



## ARIMA모델에 기반한 화재발생 빈도 예측모델의 설계

안상훈 · 강 훈 · 조재훈 · 김태옥 · †신동일

명지대학교 화학공학과

(2015년 3월 10일 접수, 2015년 4월 6일 수정, 2015년 4월 7일 채택)

## Forecasting Model Design of Fire Occurrences with ARIMA Models

Sanghun Ahn · Hoon Kang · Jaehoon Cho · Tae-Ok Kim · †Dongil Shin

Department of Chemical Engineering, Myongji University, Yongin 449-728, Korea

(Received March 10, 2015; Revised April 6, 2015; Accepted April 7, 2015)

### 요약

화재를 예방하기 위해서는 그에 필요한 정책과 평가가 필요하며, 화재 발생에 대한 적절한 모니터링 기법이 있어야 한다. 이를 위해 품질관리의 기법중의 하나인 관리도를 화재 모니터링에 적용하는 연구가 진행된 바 있다. 그 결과, 통계적으로 겨울에는 많은 화재가 발생한다는 것이 증명되었고, 매년마다 어떤 주기를 띠고 변화한다는 것이 발견되었다. 그럼에도 불구하고, 여름철과 겨울철에 동일한 기준(관리한계선)을 적용할 경우 상대적으로 발생건수가 과다한 겨울에는 과도한 예방활동이 진행되고, 여름에는 잠재이상이 있는 경우에도 발생건수 감소에 따른 착시현상으로 인해 활동이 부족해질 가능성이 있다. 이때, 각 계절별로 다른 관리 한계선을 적용시킨다면 합리적인 예측과 보다 효과적인 이상 패턴의 모니터링이 가능해질 수 있다. 따라서 본 연구에서는 발생건수를 대상으로 시간에 따라 변하는 시계열 모델을 사용하여 화재발생 빈도 예측 모델계수의 체계적인 설정과정을 ARIMA 모델을 기반으로 제안하였다. 이를 바탕으로 화재발생 패턴의 개선된 분석과 이에 기반한 보다 체계적인 예방활동을 진행할 수 있을 것으로 기대한다.

**Abstract** - A suitable monitoring method is necessary for successful policy implementation and its evaluation, required for effective prevention of abnormal fire occurrences. To do this, there were studies for applying control charts of quality management to fire occurrence monitoring. As a result, it was proved that more fire occurs in winter and its trend moves yearly-basis with some patterns. Although it has trend, if we apply the same criteria for each time, inefficient overreacting fire prevention policy will be accomplished in winter, and deficient policy will be accomplished in summer. Thus, applying different control limits adaptively for each time would enable better forecasting and monitoring of fire occurrences. In this study, we treat fire occurrences as time series model and propose a method for configuring its coefficients with ARIMA model. Based on this, we expect to carry out advanced analysis of fire occurrences and reasonable implementation of prevention activities.

**Key words** : fire occurrence pattern, time series, ARIMA model, autocorrelation

†Corresponding author: dongil@mju.ac.kr

Copyright © 2015 by The Korean Institute of Gas

## I. 서론

화재는 인류와 더불어 상존해온 재해 중의 하나이며, 자연적으로 발생하는 경우도 있지만 대개는 인위적인 행동에 의해 발생한다. 화재 발생현상은 여러 물리적 수식에 의해 설명될 수 있으나, 이것들에 의해 화재 발생패턴의 이상을 정확하게 진단하거나, 예견하기에는 부족한 측면이 있다.

화재 발생과 피해를 감소시키기 위해서는 주로 통계적인 자료를 바탕으로 정책을 수립하거나 규제를 시행하는 방안이 있는데, 이를 위해 세계 각 국가에서는 발생한 각 화재의 기록들을 축적하여 이를 통계분석하고 있다. 특히, 미국에서는 1972년부터 National Fire Incident Reporting System[1]을 만들어 매년 약 백만 건에 대한 화재 통계보고서를 발표하고 있으며, 우리나라도 연간 약 4만 건의 화재사고를 국가화재정보시스템(NFDS)에 등록하고 있다.

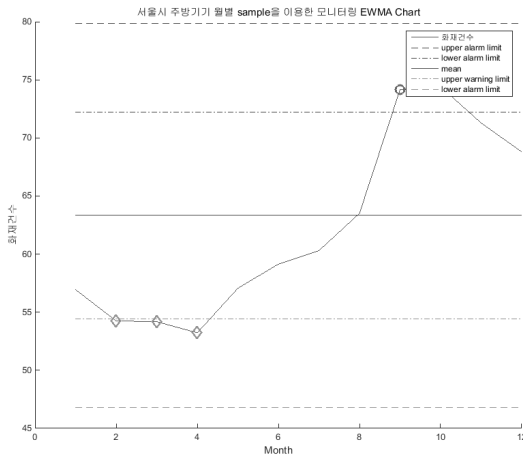


Fig. 1. EWMA control chart for fire occurrence monitoring system.

