하는 각종 외란은 수용가에게 직접적인 영향을 미친다. 이에 따라 세계 각국의 전력회사(기관)은 전기품질 및 공급신뢰도의 효율적인 관리를 위하여 다양한 평가지수를 연구개발 하였다. 전기품질에 대한 관심도가 낮았던 과거에는 일반적으로 SAIFI(System Average Interruption Frequency Index), SAIDI(System Average Interruption Duration Index), CAIDI(Customer Average Interruption Duration Index) 등 일시정전(Sustained Interruption)과 관련된 지수를 중심으로 관리하여 왔으며, 일시정전과 순간정전(Temporary Interruption)의 구분은 우리나라를 비롯하여 대부분의 국가에서는 5분을 기준으로 하고 있으며 IEEE가 제안한 대표적인 공급신뢰도 지수는 [표 2-3] 와 같다. [8]

[표2-3] 공급신뢰도(Reliability) 관리지표 [8]

용어(Terms)	정의(Definition)	적용분야
SAIDI (호당정전시간)	System Average Interruption Duration Index $SAIDI = \frac{\text{연정전호수의 정전시간}}{\text{전체 고객호수}}$	고객의 호당정전시간을 관리하는 가장 보편적 지수
SAIFI (호당정전횟수)	System Average Interruption Frequency Index $SAIFI = \frac{\text{연정전호수}}{\text{전체 고객호수}}$	일정지역내의 고객 정전빈도 관리용
CAIDI (정전고객 평균 정전시간)	Customer Average Interruption Duration Index $CAIDI = \frac{\text{연정전호수의 정전시간}}{\text{연 정전호수}}$	고객당 평균 고장복구 소요시간 관리
MAIFI (호당순간정전 횟수)	Momentary Average Interruption Frequency Index $MAIFI = \frac{\text{연순간정전호수}}{\text{전체 고객호수}}$	일정기간내의 고객정전 빈도 관리
ASAI (공급 가용율)	Average Service Availability Index $ASAI = \frac{\text{고객 공급 유효시간}}{\text{고객 공급 수요시간}}$	고객의 연간 전기공급 지속율 관리를 위한 지수

제 2 절 정전의 종류 및 분석

정전의 정의는 전력계통 운전 중 외적요인 또는 예방정비 등을 위하여 전력공급이 중단된 상태를 의미한다. 계통구분에 따라서 전원 측 정전은 발전소, 송전설비, 변전설비 고장에 따른 인출 배전선로의 정전을 의미하며, 배전 측 정전은 고압정전과 저압정전으로 나누어진다. 고압정전은 22.9kV 특고압 배전선로(6.6kV 포함)의 고장으로 인한 정전을 의미하며, 저압정전은 변압기 2차 측 220/380V(이하 저압) 저압선로 및 인입선 등으로 인한 정전을 의미한다. [7]

정전의 특성으로 구분 시, 불시정전은 배전계통의 고장, 사고 또는 재해 등으로 인한 정전으로서 계통 운영자가 사전에 예측할 수 없는 모든 정전을 의미하며, 계획정전은 배전설비의 유지보수 및 신증설 공사를 시행하기 위하여 사전에 설정된 계획에 따라 이루어지는 정전을 의미한다. [7]

마지막으로 시간으로 구분 시, 일시정전은 설비결함으로 기능에 이상이 생기거나 또는 외부영향으로 인하여 선로의 전력공급이 중단되어 고객(저・고압 포함)의 정전 지속시간이 5분 이상인 정전을 의미하며, 순간정전은 설비결함으로 기능에 이상이 생기거나 또는 외부영향으로 인하여 선로의 전력공급이 중단되어 고객 (저・고압 포함)의 정전 지속시간이 1분 이상 5분 미만인 정전을 의미한다. 또한 순시정전은 설비결함으로 기능에 이상이 생기거나 또는 외부영향으로 인하여 선로의 전력공급이 중단되어 고객 (저・고압 포함)의 정전 지속시간이 1분 미만인 정전을 의미한다. [7]

현행 전기사업자인 한전은 5분 이상의 정전을 일시정전으로 구분하고 있으나, 유럽의 경우 3분을 기준으로 일시/순간고장을 구분하고 있으며

미국의 경우에는 전력회사별로 상이하나 대부분 3~5분을 기준으로(일부 전력회사의 경우 1, 2분으로 구분) 관리하고 있다. 또한, 현재의 정전분류는 배전계통의 정전원인별 분류체계를 명확화·체계화 할 필요가 있어 정전시간 구분을 기존의 일시, 순간정전에서 [표2-4]와 같이 일시, 순간, 순시정전으로 세분화하고 있다. [7]

[표2-4] 배전계통 정전(Interruption) 분류체계 [7]

계통구분	전압구분	특성구분	시간구분
전원측 정전 (Source Side)			
배전측 정전 (Load Side)	고압정전 (Primary)	불시정전 (Unplanned)	일시정전(Sustained)
			순간정전(Temporary)
	저압정전 (Secondary)	계획정전 (Planned)	순시정전(Momentary)

본 연구에서는 배전측 정전의 고압정전을 대상으로 하였고, 정전의 시간에 따른 구분은 별도로 하지 않고, 모두 동일한 1건의 정전으로 간주하였다. 그리하여 부산, 울산지역의 총 정전건수는 3,969건으로 분류하여 연구를 진행하였다.

제 3 절 선행연구 분석

오도은·김현수(2017)은 기상기후 기반 배전 개폐기 고장예측에 관한 연구에서 배전설비 중 중요도가 높다고 판단되는 배전 개폐기의 고장원인을 분석하고, 기상기후의 변화가 배전 개폐기에 영향을 미치는 것이 무엇인지를 파악하여 데이터마이닝 기법을 이용하여 모델링 하였다. Decision tree 분석 방법과 Logistic Regression 분석 방법을 이용하여 분석한 결과 23개의 기상요인 중 최대풍속과 최대순간풍속 변화량이 배전 개폐기 고장에 가장 큰 영향을 주고 있음을 확인하였다. [9]

위 연구의 아쉬운 점은 독립변수로 기상요인 23개를 사용하였다. 다수의 독립변수는 변수 간의 변별력을 떨어뜨리며 다중공선성 등의 문제를 야기할 수 있어 통계적인 해석의 의미를 잃게 될 수 있다. 또한, 일반적으로 회귀식에 독립변수가 추가됨에 따라 R^2이 상승하는 경향이 있어, 독립변수가 23개인 정전모형의 표면적인 R^2은 높을 수 있으나, 실질적인 모형의 신뢰도가 떨어질 수 있다. 본 연구와의 큰 차이점은 위 선행연구는 배전설비 중 가공 개폐기에 한정하여 정전확률에 대해서 연구하였으며, 전체 배전설비의 정전횟수를 예측한 본 연구와는 차이점이 존재한다.

임청원·한상국(2018)은 빅데이터 기반 전력계통 고장예측시스템 구축을 위한 기상요인 및 배전설비 고장에 관한 연구에서 배전설비 중 외부에 노출되어 기후변화와 같은 환경적인 요인의 영향이 높으며 고장빈도도 상대적으로 높은 배전설비인 COS, 애자, 피뢰기의 고장원인을 각각 분석하고 기후의 변화가 각 설비에 영향을 미치는 요소가 무엇인지를 데이터마이닝 기법을 이용하여 모델링 하였다. Decision tree 분석방법과 Logistic 회귀분석 방법을 이용하여 분석한 결과 22개의 기상요인 관련 독립변수 중 풍속, 강수 및 적설과 관련된 기상요인이 배전설비 고장에

많은 영향을 주고 있음을 확인하였다. [10]

위 연구의 아쉬운 점은 앞선 선행연구와 마찬가지로 다수의 독립변수(22개)를 사용하여, 다수의 독립변수 간 변별력과 다중공선성 등의 문제를 야기할 수 있어 통계적인 해석의 의미를 잃게 된다. 본 연구와의 큰 차이점은 배전설비 중 비교적 고장빈도가 높은 설비인 COS, 애자, 피뢰기에 한정하여 정전확률에 대한 연구를 하여 전체 배전설비의 정전횟수를 예측한 본 연구와는 차이점이 존재한다.

본 연구는 2건의 선행연구의 검토를 통하여 적절한 독립변수를 선별하였다. 위의 선행연구에서 고 영향 요인으로 전자는 풍속, 후자는 풍속, 강수, 적설을 들고 있다. 본 연구는 최초에 풍속, 강수, 적설을 고려하였지만, 부산, 울산지역의 지역적인 특성을 추가적으로 고려하였다. 즉, 본 연구에서는 부산, 울산지역의 날씨 특성상 2015~2019년도 적설빈도가 단 2회(1.3cm[2018. 1. 10], 0.2cm[2019 1. 31]) 이므로 통계 관측수의 부족으로 유의확률이 산출되지 않아 통계적인 해석이 불가하여, 적설은 독립변수에서 제외하고 고 영향 기상요인인 풍속과 강수를 선별하여 연구를 진행했다.

또한 2건의 선행연구 모두 Decision tree 분석 방법과 Logistic 회귀분석 방법을 통하여 비교적 고장 빈도가 높은 배전설비의 정전확률에 대한 예측모형을 도출하였으나, 본 연구에서는 다중 선형 회귀분석과 다항 비선형 회귀분석을 통하여 부산, 울산지역의 전체 배전설비의 정전횟수를 예측하는 모형을 모델링 하였다.

그러므로 본 연구에서 고 영향군으로 분류된 소수(2개)의 독립변수를 활용하여 정전횟수 예측모델을 모델링 하였다. 또한 특정 배전설비별로 정전확률을 예측하기보다 전체 배전설비의 정전횟수를 예측하여 전기사

업자 측면에서 실제 전력산업 현장에 즉시 활용 가능하도록 하였다. [표 2-5] 은 선행연구 2건과 본 연구의 방법론, 변수 등을 비교 정리한 것이다.

