

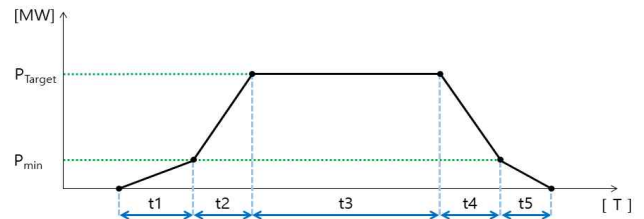
## EWMA와 3-시그마를 이용한 발전기 출력데이터 이상치 탐지와 처리방안 소개

**송재은, 공지웅, 이주호**  
한국전력거래소

### Anomaly Detection and Treatment of Generator Power Output Data Using EWMA and 3-Sigma

Jae-Eun Song, Ji-Woong Gong, Ju-Hoo Lee  
Korea Power Exchange

**Abstract** - For electricity supply-demand and power generation operation, current generator status identification is important. However, there may be a deviation between the actual output and the acquired data due to field conditions or communication problems. Among these deviation data, two cases were defined as outliers: the case where the generator's up/down ramp-rate greatly deviates and the case where the generator's output fluctuated from the constant output trend. This paper described that outlier detection and treatment using EWMA and 3-Sigma method, and it was verified with real data.



<그림 1> 시간대별 발전기 출력 곡선

#### 1. 서 론

큰 출력 변동성을 갖는 신재생 에너지원의 증가는 발전계획과 실적 간 편차에 직접적인 영향을 주고, 이는 전력수급과 계통운영에 어려움을 초래한다. 이에 안정적인 전력수급을 위해서 발전기의 실시간 운전상태와 발전출력 정보가 매우 중요하다. 그러나 현장 상황 또는 통신의 문제로 인해 발전기의 출력데이터가 이상치로 수신되거나 혹은 결측이 발생할 경우, 발전기의 현재 운전상태가 현장과 다르게 판단될 수 있다. 검출이 용이한 결측치와 달리, 이상치의 경우 탐지와 처리가 상대적으로 까다롭다. 이러한 이상치를 처리하기 위해 이동평균과 3-시그마를 이용한 연구들이 수행되었다.[1-2] 본 논문에서는 발전기 상태판별의 품질 향상을 위해, 지수가중 이동평균과 3-시그마 기법으로 이상치를 탐지하고 처리하는 방법을 제안하고자 한다.

#### 2. 본 론

##### 2.1. 발전기의 상태 판별

현재 발전계획은 익일 24시간에 대해, 수요예측과 발전기별 입찰정보를 바탕으로 하루 전 수립되어 이를 당일에 사용하고 있다. 그러나 최근 신재생 에너지원의 확대에 의해 발전계획과 실적 간 편차에 미치는 영향이 커지게 되었고, 전력수급 및 발전운영에 대한 부담이 점점 증가하고 있다.

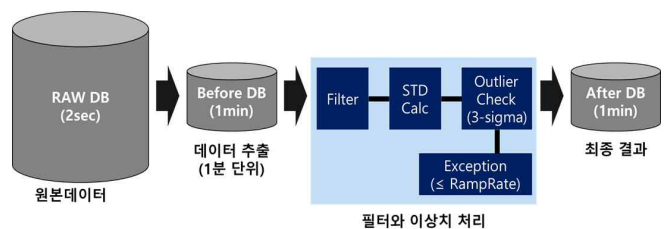
안정적인 전력수급을 위해서 계통 운영자는 제어가능한 발전기의 현재 출력을 조정하고, 추가적으로 각 발전기의 상태에 따라 운전과 정지시기를 수행해야 한다. 이 과정에서 각 발전기의 실시간 운전상태와 발전출력 정보는 매우 중요한 참조자료가 된다.