Table 1. Comparison of monitoring methods

Method		Strength	Weakness
Multivariate analysis		Possible to analyze correlated data	Hard to understand Complex model
Univariate analysis	Analysis without autocorrelation data	Intuitive, Easy to write	Unsuitable to analyze autocorrelated data
	Curve fitting	Intuitive, Easy to write	Unsuitable to nonstationary autocorrelated data
	Analysis with autocorrelation data	Intuitive	Hard to calculate

또한 캐나다에서도 산불에 대해 Shewhart X-chart, CUSUM chart 등과 같은 Statistical Process Control 기법을 사용하여 모니터링을 시도한 연구가 있으며 [2], 국내에서도 주기돈[3]의 연구를 비롯해 유사한 연구가 진행된 사례가 있다.

그러나 이들 기존 연구는 몇 가지 한계점을 보이고 있다. 즉, 화재 발생이 상대적으로 적은 여름철에는 이상을 감지하기 힘들고(false negative), 발생건수가 많은 겨울철에는 잘못된 이상을 감지하는 오류(false positive)를 갖고 있다. 즉, Shewhart chart를 비롯한 기존 control chart들은 과거의 변량과 상관성(자기상관성)을 가지는 변수를 모니터링 하는 경우 공정의 정상성(stationarity) 여부와 표준편차와 같은 통계적 변량의 비합리성과 같은 몇 가지 문제점이 있다. 이것은 화재 발생빈도가 항상 일정한 값을 유지해야 하는 데이터가 아니며, 자기상관성을 가지는 시계열 데이터이기 때문이다. 또한 Fig. 1과 같이 공정이 정상성에서 벗어나는 경우를 빠르게 탐지하기 위한 EWMA chart(지수가중이동평균 관리도)도 자기상관성을 가지면서 정상성을 보이는 계열의 데이터에 대하여는 모델의 강건성이 떨어진다는 한계가 있다[4].

화재 발생빈도를 예측하기 위해 기상정보를 사용하여 시간당 발생건수를 수치형 데이터로 간주하는 data mining기법을 이용한 기상과 화재의 상호 연관관계를 설명하는 모델을 만드는 연구도 진행된 바 있다[5].

따라서 본 연구에서는 현실적으로 화재 발생빈도를 모니터링하기 위해 화재 예측시간에 따른 변화를 반영하는 시계열 모델을 적용하는 방법을 제안하였다.

## II. 이론

Table 1과 같이 모니터링 기법에는 크게 단변량을 다루는 기법과 다변량을 다루는 기법이 있으나,

본 연구에서는 단변량만을 고려하였다. PCA와 같은 다변량 분석기법은 차원축소를 통해 상관관계가 있는 변수들을 더 적은 종류의 변수들로 변환하는 방식으로 상관성을 제거하지만, 시계열 분석기법은 변량을 잡음을 포함하는 제귀식으로 변환하는 방식으로, 결과가 알아보기 간결하다는 장점이 있다.

**2.1. Control chart[6]**

Control chart는 어떤 통계변량,  $X$ 에 대해 평균선과 관리 상한선 및 하한선을 설정한다. 이 경계선들은 변량이 정규분포를 따른다는 가정에 근거하는데, 예를 들어  $X$ 가 정규분포  $X \sim N(\bar{X}, \sigma^2)$ 를 따른다고 할 때 추출한 샘플의 변량을 95%의 신뢰도로 예측하고자 한다면 표준편차와 그 신뢰구간의 관계는 식 (1)과 같이 주어지고, 식 (1)에 의해 예측 가능한 통계변량,  $X$ 의 범위를 예측한다.

$$P(\bar{X} - 1.96\sigma \leq X \leq \bar{X} + 1.96\sigma) = 0.95 \quad (1)$$

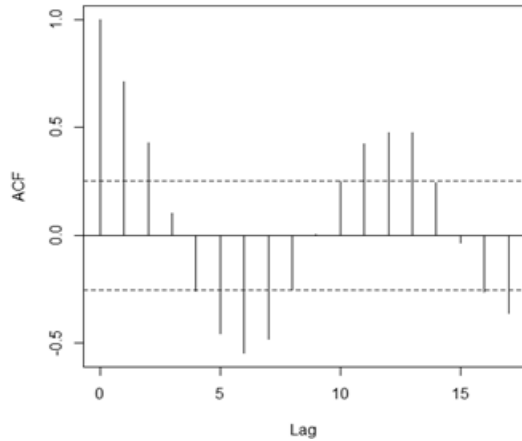
여기서  $\bar{X}$ 는 평균이고,  $\sigma$ 는 표준편차이다.