[표2-5] 선행연구와 본연구의 비교표

구 분	선행연구1 (오도은,김현수)	선행연구2 (임청원,한상국)	본 연구
방법론	-Decision Tree -Logistic 회귀분석	-Decision Tree -Logistic 회귀분석	- 다중 선형 회귀분석 - 다항 비선형 회귀분석
독립변수	기상요인 23개	기상요인 22개	기상요인 2개 (풍속, 강수)
종속변수	개폐기 정전확률(%)	COS, 애자, 피뢰기 정전확률(%)	전체설비 정전횟수(건)
기상요인 분석결과 (고영향)	풍속	풍속, 강수, 적설	-

제 3 장 연구의 설계 및 방법

제 1 절 데이터 수집 및 분석

본 연구는 IBM SPSS Statistics 26와 MATLAB R2021a 2가지 시뮬레이션 Tool을 활용하여 배전설비 정전횟수에 대한 예측모형을 설계하였다. IBM SPSS Statistics 26을 활용하여 Pearson 상관관계 분석과 다중선형 회귀분석을 시행하여 선형 회귀모형을 모델링 했으나, 과소적합(Under-fitting) 형태의 회귀모형이 도출되어, MATLAB R2021a의 Curve Fitting Toolbox 기능으로 다항 회귀분석을 시행하여 최적적합(Optimal-fitting)의 다항 비선형 회귀모형을 모델링 하였다.

본 연구의 독립변수는 기상청의 기상자료개방포털에서 종관기상관측(ASOS)자료를 추출하여 사용했다. 종관기상관측이란 정해진 시각의 대기 상태를 파악하기 위해 모든 관측소에서 같은 시가에 실시하는 지상관측을 의미한다. 시정, 구름, 증발량, 일기현상 등 일부 목측 요소를 제외하고 종관기상관측장비(ASOS, Automated Synoptic Observing System)를 이용해 자동으로 기상정보를 관측하고 그 결과로 얻어진 데이터이며, 종관기상관측장비는 [그림3-1] 과 같다. 지상 종관기상관측 자료 중 2015~2019년도 5개년 간 부산지역 일일 기상정보 사용하였다. [11]

[그림3-1] 종관기상관측장비(ASOS) [11]

종속변수는 한국전력공사의 부산울산지역본부에 소속된 13개 지사의 2015~2019년간 5개년 간 일일 정전횟수 데이터를 사용하였으며, 5개년 간 총 3,969건의 정전이 발생하였다. 13개 지사별로 정전내역을 분류하면 [표3-1] 와 같다. [12]

[표3-1] 부산울산지역본부 지사별 정전내역 [12]

지사명	정전횟수
기장지사	189
김해지사	842
남부산지사	337
동래지사	263
동울산지사	168
부산울산본부직할	221
북부산지사	280
서부산지사	334
서울산지사	186
양산지사	256
영도지사	77
울산지사	495
중부산지사	321
총계(13)	3,969

김해지사는 정전횟수가 842건으로 가장 많고, 영도지사는 77건으로 가장 적다. 이런 차이는 [표2-1]를 참고하면, 김해지사의 배전선로 회선수는 167개이고 변압기는 22,963대인 반면, 영도지사는 배전선로 회선수는 16개이고 변압기는 2,810대이므로 배전설비 규모에서 약 10배 차이가 발생하여 설비 규모의 차이에서 나온 결과로 보여진다. 또한, 13개 지사의 2015~2019년 5개년 평균 정전횟수는 305.3건이다.

[표3-2] 부산울산지역본부 정전원인별 정전내역 [12]

정전원인	정전횟수
조류접촉	1,043
자연열화	668
자연현상/풍우해	650
원인불명	389
일반인과실	252
이물접촉	248
기타	132
수목접촉	122
동일원인 동시고장	104
보수불량	98
자연현상/뇌해	73
부식	47
제작불량	42
시공불량	30
작업자과실	19
침수	17
화재	15
과부하	12
오동작	4
타사업소 공용선로 고장	2
자연현상/빙설해	1
자연현상/염진해	1
총계	3,969

[표3-2]는 정전원인별 정전내역이다. 조류접촉이 1,043건으로 가장 많았고, 빙설해와 염진해가 각각 1건으로 가장 적었다. 조류접촉은 까치 등 조류가 배전설비에 둥지를 짓거나 전력설비에 접촉하여 발생하며, 전기사업자는 조류둥지 순시 및 철거, 조류포획 등을 통해 정전예방 활동을 시행하고 있다. 총 3,969건 중 정전원인과 기상요인 간 직접적인 관련성(자연현상/풍우해, 이물접촉 등)이 1,795건으로 45.2%이 차지하고, 간접적인 관련성(조류접촉, 원인불명)이 2,174건으로 54.8% 차지하는 것으로 가정하고 본 연구를 진행하였다.

[표3-3] 부산울산지역본부 정전종류별 정전내역 [12]

정전종류	정전횟수
순시정전(1분미만)	1,515
순간정전(1분~5분)	1,960
일시정전(5분이상)	494
총계	3,969

[표3-3] 는 정전시간에 따른 정전종류를 분류한 내역이다. 1분 미만의 순시정전이 1,515건으로 38.1%이며, 1분~5분 사이의 순간정전은 1,960건으로 49.4%를 차지하였다. 순시정전과 순간정전이 총 3,475건으로 87.5%를 차지했으며, 5분 이상의 일시정전은 494건으로 11.9%를 차지하여 비교적 적은 부분을 차지했다.

[표3-1], [표3-2], [표3-3]의 3,969건의 정전내역은 2015~2019년 5년간 발생했던 총 정전횟수 이며, 이를 정전이 발생한 날짜 단위로 매칭하여 총 3,969건의 정전데이터가 1,826개(365일*5개년) 날짜에 매칭하여 연구를 진행하였다.

제 2 절 연구모형 설계

1. Pearson 상관분석

본 연구는 상관관계 방법론으로 가장 보편적인 방법론인 Pearson 상관관계 분석을 사용했으며, 고 영향 기상요인인 강수와 풍속에 속하는 다양한 세부 기상요인들과 종속변수인 일일 정전횟수의 상관관계를 분석하였다. 그 결과, 강수는 일 강수량을 풍속은 최대순간풍속을 최종 독립변수로 확정하였다.

Pearson 상관계수는 -1부터 +1사이의 값을 갖으며, 두변수의 상관관계가 positive 또는 negative의 방향으로 클수록 +1 또는 -1에 가깝게 나타난다. 반면에 두 변수 간에 상관관계가 전혀 없으면 상관계수는 0이다. [13]

Pearson 상관관계 분석에 앞서, 강수와 풍속에 속하는 다양한 세부 기상요인과 2015~2019년 5개년 간 일일 정전횟수에 대한 각각의 변수특성에 대하여 간략하게 파악하기 위하여 Excel 기술통계분석을 수행했으며, 강수관련 기술통계분석 결과는 [표3-4]와 같다.

[표3-4] “강수” 관련 기술통계분석 결과

구분	강수 계속 시간(hr)	10분 최다 강수량(mm)	1시간 최다 강수량(mm)	일 강수량 (mm)
평균	6.041	1.828	5.123	3.883
표준 편차	5.547	2.970	8.739	14.750
최소값	0.01	0	0	0
최대값	24	25.2	86.3	264.1
관측수	594	408	408	597

강수에 속하는 기상요인들로 부산, 울산지역의 강수특성에 대해서 추정 할 수 있었다. 첫째, 대부분 관측수가 408~597개 이므로 총 관측수 1,826개의 약 1/3~1/4 수준 즉, 3일 또는 4일 중에 하루는 비가 온다고 볼 수 있었으며, 둘째, 한번 비가 오기 시작하면 평균적으로 강수시간이 약 6시간 정도 지속된다는 것을 확인할 수 있었다.

[표3-5] "풍속" 관련 기술통계분석 결과

구분	최대 순간 풍속(m/s)	최대 풍속(m/s)	평균 풍속(m/s)
평균	10.531	6.353	3.120
표준 편차	3.275	2.025	1.141
최소값	3.4	2.4	0.3
최대값	33.6	19.2	9.8
관측수	1,826	1,826	1,826

[표3-5]는 풍속 관련 기술통계분석 결과이다. 위 결과를 통해 풍속관련 기상요인들로 부산, 울산지역의 바람의 특성에 대해서 추정할 수 있었다. 관측수가 1,826개로 모두 존재하는 것으로 보아, 강수와는 달리 바람이 전혀 없는 날은 없으며, 평균풍속 3.12m/s, 최대순간풍속이 10.531m/s로 [표3-6]과 같이 강풍주의보 또는 강풍경보에 해당하지 않는 일상 수준의 바람이 부는 것을 알 수 있었다.

[표3-6] 강풍주의보 와 강풍경보 기준 [14]

구분	주 의 보	경 보
강풍	평균풍속이 14m/s이상 또는 순간 풍속이 20m/s이상 예상될 때, 다만 산지의 경우 평균풍속 17m/s이상 또는 최대순간풍속 25m/s이상이 예상될 때	평균풍속이 21m/s이상 또는 순간 풍속이 26m/s이상 예상될 때, 다만 산지의 경우 평균풍속 24m/s이상 또는 최대순간풍속 30m/s이상이 예상될 때

[표3-7] “정전횟수” 관련 기술통계분석 결과

구분	정전건수
평균	2.173
표준 편차	7.555
최소값	0
최대값	215
관측수	1826

[표3-7]은 종속변수인 일일 정전횟수의 기술통계분석 결과이다. 정전횟수의 특징은 평균적으로 매일 약 2건 정도의 정전이 발생하고 있음을 알 수 있었으며, 최소값과 최대값은 각각 0과 215로 정전발생횟수의 범위가 다른 기상요인에 비하여 상당히 넓은 것을 알 수 있었다. 이로써 Excel 기술통계분석을 통하여 각각 변수의 특성에 대하여 간략하게 파악하였다.