발전기의 상태는 발전출력과 그 추이를 통해 확인할 수 있다. <그림 1>은 계통에 연결된 발전기의 시간에 따른 출력변화를 나타낸다. 발전기의 운전 구간은 총 다섯 가지로 구분할 수 있다. 발전기가 안전 운전이 가능한 최소출력( $P_{min}$ )에 도달하는 기동구간(t1), 분당증발률(MW/min)로 목표출력( $P_{target}$ )까지 도달하는 출력증가구간(t2), 그리고 목표출력을 유지하면서 연속운전(t3)하는 구간으로 구분된다. 이후 수급상황에 따라 발전기 정지가 필요할 경우, 분당감발률(MW/min)로 최소출력에 도달하는 출력감소구간(t4), 마지막으로 계통에서 분리하는 정지구간(t5)으로 구분된다.[3]

그러나 일부 발전기에서 발전소 현장 상황이나 통신의 문제로 인해, 발전기 특성을 벗어난 이상치가 관측될 수 있다. 이상치는 잘못된 발전기 상태분석을 초래할 수 있는 값이므로 반드시 출력추세에 맞추어 올바르게 처리되어야 한다. 예를 들어 발전기의 출력이 최소출력( $P_{min}$ )을 기준으로 오르내리게 된다면, 발전기의 상태는 기동구간과 정지구간을 반복하게 되어 발전기의 현재 운전상태가 현장과 다르게 판단될 수 있다. 그리고 일정한 목표출력으로 운전 중에 일시적으로 크게 변동되는 값은 데이터의 오류나 노이즈로 의심될 수 있다. 따라서 본 연구에서는 이상치의 기준을 ‘발전기의 증감발률을 크게 벗어나는 경우’와 ‘일정출력을 갖는 추세로 연속 운전하는 범위에서 크게 벗어나는 경우’ 이 2가지로 정의하여 이상치를 탐지하고 이를 검증하고자 한다.

##### 2.2. 이상치 탐지와 처리 과정

<그림 2>는 데이터의 이상치 탐지와 처리 과정을 나타내고 있다. 우선 발전소로부터 2초마다 수집된 발전기 출력데이터의 크기 축소를 위해, 1분 단위로 재추출한다. 그리고 지수가중 이동평균과 3-시그마 기법을 이용해 이상치를 탐지하고, 만일 이상치로 판별되면 이를 지수가중 이동평균의 결과로 대체하여 적용하였다.



<그림 2> 데이터 처리 과정

##### 2.2.1 지수가중 이동평균

최근 관측값에 큰 가중치를 주고 최근 변화를 반영하여 이상치를 탐지하는 지수가중 이동평균(EWMA: Exponentially Weighted Moving Average)방법은 식(1)과 같이 산출된다. EWMA는 단순 이동평균과 달리 잡음제거와 변화 민감성을 계산 결과에 반영할 수 있으며, 계절성을 보이거나 변화가 큰 시계열 자료에 적용할 수 있다. 또한 재귀식 계산 방식으로 결과계산이 빠르고, 샘플링 데이터를 모두 저장하지 않아도 되는 장점이 있다.

그러나 최적의 결과를 얻기 위한 가중치( $\lambda$ ) 설계에 있어 경험적인 부분에 많이 의존해야 하는 한계가 있다. Hunter의 연구에 따르면  $\lambda$ 는 0.2~0.3 사이의 값으로 사용할 것을 제안하였고[4], 본 연구에서는 0.3의 가중치를 적용하였으며 샘플링 구간은 10개(10분 단위)를 적용하여 그 결과를 별도 테이블에 저장하였다.