Shewhart chart는 정규분포에서 얻은 신뢰구간을 일반적인 정상관리구간으로 설정하고, 여기서 벗어난 변량을 이상으로 판단한다. 이것은 항상 일정한 수치를 요구하는 변량, 예를 들어 생산품의 질량이나, 연속식 화학공정에서의 농도 또는 온도와 같은 변량의 모니터링에 적절하다.

Control chart는 조업자나 관리자들 중에서 통계적인 교육이 부족한 인원에 대해서도 직관적인 판단 기준을 제공하여 이상이 발생할 시 적절한 조치를 취할 수 있게도 한다. 또한 시간에 따라 변화하는 변량을 모니터링 할 경우에 공정 자체의 변화에 의해 지속적인 변화가 생길 때 더욱 빨리 반응하기 위해 설계한 CUSUM(Cumulative Sum) chart나 EWMA(Exponential Weight Move Average) chart도 있다. 이것 또한 시간에 관계없이 일정한 수치를 요구하는 변량에만 적절한 문제점이 있다.

**2.2. 시계열**

시계열은 시간에 따라 기록되는 통계변량으로, 시간별 도로의 통행량이나, 연간 무역액수 등 경제적 또는 공학적인 분야에서 흔히 사용되는 개념이다. 다른 통계변량과 다르게 시계열이 가지는 가장 큰 특징은 다른 시점에서 변량과 상관관계를 가지는 자기상관성(autocorrelation)이다. 이것은 다른 시점과의 공분산을 통해 정량화될 수 있다. 즉, 식 (2)와 같이 정의되는 자기상관함수(autocorrelation function)가 일정한 크기를 넘어가면 자기상관성을 가진다.



**Fig. 2.** Plot showing autocorrelation of national monthly fire occurrences.

$$\mu = \sum_{i=1}^n \frac{(X_i - \bar{X})(X_{i+j} - \bar{X})}{\sigma^2} \quad (2)$$

여기서  $X_i$ 는 임의의 시점,  $i$ 에서 통계변량이고,  $j$ 는 상관관계를 고려할 시점과의 인덱스 차이(lag)이며,  $\bar{X}$ 는 평균, 그리고  $\sigma$ 는 표준편차이다.

본 연구에서 사용한 전국 화재발생 건수에 대하여 자기상관함수를 도시하면 Fig. 2와 같다.

Fig. 2에서 1, 2, 5, 6, 7, 10, 11, 12개월 전과의 데이터는 ACF값이 0.25 이상이 되어 유의성이 있는 자기상관성을 가진다고 볼 수 있다. 이러한 수치는 화재 발생빈도가 계절마다 주기를 띠고 늘었다 줄었다 하는 사인(sine)형 곡선을 가지는 사실과 관계가 깊다.

시계열을 예측하는 데에는 본 연구에 사용한 ARIMA(Autoregressive Integrated Moving Average)모델 외에도 최소자승법, 지수평활법, 회귀분석법 등이 있다[2]. 그러나 화재 발생빈도는 선형적으로 변화하는 변량이기보다는 평균이 이동하는 변화 경향을 보이는 비정상적인(nonstationary) 시계열이므로, 본 연구에서는 이러한 경우에 보다 적합하다고 판단되는 ARIMA모델을 사용하였다. ARIMA모델은 아래에 서술된 여러 가지 모델의 결합형태로 이루어졌으며, 주기성을 가진 시계열이나, 평균이 이동하는 시계열에 대해서도 적합한 모델이며, 경제지표의 분석이나 각종 제어분야에도 널리 활용되고 있다.

**2.3. Linear stationary model[7]**

**2.3.1. Autoregressive model**

Autoregressive model을 수식으로 나타내면 정

수형 인덱스,  $t$ 를 가진 어떤 시계열,  $X_t$ 를 유한한 개수의 자신의 과거의 값들에 상수를 곱한 것과 정규분포를 따르는 백색소음(white noise),  $a_t$ 의 합으로 나타낸다.

$$(1 - \phi_1 B^1 - \dots - \phi_p B^p) X_t = B(\phi) X_t = a_t \quad (3)$$

여기서  $B$ 는 지연연산자(lag operator)이고,  $X_t B^n = X_{t-n}$ 으로 나타낸다. 이렇게 나타낸 수식에서 유한한 지연차수,  $p$ 에 대한 모델을 AR( $p$ ) 모델이라고 한다.

### 2.3.2. Moving average model

Moving average model은 정수형 인덱스,  $t$ 를 가진 어떤 시계열 데이터,  $X_t$ 를 유한한 개수의 과거의 noise에 상수를 곱한 것들의 합으로 나타낸다.

$$X_t = (1 - \theta_1 B^1 - \dots - \theta_q B^q) a_t = \theta(B) a_t \quad (4)$$

이렇게 나타낸 수식에서 유한한 지연차수,  $q$ 에 대한 모델을 MA( $q$ ) 모델이라고 한다.