강수에 속하는 다양한 세부 기상요인들과 일일 정전횟수 간의 상관관계 분석을 위하여 강수의 세부 기상요인인 강수 계속시간(hr), 10분 최다 강수량(mm), 1시간 최다 강수량(mm), 일 강수량(mm)을 정전횟수와 Pearson 상관관계를 분석을 시행하였으며, 그 결과는 [표3-8] 과 같다.

일 강수량의 Pearson 상관계수가 0.261로 1시간 최다 강수량 0.222, 10분 최다 강수량 0.165, 강수 계속시간 0.158 보다 상관관계가 상대적으로 높은 것을 알 수 있었으며, 관측치 수(N) 또한 일 강수량 597개 > 강수 계속시간 594개 > 1시간 최다강수량 408개 = 10분 최다강수량 408개 순으로 일 강수량이 관측치 수가 가장 크다는 것을 알 수 있었다.

그리하여, 강수를 대표하는 최종 독립변수로 일 강수량을 선정하였으며, 4가지 변수는 모두 유의확률은 0.001 이하로 통계적으로 유의미한 것으로 나타났다.

[표3-8] “강수” 관련 Pearson 상관분석 결과

구분		정전횟수	강수 계속시간	10분 최다 강수량	1시간 최다 강수량	일 강수량
정전횟수	Pearson 상관	1	0.158**	0.165**	0.222**	0.261**
	유의확률		0.000	0.001	0.000	0.000
	N	1,826	594	408	408	597
강수 계속시간	Pearson 상관	0.158**	1	0.361**	0.440**	0.621**
	유의확률	0.000		0.000	0.000	0.000
	N	594	594	406	406	594
10분 최다 강수량	Pearson 상관	0.165**	0.361**	1	0.938**	0.761**
	유의확률	0.001	0.000		0.000	0.000
	N	408	406	408	408	408
1시간 최다 강수량	Pearson 상관	0.222**	0.440**	0.938**	1	0.873**
	유의확률	0.000	0.000	0.000		0.000
	N	408	406	408	408	408
일 강수량	Pearson 상관	0.261**	0.621**	0.761**	0.873**	1
	유의확률	0.000	0.000	0.000	0.000	
	N	597	594	408	408	597

[표3-9] 는 풍속의 세부 기상요인인 최대순간풍속, 최대풍속, 평균풍속을 정전횟수와 Pearson 상관관계를 분석한 결과이다.

Pearson 상관관계 분석결과, 최대순간풍속의 Pearson 상관계수가 0.281로 최대풍속 0.250, 평균풍속 0.135 보다 상관관계가 상대적으로 높으며, 관측치 수(N) 는 4개 변수 모두 1,826으로 동일한 것을 알 수 있었다. 그리하여 풍속을 대표하는 최종 독립변수로 최대순간풍속을 선정하였으며, 3개 변수는 모두 유의확률은 0.000으로 통계적으로 유의미한 것으로 나타났다.

[표3-9] "풍속" 관련 Pearson 상관분석 결과

구분		정전횟수	최대순간풍속	최대풍속	평균풍속
정전횟수	Pearson 상관	1	0.281**	0.250**	0.135**
	유의확률		0.000	0.000	0.000
	N	1,826	1,826	1,826	1,826
최대순간풍속	Pearson 상관	0.281**	1	0.874**	0.780**
	유의확률	0.000		0.000	0.000
	N	1,826	1,826	1,826	1,826
최대풍속	Pearson 상관	0.250**	0.874**	1	0.798**
	유의확률	0.000	0.000		0.000
	N	1,826	1,826	1,826	1,826
평균풍속	Pearson 상관	0.135**	0.780**	0.798**	1
	유의확률	0.000	0.000	0.000	
	N	1,826	1,826	1,826	1,826

또한, 추가적으로 독립변수 선정의 타당성을 위하여 강수와 풍속 외 나머지 21개 기상요인에 대해서도 Pearson 상관관계 분석을 시행하였고, 그 결과는 [표3-10] 와 같다.

[표3-10] 21개 기상요인 Pearson 상관분석 결과

구 분	Pearson 상관	유의확률	N
일최심신적설	-1.000	-	2
일최심적설	-1.000	-	2
합계3시간신적설	-1.000	-	2

강수와 풍속 외 21개 기상요인 중 일최심신적설, 일최심적설, 합계3시간신적설 3가지가 적설 기상요인은 상관관계는 -1.000 로 매우 강한 음의 상관관계를 갖으나, 관측수(N)가 단 2회라서 유의확률이 존재하지 않아 통계적인 해석이 불가하였다. 또한 21개 기상요인 중 최종선택 된 독립변수인 일 강수량 0.261 과 최대순간풍속 0.281 보다 높은 상관관계를 갖는 기상요인은 찾아볼 수 없어 다시 한번 최종 선택된 독립변수가 적절한 하다는 것은 입증할 수 있었다.

2. 다중 선형 회귀분석

Pearson 상관관계 분석을 통하여 일 강수량, 최대순간풍속을 최종 독립변수로 2개로 선정하였고, 이어서 다중 선형 회귀분석을 수행하였다.

영국의 생물통계학자인 갈톤(Francis Galton:1822~1911)이 아버지와 아들의 신장에 대한 관계를 알아보기 위한 연구에서 처음 제안한 것이 회귀분석이다. 그는 아버지와 아들이 신장에 대한 관계가 완전한 비례관계에 있지 않음에 주장하였다. 결국 키가 큰 아버지의 아들은 평균적으로 아버지보다 키가 작아지는 경향이 있고, 키가 작은 아버지의 아들은 아버지보다 커지는 경향이 있음을 알게 된다. 즉, 아버지와 아들의 신장에 대한 관계식은 각각의 평균점을 지나므로 아들의 신장은 인류 전체의 평균 신장으로 회귀하려는 경향이 있다는 것이다. 이러한 함수관계를 regression(회귀) 라는 용어로 처음 사용한 사람이 갈톤이다. 이처럼 신장의 회귀관계를 나타내는 직선을 회귀직선이라고 부르게 되었으며, 오늘날에는 여러 변수들이 다른 변수에 영향을 주는 함수관계를 분석하는 통계적 기법을 회귀분석이라고 한다. [15]

선형 회귀(linear regression)는 종속 변수 Y와 한 개 이상의 독립 변수(또는 설명 변수) X와의 선형 상관관계를 모델링하는 회귀분석 기법이다. 한 개의 설명 변수에 기반한 경우에는 단순 선형 회귀(simple linear regression), 둘 이상의 설명 변수에 기반한 경우에는 다중 선형 회귀라고 한다. 일반적으로 최소제곱법(least square method)을 사용해 선형 회귀 모델을 세운다. 최소제곱법 외에 다른 기법으로도 선형 회귀 모델을 세울 수 있다. 손실 함수(loss fuction)를 최소화 하는 방식으로 선형 회귀 모델을 세울 수도 있다. 최소제곱법은 선형 회귀 모델 뿐 아니라, 비선형 회귀 모델에도 적용할 수 있다. [16]

한편, R^2는 회귀분석의 결정계수(coefficient of determination)으로써 종속변수의 분산 중 독립변수들(혹은 회귀식)에 의해 설명되는 비율을 나타내며 0과 1사이의 값을 가지는데, 일반적으로 R^2가 클수록 그 회귀식은 높은 설명력을 가지기 때문에 회귀식의 적정성을 판단할 수 있는 중요한 지표가 된다. [17]

또한, 선형회귀는 주어진 데이터 집합 $\{y_i, x_{i1}, \cdots, x_{ip}\}_{i=1}^{n}$에 대해, 종속변수$y_i$ 와 p개의 독립변수 사이에 선형관계를 모델링하며, 모델은 아래와 같은 형태를 갖는다.

$$y_i = \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \epsilon_i = x_i^T \beta + \epsilon_i, \qquad i = 1, \cdots, n$$

주어진 식에서 β_i 는 각 독립변수의 계수이며, p는 선형회귀로 추정되는 모수의 개수이다. $x_i^T \beta$ 은 x_i와 β의 내적을 의미한다. ϵ_i는 오차항으로 관찰되지 않은 확률변수로, 종속변수와 독립변수 사이에 오차를 의미한다. 이것이 선형회귀라고 불리는 것은 종속변수가 독립변수에 대해 1차 선형함수의 관계에 있을 것이라고 가정하기 때문이다. [16]

IBM SPSS Statistics 26 를 통한 다중 선형 회귀분석 결과, 유의확률 0.000으로 통계적으로는 유의미함을 확인하였다. 또한, 다중분산팽창요인(variance inflation factior;VIF)는 1.061으로 10보다 훨씬 작으므로 공선성의 문제는 없다고 할 수 있다. 그러나, 도출한 회귀식의 R^2 값이 0.177으로 설명력이 낮고, 1,826개 데이터의 SSE(explained sum of squares, 잔차제곱의 합)이 82,111.037으로 매우 큰 값을 갖아 Under-fitting 형태의 과소적합 예측모형이 도출된 것을 확인할 수 있었으며, 도출된 회귀식은 아래와 같다.

※ 도출된 회귀식

⇨ 정전횟수(Y) = -9.868 + 0.096$X_{1(\text{일강수량})}$ + 1.096$X_{2(\text{최대순간풍속})}$

[표3-11] 다중 선형회귀 모형의 설명력

R	R^2	R^2_{adj}	추정값의 표준오차
0.421	0.177	0.174	11.757

[표3-11]는 다중선형회귀 모형의 설명력을 정리한 것이다. R^2 값이 0.177이고 R^2_{adj} 0.174 로 도출된 1차식 형태의 회귀식이 과소적합(Under-fitting) 형태로 설명력이 전체적으로 낮은 모형임을 확인할 수 있다. 즉, 다중 선형 회귀분석으로 도출된 회귀식은 종속변수를 설명하기에 필요한 데이터들을 충분히 반영하지 못한 상태이며, 최적적합(Optimal-fitting)을 위해서는 현재 회귀식보다 좀 더 복잡한 형태의 회귀모형이 필요하다는 것을 알 수 있었다.