$$EWMA_t = \lambda X_t + (1 - \lambda)EWMA_{t-1} \quad (1)$$

$(t = 1, 2, \dots, n)$

$EWMA_t$ :  $t$ 시점의 결과

$X_t$ :  $t$ 시점의 관측치

$\lambda$ : 가중치

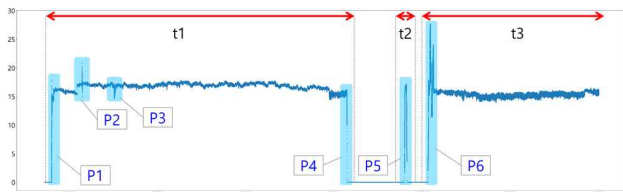
$$(0 < \lambda \leq 1)$$

### 2.2.2 이상치 판별과 예외처리

통계학에서 3-시그마 규칙은 평균으로부터 양쪽으로  $\pm 3$  표준편차의 범위에 대부분의 값들(99.7%)이 포함되어 있다는 의미로, 범위 밖(0.3%)의 값들은 이상치로 간주한다.[5] 본 논문에서는 동일 샘플링 구간으로 EWMA를 계산 결과를 재이용해, 이동표준편차를 계산하고 이동표준편차의 3배에 해당하는 상한과 하한범위를 설정하였다. 단, 상한과 하한의 차이가 발전기 증감 발률에 비해 충분히 작다면, 범위를 벗어나더라도 이는 정상데이터로 간주하여 예외처리 하였다.

### 2.3. 발전기 출력 데이터의 모의 및 분석

본 절에서는 실제 발전기 출력데이터를 이용해, 제안한 이상치 탐지 및 처리 방법을 적용하였다. 모의에 사용된 A열병합 발전기의 데이터 전체 구간은 '22.02.06~'22.03.22(45일간)에 해당한다. <그림 3>은 A열병합 발전기의 출력곡선(2초 단위, RAW DB)을 나타내며, A열병합 발전기는  $t_1$  구간 동안 연속운전을 하였다가  $t_2$  구간에서 약 6시간 동안 짧게 운전 후 정지하고, 다시  $t_3$  구간에서 연속운전 하였다. 육안으로 직접 이상치가 확인가능한 구간을 하늘색 블록으로 표시(P1~P6)하였다. 앞서 이상치는 출력의 변화율이 급격한 곳과 일정한 출력추세를 벗어나는 2가지 구간으로 정의하였다. P1, P4, P5의 경우 기동 및 정지구간으로 출력의 변화율이 큰 구간이며, P2, P3는 일정 출력추세를 벗어나는 구간이다. 마지막으로 P6는 2가지 특징이 혼합된 구간이다.

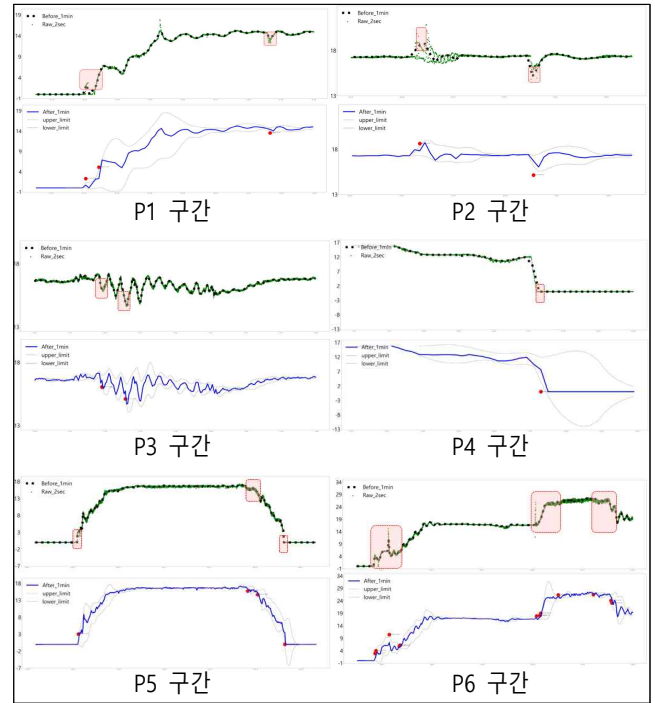


<그림 3> A열병합 발전기 출력데이터

<그림 4>는 P1~P6에 해당하는 구간에 대해, 이상치 탐지 및 처리결과를 확대하여 보여주고 있다. 각각의 그림은 위아래로 구분되어 있는데, 상단의 그림은 2초 단위의 RAW DB(녹색)와 1분 단위로 재추출한 데이터(검정색)이며 이상치로 의심되는 구간은 적색 박스로 표시하였다. 하단의 그림은 상, 하한 범위를 회색 실선으로 표시하였으며, 상단 그림에서 이상치로 의심되는 구간의 실제 이상치를 적색으로 표시하였다. 그리고 이상치를 EWMA 결과로 대체한 최종 결과의 흐름 곡선은 청색 실선으로 표시하였다.