### 2.3.3 Autoregressive moving average model

앞서 설명한 두 수식을 합친 형태로, 이 모델은 시계열이 자기상관성을 가지고, 추세를 가지지 않아서 모델과의 차이인 소음,  $a_t$ 가 정규분포를 따를 때 사용한다.

$$\phi(B) X_t = \theta(B) a_t \quad (5)$$

이와 같이 나타낸 수식에서 유한한  $p$  및  $q$ 에 대한 모델을 ARMA( $p, q$ ) 모델이라고 한다.

## 2.4. Linear nonstationary model[7]

### 2.4.1. Autoregressive integrated moving average (ARIMA) model

시계열이 잡음은 있더라도 추세는 없는 정상성을 가진다면 ARMA 모델로도 충분하지만, 시계열이 추세를 가질 경우에는 잔차를 제거하기 위하여 ARIMA 모델을 도입해야 한다.

ARIMA모델은 잔차를  $\Delta X_t = X_t - X_{t-1} = (1 - B) X_t$ 로 정의하고, 이를 앞서 설명한 ARMA( $p, q$ )모델과 결합하여 식 (6)과 같이 나타낸다.

$$(1 - \phi_1 B - \dots - \phi_p B^p)(1 - B)^d X_t$$

$$= (1 - \theta_1 B - \dots - \theta_q B^q) a_t \quad (6)$$

### 2.4.2. Seasonal ARIMA model

특정 주기를 가지고, 변화를 나타내는 변수에 대한 ARIMA( $p, d, q$ )( $P, D, Q$ ) $m$ 은 식 (7)과 같이 나타낸다.

$$\phi(B)\Phi(B^m)(1 - B)^d(1 - B^m)^D X_t = \theta(B)\Theta(B^m)a_t \quad (7)$$

여기서  $m$ 은 주기의 인덱스로, 본 연구에서와 같이 12개월을 주기로 나타낸다고 가정할 경우에는 12이다. 기존 계수인  $p, d$  및  $q$ 와 계절에 따른 지연계수  $P, D$  및  $Q$ 를 결정하면 각 계수( $\phi, \Phi, \theta, \Theta$ )들에 대해 시행오차를 통해 각 모델에 의한 값과 실제 값의 편차의 제곱합을 최소화 시키는 방법으로 진행한다.

## 2.5. 모델 검증

모델이 완성되면 그 모델의 타당성을 확인해야 하는데, 타당성은 모델로 예측한 값과 실제 값의 오차의 자기상관성 위주로 판단한다. ARIMA 모델이 자기상관성을 제거한 변량을 관찰하기 위하여도 설계한 모델이 자기상관성을 가질 경우 모델이 아무리 test set에 대해 예측을 잘했다더라도 잘못된 모델이기 때문이다.

## III. 데이터 분석

본 연구에서는 뉴질랜드 오클랜드 대학에서 시작된 프로젝트로 개발되어 주로 경제학 분야에서 많이 사용되는 오픈소스 통계 프로그램인 R의 forecast package를 사용하여 데이터를 분석하였다[8,9].

2007년 1월부터 2011년 12월까지의 전국 화재건수는 소방방재청의 NFDS에서 받은 데이터를 사용하였으며, 전국에서 발생한 2007년 1월부터 2011년 12월까지의 5년간의 월별 화재발생량의 데이터로 예측모델을 설계하여 2012년 1월부터 2012년 12월까지의 1년간의 데이터로 예측의 타당성을 평가하는데 사용하였다. 그 이유는 training set과 test set이 서로 다른 데이터를 사용해야 하기 때문이다[10].

## IV. 결과 및 고찰

Fig. 3은 화재발생 예측모델을 결정하기 위하여 2007년~2011년의 5년간 각 월의 화재 발생빈도를 나타낸 것이다. 본 연구에서는 2012년의 월별 화재발생 건수를 예측하고자 하기 때문에 월의 시간단위는

12개월을 기준으로 추세를 가지고 순환하는 시계열이므로, 12개월의 Seasonal ARIMA model, ARIMA(p,d,q)(P,D,Q)<sub>12</sub>로 가정하고, 각 상수들을 결정하였다. 이때, 이들 상수들은 시계열의 자기상관함수의 그래프를 해석하여 결정하였다. 이를 위해 12개월 전과의 편차와 그 자기상관계수(ACF)와 편자기상관계수(PACF)를 도식한 결과는 Fig. 4와 같다.