[표3-12] 다중 선형회귀 모형의 ANOVA

모형	제곱합	자유도	평균제곱	F	유의확률
회귀	17,681.771	2	8,840.885	63.956	0.000
잔차 (SSE)	82,111.037	594	138.234		
전체	99,792.807	596			

[표3-12]는 다중 선형회귀 모형의 ANOVA를 정리한 것이다. 1,826개 데이터의 SSE이 82,111.037으로 매우 큰 값을 갖는 모형임을 알 수 있었으며, ANOVA 분산분석 유의확률 0.000에서 F분포 63.956으로 통계적으로 유의하다고 볼 수 있다.

[표3-13] 다중 선형회귀 모형의 계수

구분	비표준화 계수		표준화계수 베타	t	유의확률	VIF
	B	표준화오류				
상수	-9.868	1.423		-6.935	0.000	
X_1	0.096	0.021	0.179	4.671	0.000	1.061
X_2	1.096	0.123	0.340	8.881	0.000	1.061

[표3-13] 는 다중 선형회귀 모형의 계수를 보여준다. 비표준화 계수는 상수, X_1, X_2 가 t분포 값에 따라 0.0000 유의하며, 상수는 -9.868 이고, X_1(일 강수량)의 계수는 0.096, X_2(최대순간풍속)의 계수는 1.096 로 X_2 가 X_1에 비하여 정전횟수에 상대적으로 큰 영향을 미친다고 볼 수 있다. 또한, 다중분산팽창요인(VIF)은 1.061로 10보다 훨씬 작으므로 다중공선성의 문제는 없다고 할 수 있다.

3. 다항 비선형 회귀분석

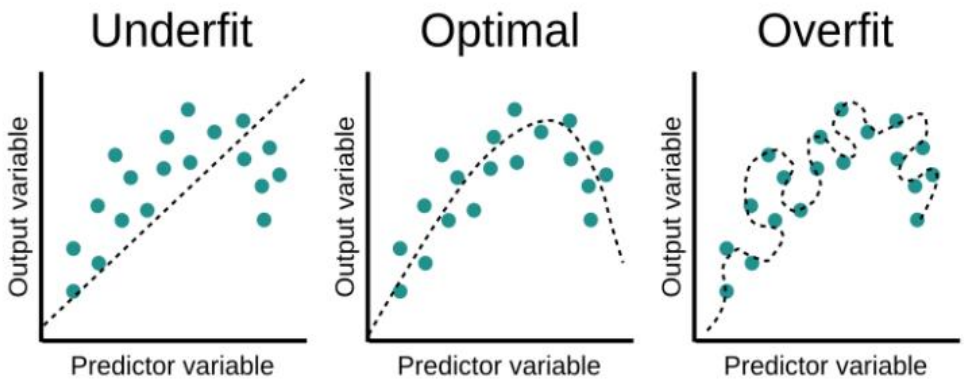
앞서 IBM SPSS Statistics 26 로 수행했던 다중 선형 회귀분석의 Under-fitting 한계점으로 인하여, MATLAB R2021a 의 Curve Fitting Toolbox 기능을 활용하여 다항 비선형 회귀분석을 시행하였다.

복잡하고 다양한 실제 데이터를 모형화 하는데 있어 선형회귀 모형이 부적합한 경우가 종종 있다. 이처럼 비선형 회귀분석은 어떤 인자나 변량 사이의 상관관계를 분석하기 위한 한 수법이며, 어떤 변량 x로부터 다른 변량 y를 추정하는 경우에 단순히 양자의 상관의 강약뿐만 아니라 x에 대한 y의 평균적인 경향이나 불균형의 정도가 2차 곡선 등 고차의 곡선으로 특정짓는 관계에 있는 변량군의 분석을 말하며, 비선형 회귀분석과 선형 회귀 분석은 다음과 같은 점에서 유사하다. 첫째, 하나의 반응 변수와 하나 이상의 예측 변수 사이의 관계를 수학적으로 설명이 가능하고, 둘째, 곡선 형태의 관계를 모형화 할 수 있으며, 셋째, 잔차의 제곱합(SSE)을 최소화하며, 넷째, 잔차 그림을 사용하여 확인할 수 있는 동일한 가정을 가진다. [17]

또한, Curve fitting 기능은 일련의 데이터 포인트에 가장 적합한 곡선 또는 수학적 함수를 구하는 과정이며, 회귀분석에 주로 사용한다. 방법으로는 주로 다항식(Polynomial)을 사용하며, 간단한 경험적 모델이 필요할 때 자주 사용된다. Polynomial Fitting의 장점으로는 데이터에 대한 적절한 유연성이 있다는 점과 Fitting 과정이 간단하며 모형 적합도는 SSE 등으로 빠르게 파악할 수 있다는 장점이 있다. 그러나 Polynomial Fitting의 단점은 첫째, 고차항의 Fitting시, 불안정해질 수 있다는 점과 둘째, 모든 차수의 다항식은 데이터 범위 내에서는 Optimal-fitting을 제공할 수 있지만, 해당 범위 밖에서는 크게 발산할 수 있다는 점이다.

[18]

일반적으로 회귀분석을 시행하면 [그림3-1] 과 같이 만나게 되는 2가지 문제점으로 과소적합(Under-fitting)과 과대적합(Over-fitting)이 있다. 이 문제들을 해결하여 최종적으로 Optimal-fitting을 도출하는 것이 회귀분석의 목표이다.



[그림3-1] 과소적합, 최적적합, 과대적합 비교 [19]

[표3-1]의 첫 번째 그림은 과소적합(Under-fitting) 모형의 사례로 회귀식이 너무 단순해서 데이터의 내재된 구조를 회귀식으로 표현하지 못할 때 발생하며, 과소적합의 해결방법으로는 첫째, 파라미터가 더 많은 복잡한 모델을 선택하는 방법이 있고, 둘째, 모델의 제약을 줄이는 방법이 있으며, 셋째, Over-fitting이 되기 전까지 충분히 회귀식을 도출하는 방법이 있다. [20]

반면에, [표3-1]의 세 번째 그림은 과대적합(Over-fitting) 모형의 사례이다. 과대적합은 과소적합의 반대개념으로 모형이 원본 데이터에는 너무 잘 맞지만 일반성이 떨어지는 현상을 의미한다. 과대적합의 해결방법으로는 첫째, 원본 데이터를 더 많이 모으거나, 둘째, 정규화 또는 제약조건 등 다양한 방법을 이용해서 적당한 복잡도를 갖는 모델을 찾는 방

법이 있으며, 셋째, 원본 데이터의 오류수정과 이상치 제거로 노이즈를 줄이는 방법이 있다. [20]

일반적으로 회귀분석에 있어, 과소적합(Under-fitting)과 과대적합(Over-fitting) 둘 다 발생하면 안된다. 그러나 Under-fitting 보다는 Over-fitting인 상태가 더 좋을 수 있다. 왜냐하면 Over-fitting은 최소한 원본 데이터에서라도 모형의 성능이 잘 나오기 때문이다. 그러나 회귀분석의 목표인 일반화를 위해서는 과소접합, 과대적합을 모두 피해서 최적적합인 Optimal-fitting을 찾는 과정이 필요하다. [20]

본 연구에서는 앞선 다중 선형 회귀분석의 Under-fitting 문제를 해결하기 위하여 파라미터가 많은 복잡한 모델인 다항식(Polynomial) 모델의 사용하였다. 일반적으로 비선형 곡선에 대한 다항식 모델은 아래와 같이 지정된다.

$$y = \sum_{i=1}^{n+1} p_i x^{n+1-i}$$

$n+1$은 다항식의 위수이며 위수는 fitting 할 계수의 개수를 나타낸다. 또한, n은 다항식의 차수며 차수는 예측변수의 가장 높은 거듭제곱을 나타낸다. 예를 들어, 3차 다항식은 아래와 같이 지정된다. [18]

$$y = p1x^3 + p2x^2 + p3x + p4$$

다항 비선형 회귀식을 산출하기 위하여 일 강수량 X와 최대순간풍속 Y를 각각 1~5차 함수형태 일때의 SSE와 R^2값을 도출했으며, 총 25번의 시뮬레이션 결과를 정리하면 [표3-14]와 같다.

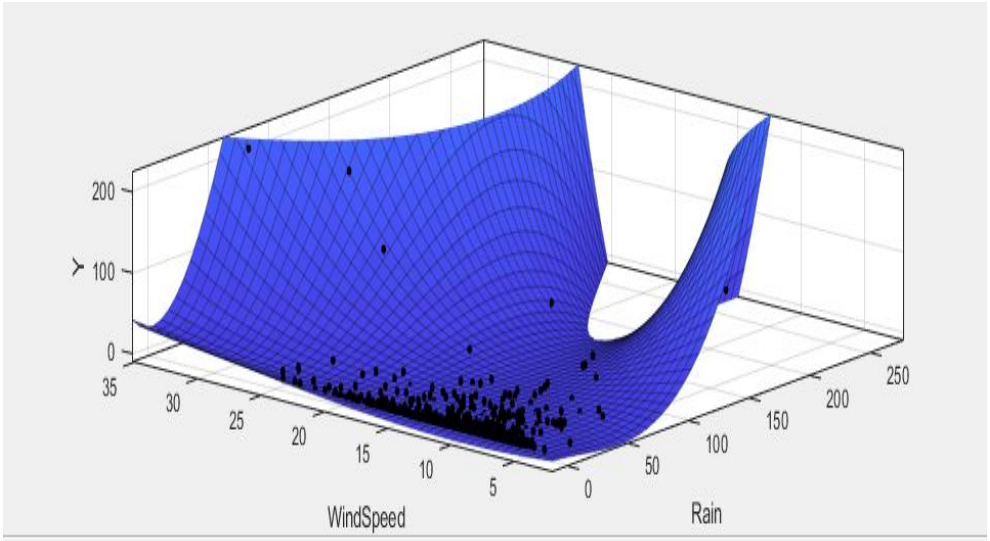
[표3-14] 다항 비선형 회귀분석 결과

X (일 강수량)	Y (최대순간풍속)	SSE	R^2
1차	1차	82,111	0.177
	2차	49,780	0.522
	3차	23,710	0.7724
	4차	17,270	0.8343
	5차	15,530	0.851
2차	1차	54,610	0.4759
	2차	48,240	0.537
	3차	23,600	0.773
	4차	15,900	0.8474
	5차	12,620	0.8789
3차	1차	54,600	0.476
	2차	23,910	0.7705
	3차	23,610	0.7734
	4차	13,080	0.8744
	5차	11,780	0.8869
4차	1차	42,960	0.5877
	2차	15,580	0.8505
	3차	1,271	0.9878
	4차	1,437	0.9862
	5차	2,071	0.9801
5차	1차	42,220	0.5948
	2차	14,260	0.8631
	3차	1,781	0.9829
	4차	2,109	0.9798
	5차	2,241	0.9785

[표3-14] 의 다항 비선형 회귀 결과, 일 강수량 X는 4차 함수형태, 최대순간풍속 Y는 3차 함수형태로 총 14개 다항식으로 이루어진 비선형회귀모형이 R^2값이 0.9878로 가장 높고, 1,826개 데이터의 SSE(explained sum of squares, 잔차제곱의 합)은 1,271으로 가장 적은 것으로 보아, 최적적합(Optimal-fitting) 모형인 것을 알 수 있었으며 도출된 회귀식은 아래와 같다.