2초 데이터로부터 1분 단위의 데이터를 추출하였으므로, 이상치를 추출한 경우에는 이를 잘 탐지하였고 이상치 주변의 데이터를 추출한 경우 전체 추세 흐름의 변동성에 큰 영향을 미치지 않아 정상 데이터로 올바르게 판별되었다. 그리고 이상치로 탐지된 데이터는 EWMA 결과로 대체되어 추세가 급격하게 변화는 문제가 보완되는 것이 확인되었다. 먼저 P1, P4, P5는 기동 및 정지구간으로 발전기가 계통연결 직후 출력이 크게 올랐을 때를 이상치로 판별하고 출력을 낮은 값으로 대체하여 적용하였다. 반면 발전기가 정지하는 시점의 출력(OMW)이 이상치로 판별되었는데, 일반적으로 발전기는 정지 직전 특정 출력을 내는 상

태에서 계통분리가 되기 때문에 이는 정상데이터로 해석되어야 한다. 또한 P2는 일정출력이 유지되어야 할 구간에서 일시적으로 벗어나는 시점의 데이터를 올바르게 탐지하였으나, P3처럼 연속적인 진동을 가지고 출력이 변동할 경우 부분적으로만 이상치가 탐지되는 것이 확인되었다. 그리고 2가지 특징이 혼합된 P6구간에서는 기동 직후 발전기의 출력이 크게 위아래로 변동되는 시점에서 이상치를 올바르게 판별하였으며, 이후 일정 출력 추세를 벗어나는 이상치 또한 정상적으로 감지되었음이 확인되었다.



<그림 4> 이상치 처리 전후 결과 비교

## 3. 결 론

본 논문은 발전기 출력데이터에서 이상치를 처리하기 위해, 지수가중 이동평균 및 3-시그마 기반의 이상 탐지 및 처리 기법을 제안하고 실제 발전출력 데이터를 이용해 이를 검증하였다.

기동 직후 출력은 이상향이나, 불규칙하게 출력이 변동되는 시점에서 이를 올바르게 이상치로 탐지하였으며, 일정출력 추세를 범위를 벗어나는 경우 또한 이상치로 잘 탐지됨이 모의결과를 통해 확인되었다. 그러나 발전출력이 우하향하는 정지시점에서 불필요한 이상치가 탐지되었고 이는 향후 보완이 필요하다.

지수가중 이동평균은 가중치에 따라 그 성능이 결정되므로 다양한 발전기에 이를 적용하기 위해서는 가중치 설계에 대한 연구와 검토가 필요할 것으로 생각된다.

### [참고 문헌]

- [1] 손시운, “이동 평균과 3-시그마를 이용한 하둑 로그 데이터의 이상 탐지”, 한국정보처리학회, p.283-288, 2016
- [2] 고민정, “이동 평균을 이용한 인터넷 경제 시스템의 낙찰 예정가 자동생성”, 한국전자거래학회, p.17-31, 2004
- [3] 전력거래소, “계통평가세부운영규정”, 제3장 발전기 기술적 특성자료 작성 및 적용기준, 3.3.19, 2022
- [4] Hunter, J. S. (1986). The Exponentially Weighted Moving Average. Journal of Quality Technology, 18(4), 203-210. doi:10.1080/00224065.1986.11979014
- [5] F. Pukelsheim, “The three sigma rule,” The American Statistician, Vol.48, Issue 2, pp.88-91, 1994.