Fig. 4는 지연에 따른 12개월 전과의 차이의 자기상관함수(ACF)와 편자기상관계수(PACF)를 도식한 것으로, 12개월에서 ACF와 PACF에서 피크가 관찰

되었으므로, 1년 전의 데이터와의 자기상관성을 가진다고 판단하여 계절(season) term에서 MA(1)을 택하였다. 그리고 1개월에서의 피크가 관찰됐으므로, non-seasonal term에서 MA(1)을 택하여 모델의 차수를 ARIMA(0,0,1)(0,1,1)<sub>12</sub>로 결정하고, 오차의 제곱합을 최소화했을 때 식 (8)과 같은 모델을 얻을 수 있었다.

$$z_t(1 - B^{12}) = (1 + 0.3422B)(1 + 0.7597B^{12})a_t \quad (8)$$

모델의 타당성은 Fig. 5(c)와 같이 모델에 의한 예측값과 실제값과의 편차의 자기상관을 해석하기 위한 Ljung-Box test를 수행하여 어떤 달의 화재 발생건수가 다른 달과의 상관에서 얼마나 벗어났는지 판단하였다. 이때, p값이 클수록 많이 벗어남을 의미한다.

Fig. 5(b)는 모델로 계산한 2007년~2011년의 각 월에 대한 화재 발생건수와 실제 발생건수의 차이에 대한 자기상관계수를 나타낸 것으로, 예측모델의 실제 예측여부와는 별개로 모델에 의한 예측값과 실제값과의 편차에 대하여 자기상관계수 값을 구했을 때 자기 자신 외에 다른 시점에서 일정한 값을 넘어가면 모델이 잘못되었다고 판단한다. 그림에서 편차의 유의미(0.25이상)한 자기상관함수가 관찰되지 않았으므로, 모델은 타당하다고 판단하였다.

완성된 ARIMA model로 생성된 예측모델에 의해 예측한 2012년의 화재 발생빈도와 실제 화재 발생빈도를 Fig. 6에 나타내었다. Fig. 6에서 가로축은 0지점을 2007년 1월로 하여 72지점을 2012년 12월로

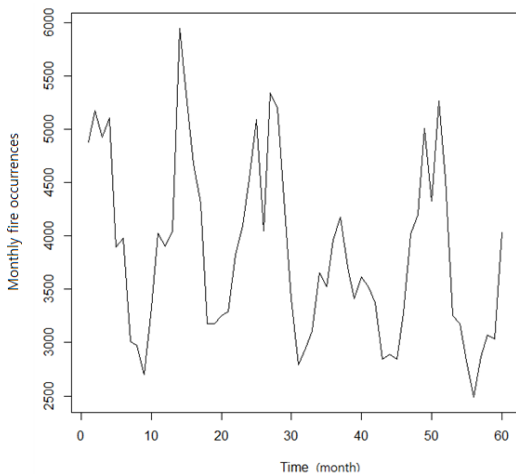


Fig. 3. Fire occurrences in the year 2007.1~2011.12.

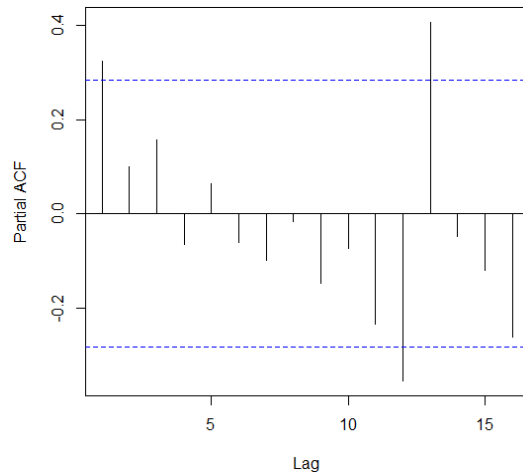
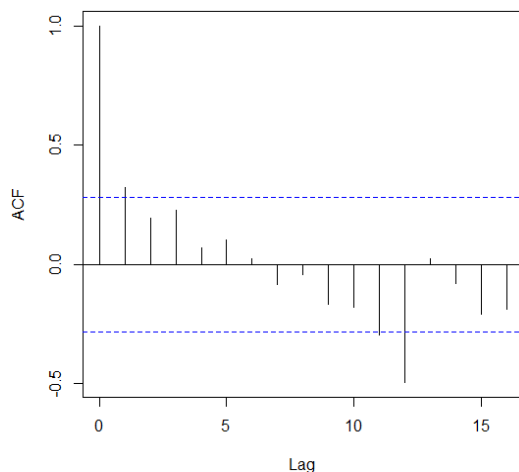


Fig. 4. Autocorrelation of difference with the data of 12-month ahead.