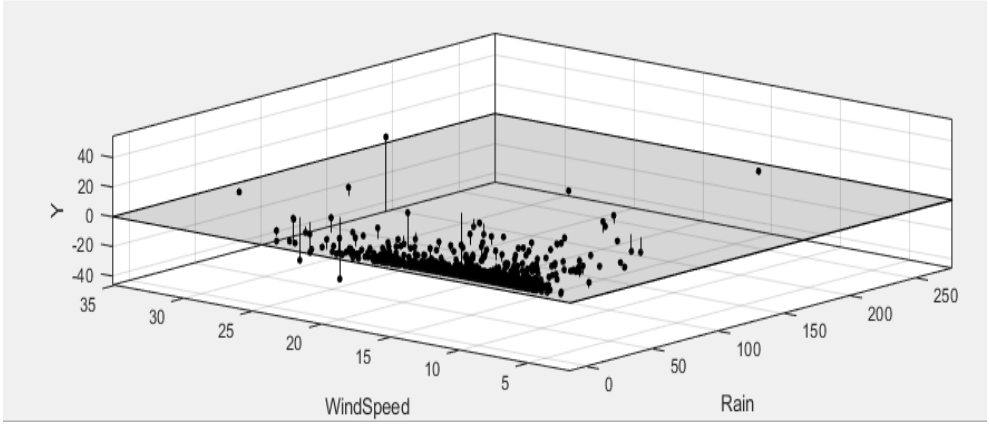
※ 도출된 회귀식

⇨ 정전횟수(Z) = $1.039 + -0.5162X + 0.1221Y + -0.001575X^2 + 0.1709XY + -0.02382Y^{22} + 0.0001256X^3 + -0.001415X^2Y + -0.01175XY^2 + 0.001124Y^3 + 2.07E-09X^4 + -0.000008263X^3Y + 0.0000967X^2Y^2 + 0.0001994XY^3$



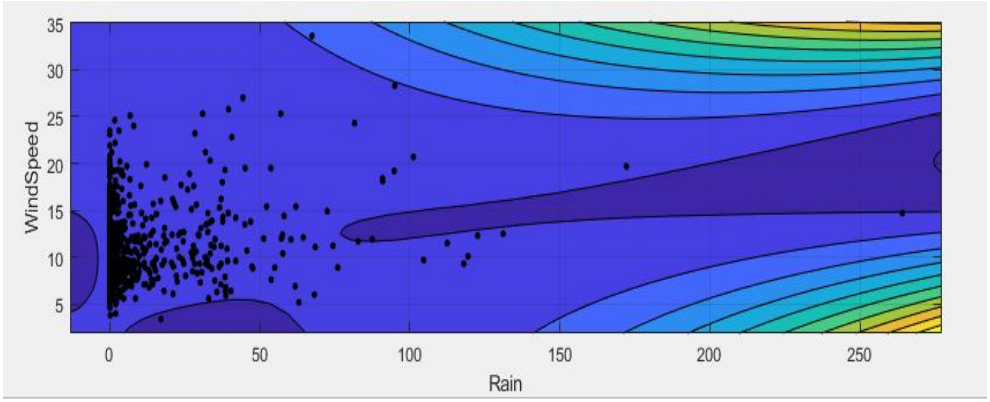
[그림3-2] MATLAB 다항 비선형 회귀 모형의 시각화

[그림3-2]는 다항 비선형 회귀모형을 시각화 한 자료이다. X와 Y가 각각 4차, 3차 함수를 띄고 있기 때문에 전체적으로 곡면형태의 회귀모형이 산출되었음을 확인할 수 있었다.



[그림3-3] MATLAB 다항 비선형 회귀모형의 잔차

[그림3-3]은 다항 비선형 회귀모형의 잔차를 시각화한 값이다. 회귀모형의 잔차는 데이터와 회귀식에 대한 피팅 간의 차이로 정의된다. 즉, 잔차 = 데이터 - 피팅으로 나타낼 수 있으며, 데이터에 피팅한 모델이 올바르다고 가정했을 때, 잔차는 랜덤오차를 근사한다. 따라서 잔차가 무작위적인 양상을 보인다면 이는 모델이 데이터를 잘 피팅한다는 것을 의미한다. 본 연구모형은 육안상으로 잔차가 0을 기준으로 위아래 방향으로 무작위하게 흩어져 있어, 이는 모형이 데이터를 잘 설명한다고 볼 수 있다. [18]



[그림3-4] MATLAB 다항 비선형 회귀모형의 등고선

[그림3-4]는 다항 비선형 회귀모형의 등고선이다. X축인 일 강수량(0~250mm) 과 Y축인 최대순간풍속(0~35m/s)의 적정한 스케일링을 통하여 산점도 형태로 나타낸 것이다. 흩어진 형태의 데이터 대부분이 파란색 부분에 위치하여, 도출된 모형이 데이터를 잘 설명하고 있음을 다시 한 번 확인 할 수 있었다.

제 4 장 연구결과

제 1 절 정전예측 모형 결과 도출

IBM SPSS Statistics 26로 수행했던 다중 선형 회귀분석의 한계점으로 인하여, MATLAB R2021a의 Curve Fitting Toolbox 기능으로 비선형 회귀분석을 시행한 결과, 일 강수량 X는 4차 함수형태, 최대순간풍속 Y는 3차 함수형태의 다항 비선형 회귀분석 모형이 Optimal-fitting 한 배전설비 정전횟수 예측모형으로 도출되었다.

다중 선형 회귀분석은 R^2이 17.7% 로 다항 비선형 회귀분석의 R^2인 98.78%에 비해 설명력이 1/6 수준이며, 다중 선형 회귀분석의 SSE 값은 82,111으로 다항 비선형 회귀분석 SSE 값인 1,271 대비 잔차의 제곱의 합이 약 64배 이상 높은 것을 확인하였다. 다중 선형 회귀분석 모형과 다항 비선형 회귀분석 모형을 비교한 결과는 [표4-1]과 같다.

[표4-1] 다중 선형회귀 및 다항 비선형회귀 결과

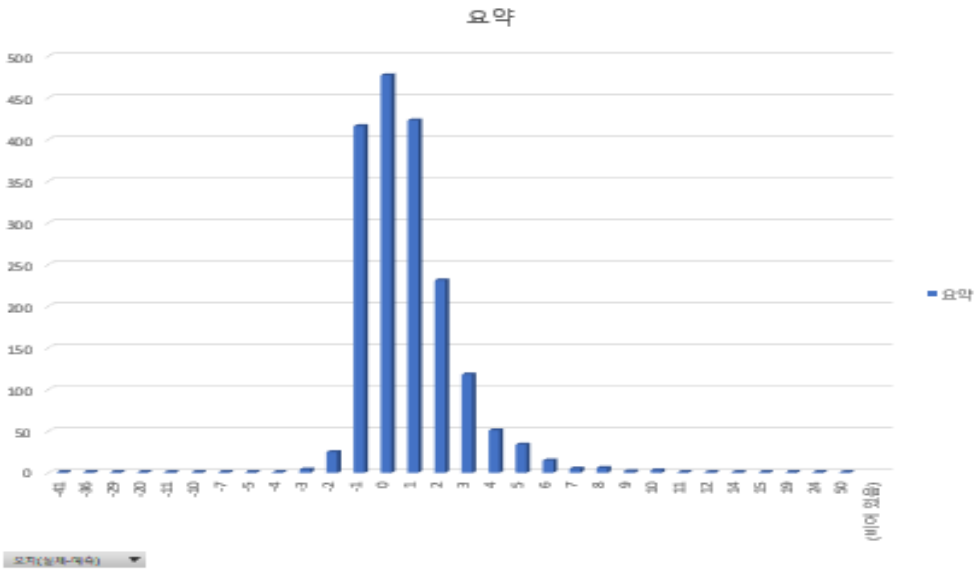
구분	다중 선형 회귀분석 (SPSS)	다항 비선형 회귀분석 (MATLAB)
R^2	17.7%	98.78%
SSE	82,111	1,271

제 2 절 정전예측 모형 검증

1. 2015~2019년 데이터 기반 정전모형검증

앞서 도출된 배전설비 정전예측 모형을 검증하기 위하여 정전예측 모형 설계를 위하여 사용한 원본 데이터인 2015~2019년도 기상청 데이터 중 일 강수량과 최대순간풍속을 입력변수로 활용하여 일일 정전횟수를 예측하였다. 그리고 실제값과 예측값의 오차를 알아보기 위하여 동 기간의 한국전력공사 부산울산지역본부의 일일 실제 정전횟수에서 예측한 정전횟수 간의 차이를 살펴보았다.

그 결과, (실제 정전횟수 - 예측 정전횟수) 기준으로 ±4건 이내에 포함된 정전횟수는 1,826개 중 1,746개로 95.62%를 차지하여 예측모형의 성능이 상당히 높은 것을 확인할 수 있었으며, 정전 예측오차 횟수는 [그림4-1]과 같이 정규분포와 유사한 형태를 취하고 있음을 확인할 수 있었다.



[그림4-1] 2015~2019년 정전모형 검증그래프

(실제 정전횟수 - 예측 정전횟수) ±4건을 기준으로 모형의 성능을 검증한 이유는 한국전력공사 부산울산본부는 13개의 지사로 구성되어 지사간 인력 또는 장비협력 등을 통하여 부산울산본부의 소속된 지사 갯수의 1/3수준인 즉, ±4건 이내의 정전발생은 원활하게 대처 가능하다고 가정하였다. 2015~2019년 정전모형 검증결과는 [표4-2]와 같다.

[표4-2] 2015~2019년 정전모형 검증결과

정전예측오차 (실제 정전횟수- 예측 정전횟수)	해당갯수	비고
-41	1	구간합계 (8건/1,826건) = 0.44%
-36	1	
-29	1	
-20	1	
-11	1	
-10	1	
-7	1	
-5	1	
-4	1	구간합계 (1,746건/1,826건) = 95.62%
-3	4	
-2	25	
-1	416	
0	477	
1	423	
2	231	
3	118	
4	51	
5	34	구간합계 (72건/1,826건) = 3.94%
6	15	
7	5	
8	6	
9	2	
10	3	
11	1	
12	1	
14	1	
15	1	
19	1	
24	1	
50	1	
총합계	1,826	

[표4-2]의 (실제 정전횟수 - 예측 정전횟수)가 -4건을 초과하는 부분은 1,826개 중 8개이며 0.44%로 실제 정전횟수 대비 예측한 정전횟수가 많은 경우라고 볼 수 있다. 이는 전기사업자가 태풍 등 기상예보를 반영하여 재난상황에 선제적으로 대비 한 결과로 해석할 수 있다. 즉, (실제 정전횟수 - 예측 정전횟수)의 오차가 -4건을 초과하는 경우에는 전기사업자 측면에서 인력과 장비가 충분하게 확보되어 실제정전을 원활히 대비할 수 있으므로 큰 문제가 되지 않는다.

예를 들어, [표4-2]의 2015~2019년 정전모형 검증결과 중 (실제 정전횟수 - 예측 정전횟수)의 오차가 음수 중에 가장 큰 -41건인 경우는 2015년 10월 1일의 날씨상태로 일 강수량 57mm 과 최대순간풍속 25.3m/s으로 예측한 정전횟수는 43건이나, 실제 발생했던 정전횟수는 단 2회에 그쳐 오차가 -41건이 발생한 경우다.

또한, [표4-2]의 2015~2019년 정전모형 검증결과 중 (실제 정전횟수 - 예측 정전횟수)의 오차가 음수 중에 두 번째로 큰 -36건인 경우는 2018년 11월 8일의 날씨상태로 일 강수량 101.2mm 과 최대순간풍속 20.7m/s으로 예측한 정전횟수는 39건이나, 실제 발생했던 정전횟수는 3건에 그쳐 오차가 -36건이 발생한 경우다.

그러나, [표4-2]의 (실제 정전횟수 - 예측 정전횟수)가 +4건을 초과하는 부분은 1,826개 중 72개이며 3.94%로, 예측 정전횟수 대비 실제 정전횟수가 많은 결과이다. 이는 예측한 정전횟수를 기반으로 인력과 장비를 확보할 경우, 인력과 장비가 충분히 확보되지 않아 현장에서 큰 혼란을 야기할 수 있다. 즉, (실제 정전횟수 - 예측 정전횟수)가 +4건을 초과하는 경우에는 전기사업자가 실제 정전에 적절하게 대응을 하기 어려울 수 있다.

실제로 [표4-2]의 (실제 정전횟수 - 예측 정전횟수)가 +50건인 경우는 2019년 9월 22일에 2019년 제17호 태풍인 “타파”가 한반도에 접근했을 때의 결과이다. 타파는 한반도 상륙하지는 않았지만 대형크기의 태풍이라서 부산지역을 지나면서 간접적인 영향을 주었다.

중형의 강한 태풍 (KMA) 대형의 강한 태풍 (JMA) 1등급 태풍 (SSHS)		
활동 기간	9월 19일 ~ 9월 23일	
최저 기압	970hPa	
최대 풍속 (10분 평균)	KMA	37m/s
	JMA	35m/s (65kt)
태풍 피해자	3명 사망 피해총액 790만 달러	

[그림4-2] 2019년 제17호 태풍「타파」개요 [21]

이 날의 기상조건은 일 강수량 81.6mm 과 최대순간풍속 24.3m/s 으로 예측한 정전횟수는 70건이나, 실제 발생했던 정전횟수는 120건으로 예측한 정전횟수보다 실제 정전횟수가 +50건 많았던 경우였다.

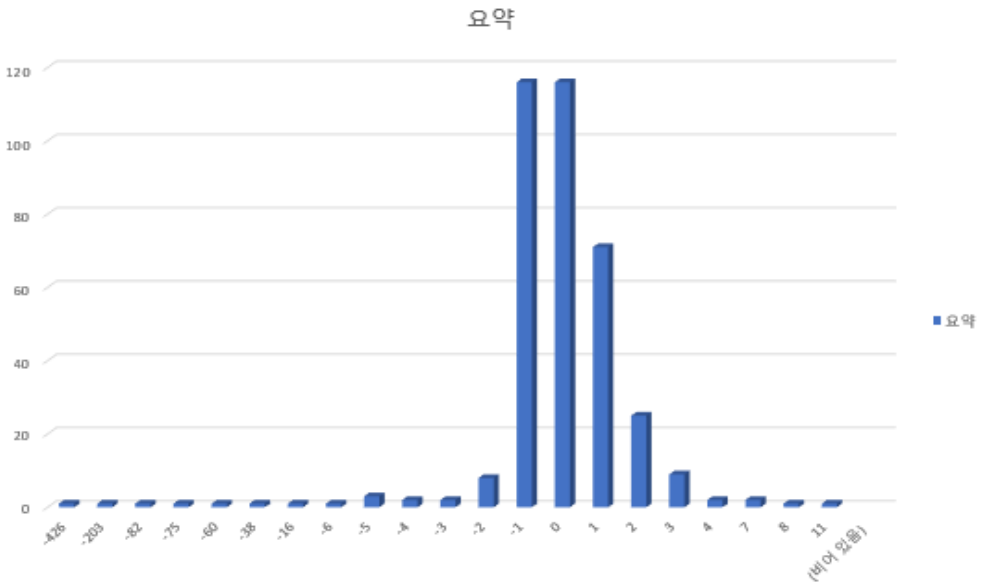
또한, [표4-2]의 2015~2019년 정전모형 검증결과 중 (실제 정전횟수 - 예측 정전횟수)의 오차가 양수 중에 두 번째로 큰 +24건인 경우는 2015년 7월 12일의 날씨상태로 일 강수량 28mm 과 최대순간풍속 17.6m/s으로 예측한 정전횟수는 2건이나, 실제 발생했던 정전횟수는 26건으로 오차가 +24건이 발생한 경우다.

즉, (실제 정전횟수 - 예측 정전횟수)가 +4건을 초과하는 경우에는 전기사업자가 실제 정전에 적절하게 대응을 하기 어려운 상황이 발생하게 된다. 또한, 배전설비의 정전으로 인하여 배전계통의 공급신뢰도와 전기품질이 떨어지게 되어, 위 정전횟수 예측모형의 한계점을 볼 수 있는 사례이다.

2. 2020년 데이터 기반 정전모형 검증

2015~2019년 정전모형 검증결과 정전예측모형이 (실제 정전횟수 - 예측 정전횟수)를 기준으로 ±4건 이내에 포함된 예측된 정전횟수는 1,826개 중 1,746개로 95.62%를 차지하여 예측모형의 성능이 상당히 높은 것을 확인할 수 있었다. 그러나 정전모형 설계를 위하여 사용한 2015~2019년 데이터는 원본 데이터이므로 예측모형과는 당연히 오차가 적게 발생할 수 밖에 없다.

본 모형의 실질적인 목적은 전기사업자 측면에서 기상예보를 바탕으로 배전설비 정전횟수를 예측할 수 있어야 한다. 그러므로 2020년도 기상청 기상 데이터 중 일 강수량과 최대순간풍속을 입력변수로 활용하여 2020년의 일일 정전횟수를 예측하였고, 2020년 동 기간의 한국전력공사 부산울산지역본부의 실제 일일 정전횟수를 산출하여 정전횟수를 오차를 산출했으며, 그 결과는 [표4-3]과 [그림4-3]과 같다.



[그림4-3] 2020년 정전모형 검증그래프

[그림4-3]과 같이 정전 예측오차는 앞서 2015~2019년 데이터 기반으로 정전모형을 검증했던 결과와 유사하게 정규분포 형태를 취하고 있음을 확인할 수 있었으며, [표4-3]의 (실제 정전횟수 - 예측 정전횟수)를 기준으로 ±4건에 이내에 포함된 정전횟수는 366개 중 351개이며 95.62%으로 예측모형의 성능이 상당히 높은 것을 확인할 수 있었다.