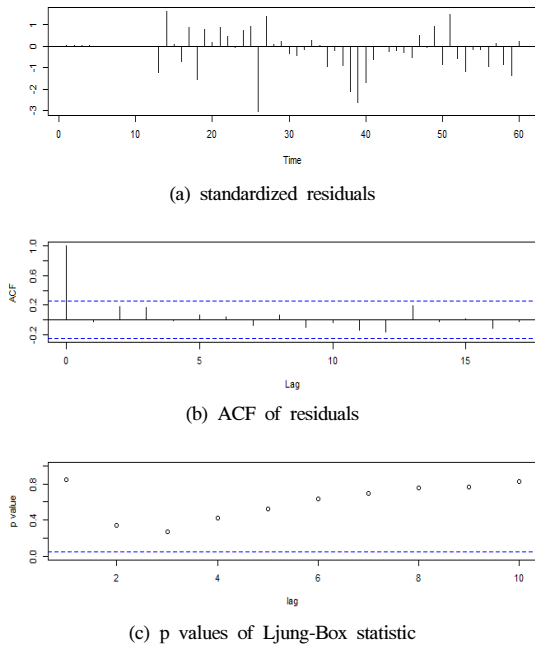


Fig. 5. Criteria of prediction model.

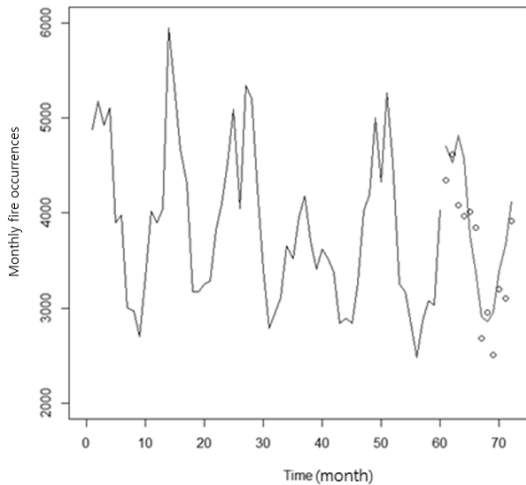


Fig. 6. Predicted/Actual number of fire occurrences of the year 2012.

표기한 것이다. 이때, 선이 끊어진 부분부터가 예측한 2012년도 각 월의 화재 발생건수이고, 원모양의 점은 실제 2012년에 발생한 화재의 건수이다. Fig. 6에서와 같이 2012년의 화재 발생량과 제안한 예측모델은 매우 잘 부합함을 확인할 수 있다.

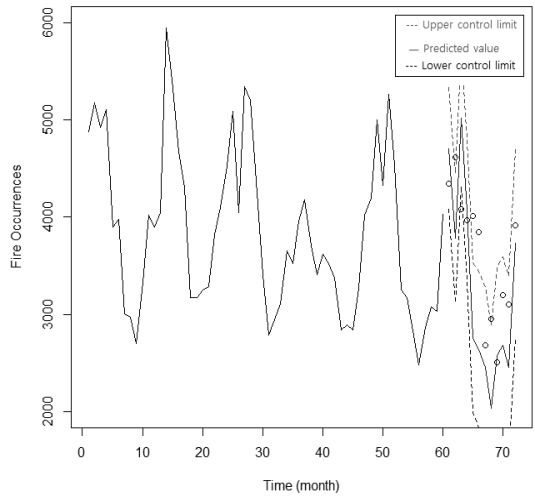


Fig. 7. Control chart for the number of fire occurrences of the year 2012.

Fig. 6에서 정규분포의 누적확률이 50%를 넘어가는  $mean \pm 0.68\sigma$ 를 관리한계선으로 나타내면 Fig. 7과 같다.

Fig. 7에서와 같이 6월에는 관리상한선을 초과하였고, 9월에는 관리하한선을 초과하여 여름의 화재 발생이 예상에서 벗어났다고 판단할 수 있었다.

추가로, 위에서 제안된 모델보다 타당한 모델이 있는지 검토하였다. 모델의 타당성을 비교하는 방법으로는, 2012년의 데이터를 검증(test) set으로 사용하여 ARIMA(p,d,q)(P,D,Q)<sub>12</sub>모델의 차수를 조정하는 방법도 있고, 더욱 낮은 AICc(Akaike information criterion with a correction for finite sample sizes)값을 가지는 모델을 채택하는 방법도 있다[11].

Fig. 8 및 Fig. 9는 모델의 계수를 변경하여 ARIMA(0,1,1)(0,2,1)<sub>12</sub>모델로 시도한 결과와 이것의 잔차의 자기상관함수를 도시한 것인데, 계산한 결과 AICc값은 기존 ARIMA(0,0,1)(0,1,1)<sub>12</sub>의 751.37보다 낮은 598.32를 얻을 수 있어서 더욱 좋은 모델이라고 판단하기 쉬우나, Fig. 9에서와 같이 잔류편차가 유의미한 자기상관을 가져서 좋은 모델로 보기 힘들다고 판단하였다. 이를 요약하여 정리하면 Table 2와 같다.

또한 기존 Shewhart chart는 95% 신뢰구간을 벗어났을 때 알람을 발생시켰었는데, 제품의 규격과 다르게 화재라는 사안의 특성을 감안할 때 너무 비현실적인 값이라 생각하여 50%로 알람 경계선을 결정하였다. 그 이유는 이 시스템은 천체지면에 준하는 이상 화재발생 현상에 대해서만 사용하는 것이

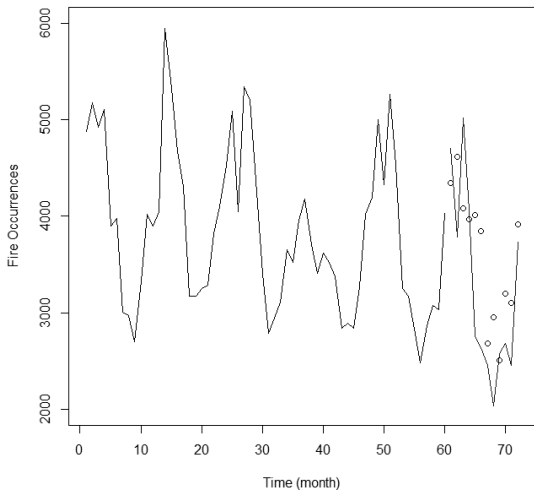


Fig. 8. Predicted result by model ARIMA (0,1,1)(0,2,1)<sub>12</sub>.

아닌, 보다 넓은 범위의 화재발생 현상에 적용시키기 위한 시스템이기 때문이다.

화재발생이 정규분포를 따른다고 가정하여 그를 논거로 한 관리도가 옳게 설계되었다면 더 적은 발생의 변화에도 빠르게 모델이 반응하는 특성은 정상 관리구간이 얼마나 좁은가로 판단할 수 있다. 즉, 합리적으로 설계된 모델은 원래 화재가 적게 발생하는 여름철은 더욱 다른 계절보다도 더욱 미소한 변화에도 반응할 수 있어야 하고, 반대로 많이 발생하는 봄과 가을철은 그보다 큰 변화에 반응한다.

2012년의 화재 발생건수에 대해 신뢰도 50%를 기준으로 한 기존 Shewhart 관리도와 ARIMA모델을 적용한 관리도의 정상관리구간을 나열하면 Table 3과 같다.

Table 3에서와 같이 여름철에는 정상관리구간이 좁게 줄어들어 작은 변화에도 빠른 반응을 보이고, 전반적으로 같은 신뢰도를 가짐에도 보다 적은 변화에 빠른 응답을 보임을 알 수 있다.

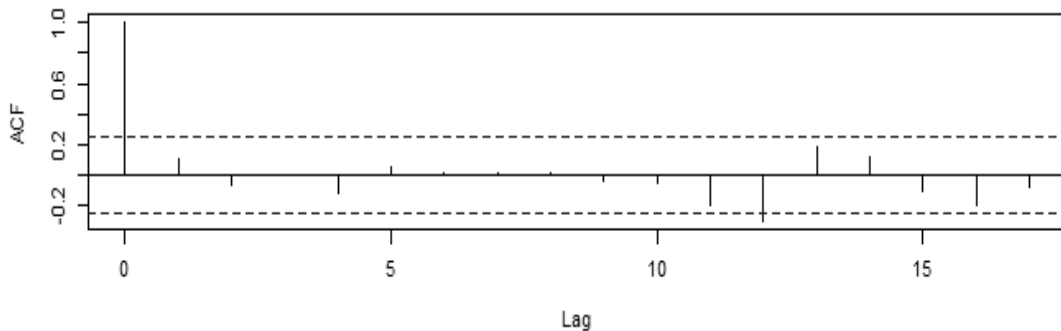


Fig. 9. ACF of residuals of ARIMA(0,1,1)(0,2,1)<sub>12</sub>.

Table 2. Comparison of ARIMA (0,0,1)(0,1,1)<sub>12</sub> and ARIMA (0,1,1)(0,2,1)<sub>12</sub>.