[표4-3] 2020년 정전모형 검증결과

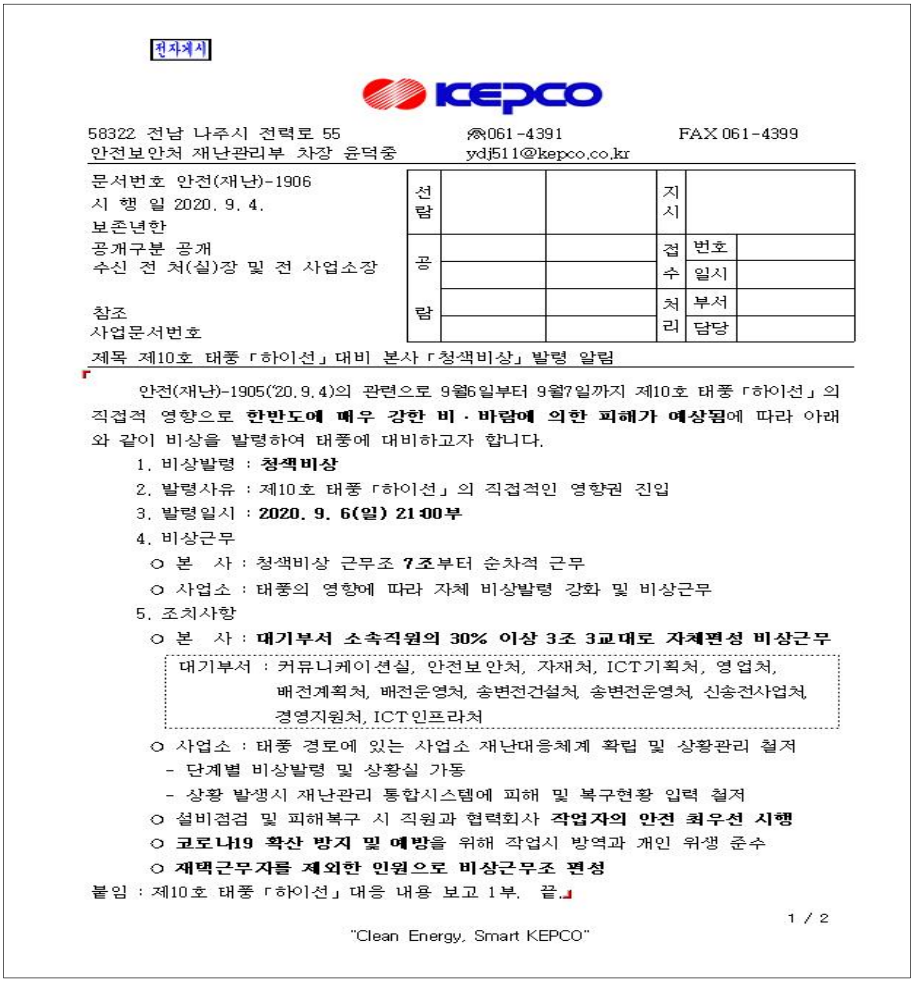
정전예측오차 (실제 정전횟수- 예측 정전횟수)	해당갯수	비고
-426	1	구간합계 (11건/366건) = 3.00%
-203	1	
-82	1	
-75	1	
-60	1	
-38	1	
-16	1	
-6	1	
-5	3	
-4	2	구간합계 (351건/366건) = 95.91%
-3	2	
-2	8	
-1	116	
0	116	
1	71	
2	25	
3	9	
4	2	
7	2	구간합계 (4건/366건) = 1.09%
8	1	
11	1	
총합계	366	

[표4-3]의 (실제 정전횟수 - 예측 정전횟수)의 오차가 -4건을 초과하는 부분은 366개 중 11개이며 3.00%으로, 실제 정전횟수 대비 예측한 정전횟수가 많은 경우라고 볼 수 있다. 이는 전기사업자가 태풍 등 기상예보를 반영하여 재난상황에 선제적으로 대비 한 결과로 해석할 수 있다. 즉, (실제 정전횟수 - 예측 정전횟수)의 오차가 -4건을 초과하는 경우에는 전기사업자 측면에서 인력과 장비가 충분하게 확보되어 실제 정전을 원활히 대비 할 수 있으므로 큰 문제가 되지 않는다.

4등급의 슈퍼 태풍 (SSHS)		
활동 기간		9월 1일 ~ 9월 8일[2]
최저 기압		910[9]hPa
최대 풍속 (10분 평균)	KMA	55m/s
	JMA	55m/s (105kt)
태풍 피해자		4명 사망 피해총액 1억 달러

[그림4-4] 2020년 제10호 태풍「하이선」개요 [20]

예를 들어 [표4-3]의 음수 중에 가장 큰 오차인 -426건은 [그림4-4]의 2020년 9월 7일에 발생한 제10호 태풍 “하이선” 관련 예측 데이터로 일 강수량은113.6mm 이 관측 되었고, 최대순간풍속은 32.2m/s 이 관측 되었다. 4등급의 슈퍼태풍의 규모에 맞게 정전횟수는 436건을 예측하였다. 그러나 전기사업자가 태풍발생 3일 전 부터 [그림4-5] 과 같이 태풍 대비 비상발령 및 비상근무 시행 등 철저한 대비를 하여 실제 정전건수는 10건이 발생하여 -426건의 오차가 발생하게 되었다.



전자게시

KEPCO

58322 전남 나주시 전력로 55 ☎061-4391 FAX 061-4399
안전보안처 재난관리부 차장 윤덕중 ydj511@kepco.co.kr

문서번호 안전(재난)-1906
시 행 일 2020. 9. 4.
보존년한
공개구분 공개
수신 전 처(실)장 및 전 사업소장

참조
사업문서번호

선람			지시		
공람			접수	번호	
				일시	
			처리	부서	
				담당	

제목 제10호 태풍「하이선」대비 본사「청색비상」발령 알림

안전(재난)-1905('20.9.4)의 관련으로 9월6일부터 9월7일까지 제10호 태풍「하이선」의 직접적 영향으로 **한반도에 매우 강한 비·바람에 의한 피해가 예상됨**에 따라 아래와 같이 비상을 발령하여 태풍에 대비하고자 합니다.

1. 비상발령 : **청색비상**
2. 발령사유 : 제10호 태풍「하이선」의 직접적인 영향권 진입
3. 발령일시 : **2020. 9. 6(일) 21:00부**
4. 비상근무
   - ○ 본 사 : 청색비상 근무조 **7조**부터 순차적 근무
   - ○ 사업소 : 태풍의 영향에 따라 자체 비상발령 강화 및 비상근무
5. 조치사항
   - ○ 본 사 : **대기부서 소속직원의 30% 이상 3조 3교대로 자체편성 비상근무**

     대기부서 : 커뮤니케이션실, 안전보안처, 자재처, ICT기획처, 영업처, 배전계획처, 배전운영처, 송변전건설처, 송변전운영처, 신송전사업처, 경영지원처, ICT인프라처
   - ○ 사업소 : 태풍 경로에 있는 사업소 재난대응체계 확립 및 상황관리 철저
     - 단계별 비상발령 및 상황실 가동
     - 상황 발생시 재난관리 통합시스템에 피해 및 복구현황 입력 철저
   - ○ 설비점검 및 피해복구 시 직원과 협력회사 **작업자의 안전 최우선 시행**
   - ○ **코로나19 확산 방지 및 예방**을 위해 작업시 방역과 개인 위생 준수
   - ○ **재택근무자를 제외한 인원으로 비상근무조 편성**

붙임 : 제10호 태풍「하이선」대응 내용 보고 1부. 끝.

1 / 2

"Clean Energy, Smart KEPCO"

[그림4-5] 태풍 “하이선” 대비 공문 [23]

또한, [표4-3] 의 음수 중에 두 번째로 큰 오차인 -203건은 2020년 7월 10일의 기상 데이터로 일 강수량 208.7mm 과 최대순간풍속 9.5m/s인 경우로 많은 양의 강수량으로 인하여 예측한 정전횟수는 204건이나, 실제 발생했던 정전횟수는 1건으로 오차가 -203건이 발생한 경우다. 위 사례는 풍속은 9.5m/s로 평범한 수준이나, 일 강수량 208.7mm 으로 상당히 많은 양의 비가 내린 경우로, 일정 데이터 범위 내에서는 적합한 fitting을 제공하지만 해당 범위 밖에서는 크게 발산할 수 있는 다항 비선형 회귀모형의 한계점을 볼 수 있는 사례이다.

[표4-3]의 (실제 정전횟수 - 예측 정전횟수)가 +4건을 초과하는 부분은 366개 중 4개이며 1.09%로, 예측 정전횟수 대비 실제 정전횟수가 많은 결과이다. 이는 예측한 정전횟수를 기반으로 인력과 장비를 확보할 경우, 인력과 장비가 충분히 확보되지 않아 현장에서 큰 혼란을 야기할 수 있다. 즉, (실제 정전횟수 - 예측 정전횟수)가 +4건을 초과하는 경우에는 전기사업자 측면에서 실제 정전에 적절하게 대응을 하기 어려울 수 있다.

실제로 [표4-3] 의 양수 중에 가장 큰 오차인 +11건은 2020년 10월 5일에 발생한 경우로, 일 강수량 36.3mm 과 최대순간풍속 9.8m/s 의 상황이다. 이 날은 태풍 발생 등 비상상황은 아니었으나, 예측한 정전 2건 대비 실제 발생한 정전은 13건으로 +11건의 오차가 발생한 상황이다. 이는 전기사업자가 실제 정전에 적절하게 대응을 하기 어려운 상황으로 도출된 정전횟수 예측모형의 한계점을 엿볼 수 있는 경우다.

제 5 장 결론

제 1 절 연구의 결과

본 연구는 한국전력공사 부산울산지역본부의 소속된 13개 지사의 2015~2019년 5개년 간 실제 정전횟수를 종속변수로 설정하였고, 선행연구를 참고하여 기상청의 종관기상관측(ASOS) 자료 중 풍속, 강수, 적설 3개를 고 영향 기상요인으로 설정하였다. 그러나 부산·울산지역의 특성상 적설빈도가 2015~2019년 5개년 간 단 2회 발생(2018. 1. 10, 2019 1. 31)하여, 통계 관측수의 부족으로 유의확률이 산출되지 않아 통계적인 해석이 불가하였다. 그래서 적설을 제외한 풍속과 강수를 고 영향 기상요인으로 설정하였다.

선행연구에 근거하여 2개의 고 영향 기상요인을 설정하였고, Pearson 상관분석을 통하여 풍속과 강수 중 상관관계가 높은 기상요인인 최대순간풍속과 일 강수량을 각각 최종 독립변수로 설정하였다. 선행연구처럼 특정 배전설비의 정전확률 예측모형이 아닌 전체 배전설비의 정전횟수 예측모형을 모델링하기 위하여 IBM SPSS Statistics 26을 활용하여 다중 선형 회귀분석을 수행하였고, 다중 선형 회귀분석으로는 과소적합 형태의 정전 예측모형이 도출되어, MATLAB R2021a의 Curve Fitting Toolbox을 활용하여 다항 비선형 회귀분석을 시행하여 Optimal-fitting의 예측모형을 도출하였다.

도출된 정전횟수 예측모형은 1차로 원본 데이터인 2015~2019년 5개년 기상 데이터를 기반으로 예측한 정전횟수와 실제 정전횟수를 비교하여 검증하였고, 2차로 2020년의 기상 데이터를 기반으로 예측한 정전횟수와 실제 정전횟수를 비교하여 정전모형을 검증하였다.

2020년의 기상 데이터로 검증결과, (실제 정전횟수 - 예측 정전횟수)를 기준으로 ±4건 이내에 포함된 예측된 정전횟수는 366개 중 351개로 95.62%를 차지하여 정전 예측모형의 성능이 상당히 높은 것을 확인할 수 있었다.

(실제 정전횟수 - 예측 정전횟수)의 오차가 -4건을 초과하는 부분은 366개 중 11개이며 3.00%로, 실제 정전횟수 대비 예측한 정전횟수가 많은 경우라고 볼 수 있다. 이는 전기사업자가 태풍 등 기상예보를 반영하여 재난상황에 선제적으로 대비 한 결과로 해석할 수 있다. 즉, (실제 정전횟수 - 예측 정전횟수)의 오차가 -4건을 초과하는 경우에는 전기사업자 측면에서 인력과 장비가 충분하게 확보되어 실제 정전을 원활히 대비할 수 있으므로 큰 문제가 되지 않는다.

실제로 2020년도 기준으로 음수 중에 가장 큰 오차인 -426건은 2020년 9월 7일 에 발생한 제10호 태풍 "하이선" 관련 데이터로 전기사업자가 1주일 전부터 태풍대비 비상발령 및 비상근무 시행 등 철저한 대비를 해서 예측한 정전만큼 실제 정전이 발생하지 않은 것으로 보인다.

그러나, (실제 정전횟수 - 예측 정전횟수)가 +4건을 초과하는 부분은 366개 중 4개이며 1.09%로, 예측 정전횟수 대비 실제 정전횟수가 많은 경우이다. 이는 예측한 정전횟수를 기반으로 인력과 장비를 확보할 경우, 인력과 장비가 충분히 확보되지 않아 현장에서 큰 혼란을 야기할 수 있다. 즉, (실제 정전횟수 - 예측 정전횟수)가 +4건을 초과하는 경우에는 전기사업자가 실제 정전에 적절하게 대응을 하기 어려울 수 있으며, 도출된 정전횟수 예측모형의 한계점을 엿볼 수 있는 경우이다.

제 2 절 한계점 및 향후 연구방안

본 연구의 한계점은 첫째, 기상청의 기상 데이터는 부산지역 단일지역에서 측정된 반면, 실제 배전설비 정전발생 위치는 한국전력공사 부산울산지역본부 소속 13개 지사 전 지역에서 흩어져 있어 기상측정 위치와 실제 정전이 발생한 위치 간에 상이한 점이 있다.

둘째, 희소한 경우지만 2020년의 기상 데이터를 기반으로 검증결과 (실제 정전횟수 - 예측 정전횟수) 간 오차가 +4건을 초과하는 부분이 366개 중 4개로 1.09% 존재하여, 정전예측 모형에 전적으로 의존할 경우 예측한 정전횟수를 초과하여 실제 정전이 발생한 경우에 한해서는 전기사업자의 적절한 대응이 곤란할 수 있다.

그러나, 2015~2019년 데이터(1차)와 2020년 데이터(2차)로 검증결과, 2차례 검증결과 모두 (실제 정전횟수 - 예측 정전횟수) 기준으로 ±4건 이내의 오차가 95% 이상을 차지하여 모형의 성능이 상당히 우수한 것을 확인할 수 있었다.

그러므로 본 연구의 정전횟수 예측모형은 부산, 울산지역에 한정하는 것이 아닌 전국적인 확대가 필요하다고 생각하며, 전기사업자가 배전설비를 운영 또는 관리함에 있어 참고자료로 활용하기에 충분하다고 생각한다.

향후 연구에서는 데이터의 특성상 시계열의 특성도 갖고 있어 ARIMA 등 시계열 분석을 통하여 정전횟수를 예측하는 방법도 필요하다고 생각한다.

참 고 문 헌

[1] 이동성·문태훈, “도시 및 기후특성이 에너지 회복력에 미치는 영향”, 대한국토·도시계획학회지, 2019. 8

[2] 기상청, “2020년 제10호 태풍「하이선」경로”

[3] 한국전력공사, “e-영업통계시스템” 2021. 7

[4] 한국전력공사, “부산울산지역본부 일반현황”, 2021. 5

[5] International Electro technical Commission, "IEC-61000-4-30"

[6] IEEE1159, "IEEE Recommended Practice for Monitoring Electric Power Quality“

[7] 한국전력공사, “배전계통 공급신뢰도 관리기준”, 2020. 6

[8] IEEE1366, "IEEE Guide for Electric Power Distribution Reliability Indices"

[9] 오도은·김현수, “기상기후 기반 배전 개폐기 고장예측에 관한 연구”, "J. of Advanced Engineering and Technology", 2017. 3

[10] 임청원·한상국, “빅데이터 기반 전력계통 고장예측시스템 구축을 위한 기상요인 및 배전설비 고장에 관한 연구”, “한국데이터정보과학회지”, 2018. 11

[11] 기상청 기상자료개방포털, “종관기상관측장비(ASOS)”

[12] 한국전력공사, “2015~2019년 부산울산지역본부 정전내역”

[13] 집현재(이학식, 임지훈), “BASIC SPSS MANUAL" 2012. 9

[14] 한국전력공사, “배전설비 피해복구 표준절차서”, 2020. 12

[15] 한나래아카데미, “SPSS 회귀분석 제3판”

[16] 위키피디아, “선형회귀”

[17] https://support.minitab.com, “비선형 회귀분석이해”

[18] https://kr.mathworks.com/help/curvefit/polynomial, “다항식 모델”

[19] https://www.educative.io, "Over-fitting and Under-fitting"

[20] O'Reilly Media, "Hands-On Machine Learning with Scikit-Learn, Keras, and Tensorflow"

[21] 위키피디아, “2019년 제17호 태풍「타파」개요”

[22] 위키피디아, “2020년 제10호 태풍「하이선」개요”

[23] 한국전력공사, “제10호 태풍「하이선」대비 본사「청색비상」발령 알림”, 2020. 9

Abstract

A Study on the Model of the Relationship between the Number of Power Outages and Weather Factors in the Distribution System Using Actual Power Outage Data

Choi Moon Ju

Department of Engineering Practice

Graduate School of Engineering Practice

Seoul National University

This study models the relationship between the number of power distribution facilities outages and weather factors by utilizing actual power outage data in Busan and Ulsan.

The actual number of power outages of 13 branches belonging to the Busan Ulsan Regional Headquarters of Korea Electric Power Corporation(KEPCO) for 5 years between 2015 and 2019 was set as dependent variables. With reference to previous studies, wind speed, precipitation and snow accumulation 3 factors were set as highly

affecting power distribution facilities outages. However, due to the characteristics of Busan and Ulsan, snow accumulation occurred only twice between 2015 and 2019(2018. 1. 10, 2019 1. 31). So, The lack of statistical observations resulted in no significant probability, which made statistical interpretation impossible. Therefore, 2 factors (wind speed and precipitation) excluding snow accumulation, were set as high-impact weather factors. Also, Through Pearson's correlation analysis, the maximum instantaneous wind speed and daily precipitation were set as the final independent variables, respectively

We performed multiple linear regression using IBM SPSS statistics 26, and multiple linear regression resulted in an under-fitting model. Thus, a polynomial non-linear regression analysis using the curve fitting toolbox of MATLAB R2021a could be performed to derive a predictive model of the optimal-fitting model.

First, based on data for 2015-2019, the power outage prediction model was verified by comparing the predicted number of power outages and the actual number of power outages.

Secondly, based on data for 2020, the power outage prediction model was verified by comparing the predicted number of power outages and the actual number of power outages

The data for 2020 showed that within ±4 of (actual number of power outages - predicted number of power outages) is 351 out of 366, accounting for 95.62%, which resulted in significantly higher performance of the model.

More than -4 of (actual number of power outages - predicted number of power outages) is 11 out of 366 cases, 3.00%, which is considered to be the case where the predicted number of power outages is higher than the actual number of power outages. This is the result of electricity operators proactively preparing for disaster

situations by reflecting weather forecasts such as typhoons, and it is not a big problem because enough manpower and equipment are secured from the side of electricity operators to smoothly prepare for the actual power outage.

However, More than +4 of (actual number of power outages - predicted number of power outages) is 4 out of 366 cases, 1.09%, which is considered to be the case where the actual number of power outages is higher than the predicted number of power outages. This can cause major confusion in the field as manpower and equipment are not sufficiently secured when securing personnel and equipment based on the predicted number of power outages. In other words, if the (actual number of power outages - predicted number of power outages) is more than +4, it may be difficult for the electricity operators to respond appropriately to the actual power outages, and This is a case where the threshold of the derived power outage prediction model can be seen.

Future research projects are not limited to Busan and Ulsan, but I think it is necessary to identify regional weather factors nationwide and derive a prediction model for the number of power outages suitable for weather characteristics in each region.

keywords : Distribution system, power outage data,
power outage count, weather factors, regression
Student Number : 2020-26551