모델	ARIMA (0,0,1)(0,1,1) <sub>12</sub>	ARIMA (0,1,1)(0,2,1) <sub>12</sub>
Test set 정확도 $(\sum \frac{(P_{k,predicted} - P_k)^2}{n})$	매우 정확함 (163116.75)	정확함 (530324.58)
AICc	751.37	598.32
Ljung-box test	통과	통과
잔차의 ACF값	자기상관 없음	자기상관 있음

Table 3. Comparison of controlled limit width of two models

자기상관을 반영한 Shewhart 관리도			자기상관을 무시한 Shewhart 관리도		
LCL	UCL	구간의 길이	LCL	UCL	구간의 길이
4003.57	5045.06	1041.49	2724.90	4587.60	1862.70
4255.44	4887.30	631.87	2724.90	4587.60	1862.70
3718.44	5182.85	1464.42	2724.90	4587.60	1862.70
3606.44	4923.10	1316.66	2724.90	4587.60	1862.70
3653.44	4126.50	473.06	2724.90	4587.60	1862.70
3486.44	3746.14	259.70	2724.90	4587.60	1862.70
2321.44	3270.16	948.72	2724.90	4587.60	1862.70
2591.44	3222.61	631.17	2724.90	4587.60	1862.70
2145.44	3313.23	1167.80	2724.90	4587.60	1862.70
2838.44	3747.60	909.16	2724.90	4587.60	1862.70
2740.44	4028.83	1288.40	2724.90	4587.60	1862.70
3556.67	4482.35	925.68	2724.90	4587.60	1862.70

## V. 결론

본 연구에서는 ARIMA 모델에 기반한 화재 발생 빈도 예측모델을 제안하고, 2007년~2011년의 화재 데이터로부터 2012년 화재 발생건수를 예측하는 모델의 검증과정을 통해 제안방법의 타당성을 입증하였다. 그 결과, 모델의 적절한 차수를 결정하는 것이 타당한 모델을 만드는 데 중요하며, 차수의 선정기준 중에서 결과론적인 정확도, AICc값, Ljung-Box test의 p value 및 잔차의 ACF값의 자기상관성에 대하여 전반적으로 합리적인 선택이 있어야 함을 알 수 있었다. 또한 ARIMA(0,1,1)(0,2,1)<sub>12</sub> 모델은 ARIMA(0,0,1)(0,1,1)<sub>12</sub> 모델에 비해 AICc값이 더 낮았으나, ACF의 잔차값이 자기상관성을 가져서 합리적인 모델이라고 판단할 수 없었으나, ARIMA(0,0,1)(0,1,1)<sub>12</sub>은 보다 더 합리적인 모델이라고 판단되었다.

따라서 본 논문에서 제안한 ARIMA 모델의 설계 방법에 따라 결정된 차수를 활용하여 화재 데이터를 생성된 모델에 대입하여 존재하는 시점 이후의 시간에 대하여 발생할 화재 발생 건수를 예측하고, 실제로 시간이 경과하였을 때 실제 화재 발생건수와 예측으로부터 벗어난 정도를 판단하여 보다 합당한 화재 예방활동을 수행할 정책을 수립하는데 활용될 수 있을 것으로 판단한다.

## 감사의 글

본 연구는 국민안전처 사회재난안전기술개발사업의 일부로 ‘유해화학물질 효과적 대응을 위한 방재기술연구’ 과제[NEMA-인적\_2014-41]의 지원을 받아 수행되었기에 감사드립니다.

## REFERENCES

- [1] National Fire Incident Reporting System, Retrieved from <http://www.dfs.dps.mo.gov/documents/nfirs-reference-guide.pdf>
- [2] Justin, J. P., *Spatial and Temporal Patterns of Forest Fire Activity in Canada*, M.S. Thesis, University of Toronto, (2001)
- [3] Joo, K. D., *A Study on a Real-time Detection and Monitoring System for Abnormal Fire Occurrences Based on Big Data Mining*, M.S. Thesis, Myongji University, (2012)
- [4] Apley, D. W. and Lee, H. C., "Robustness Comparison of Exponentially Weighted Moving-Average Charts on Autocorrelated Data and on Residuals," *Journal of Quality Technology*, 40(4), 428-447, (2008)
- [5] Song, D. W., *Predicting the Risk of Fire Occur-*



- rence according to the Weather Information using Statistics and Data Mining Techniques, Ph.D Thesis, Seoul National University of Science and Technology, (2014)
- [6] Himmelblau, D. M., *Fault Detection and Diagnosis in Chemical and Petrochemical Processes*, Elsevier, pp.127-116, (1978)
- [7] Box, George E. P., *Time Series Analysis*, Wiley, (2008)
- [8] Hyndman, R. J., *Forecasting Functions for Time Series and Linear Models*, Retrieved from <http://cran.r-project.org/web/packages/forecast/index.html>, (2015)
- [9] Hyndman, R. J. and Khandakar, Y., "Automatic Time Series Forecasting: The forecast package for R", *Journal of Statistical Software*, **27**(3), 1-22, (2008)
- [10] Witten, I. A. *Data mining*, Elsevier, pp.148, (2011)
- [11] Lim, W. C., "Reliability-Based Design Optimization Using Akaike Information Criterion for Discrete Information", *Transactions of the Korean Society of Mechanical Engineers A*, **36**(8), 921-927, (2012)