

1970-2005년 동안의 산불 발생건수 및 연소면적에 대한 시계열모형 추정

李丙斗^{1*} · 鄭主相²

¹국립산림과학원 산불연구과, ²서울대학교 산림과학부

Estimation of Time Series Model on Forest Fire Occurrences and Burned Area from 1970 to 2005

Byungdoo Lee^{1*} and Joosang Chung²

¹Division of Forest Fire, Korea Forest Research Institute, Seoul 130-712, Korea

²Department of Forest Sciences, Seoul National University, Seoul 151-921, Korea

요 약: 효율적인 산불예방과 진화활동을 위해서는 산불 발생 및 확산특성에 대한 분석이 요구된다. 본 연구에서는 1970년부터 2005년까지 총 36년간의 월별 산불 발생건수 및 연소면적 자료를 이용하여 산불특성을 분석하고, Box-Jenkins의 ARIMA 모형을 이용하여 시계열 예측모형을 추정하였다. 분석 결과 발생건수 및 연소면적은 월별로 유의한 차이가 있었는데, 3월과 4월 두 달 동안 전체 발생건수의 59%, 연소면적의 72%가 집중되었다. 발생건수와 연소면적 모두에 있어서 ARIMA(1, 0, 1) 모형이 적합한 모델로 선정되었으며, 산불은 1개월과 12개월 전의 발생건수, 연소면적 수치와 밀접한 관련이 있음을 알 수 있었다.

Abstract: It is important to understand the patterns of forest fire in terms of effective prevention and suppression activities. In this study, the monthly forest fire occurrences and their burned areas were investigated to enhance the understanding of the patterns of forest fire in Korea. The statistics of forest fires in Korea, 1970 through 2005, built by Korea Forest Service was analyzed by using time series analysis technique to fit ARIMA models proposed by Box-Jenkins. The monthly differences in forest fire characteristics were clearly distinguished, with 59% of total forest fire occurrences and 72% of total burned area being in March and April. ARIMA(1, 0, 1) was the best fitted model to both the fire occurrences and the burned area time series. The fire time series have a strong relation to the fire occurrences and the burned area of 1 month and 12 months before.

Key words : forest fire, burned area, statistics, time series analysis, ARIMA

서 론

산불로 인한 사회적, 경제적 피해가 증가하면서 효율적인 산불예방 및 진화정책을 수립하기 위해 산불의 특성을 파악하는 것이 중요해지고 있다. 산불의 특성은 일정한 지역별 양상을 보여주는데 Gill(1975)은 이러한 양상을 산불형(fire regime)이라고 정의하였다. 산불형을 구성하는 요소로는 산불강도(intensity), 산불범위(extent), 산불빈도(frequency), 계절성(seasonality), 산불형태(type) 등을 들 수 있는데, 산불강도와 범위는 공간적 특성을 표현하며, 빈도, 계절성은 시간적 특성을 나타낸다. 산불형태는 지표화, 수관화 등 산불의 연소 부위와 밀접한 관련을 갖는다.

이러한 산불 특성을 파악하기 위해 발생건수, 연소면적 등의 산불 통계자료를 참고하는 것은 효율적이다(Barney and Stocks, 1983; Van Wagner, 1988; Kasischke *et al.*, 2002). 이것은 산불 통계자료의 분석을 통해 산불의 발생 및 확산특성을 파악함으로써 시공간상에서 한정된 진화자원을 효과적으로 활용하는데 도움이 되기 때문이다. 이렇듯 산불 통계자료는 산불정책의 기초자료를 제공한다. 는 점에서 다양하게 분석되고 있는데, 대개 월별, 연도별과 같은 시간 특성을 분석하기 위해 시계열 분석법이 이용되고 있다(Stephens, 2005).

현재 산불 주관부서인 산림청에서는 산불통계시스템을 운영하여 산불별로 발화원인, 발화 및 진화 일시, 발화장소, 진화자원 동원현황, 피해면적 등을 기록하고 있다(산림청, 2005). 따라서 본 연구에서는 이러한 자료를 이용하

*Corresponding author
E-mail: byungdoo@foa.go.kr

여 시간에 따른 산불의 발생 및 연소특성을 분석하고자 하였다. 이를 위해 1970년부터 2005년까지 36년 동안의 월별 산불 발생 및 연소면적의 시계열 특성을 분석해 보고 Box-Jenkins(Box and Jenkins, 1976)의 ARIMA 모형을 이용하여 예측모형을 개발하고자 하였다.

재료 및 방법

1. 산불 발생 및 연소면적 자료

산불의 시간적 특성 및 시계열 모형을 추정하기 위해 산림청의 산불대장과 통계자료를 바탕으로 발생건수 및 연소면적을 월단위로 수집하였다. 시간적 범위는 1970년부터 2005년까지 총 36년 동안으로, 자료 수는 각각 432(36년×12개월)개였다. 이 기간 동안 산불의 월별 차이가 있는지를 알아보기 위해 F-test와 Duncan의 다중범위검증(multiplexer test)(Duncan, 1955)을 수행하였다.

2. 시계열 모형의 개발

Box-Jenkins의 ARIMA 모형은 편리성과 유용성으로 인하여 여러 시계열 자료의 분석에 폭넓게 이용되고 있다(조신섭과 손영숙, 1999). ARIMA 모형에서 시계열 모형 구축 절차는 모형 식별, 모수 추정, 모형 진단으로 구분된다. 이 구축 절차에 따라 모형 식별 단계에서 식별 통계량을 이용하여 잠정 모형을 추정한 다음, 모수를 추정하였다. 마지막으로 모형 진단 단계에서 모형의 적합성을 판정하고 적합하지 않다고 판단되는 경우 모형 식별 단계부터 다시 반복하였다. ARIMA 모형에서는 시계열의 정상성(stationarity)을 가정하고 있으므로 산불 발생건수와 연소면적 자료에 대하여 백색잡음(white noise)과 단위근 검증을 통해 정상성 여부를 판단하였다. 정상성이 아닌 자료에 대해서는 로그변환을 통해 분산을 안정화시켜 정상 시계열로 변환하였다.

모형식별 단계에서 ARIMA 모형의 차수 p와 q는 자기상관그래프(autocorrelation graph)와 부분자기상관그래프(partial autocorrelation graph)를 보고 결정하였다. 여러 모형 중 AIC(Akaike information criterion)(Akaike, 1971; 1974)와 SBC(Schwartz bayesian criterion) (Schwartz, 1978) 통계량을 분석해 이 수치가 작은 모형을 선택하였다. 잠정 모형의 모수는 t값보다 클 유의확률을 산출해 유의성을 판단하였다. 모형의 진단을 위해서 적합 잔차의 임의성 여부를 포트맨토 검정법(Portmanteau test)을 이용해 분석하였다. 포트맨토 검정법의 결과 모형이 적합하지 않다고 판단되는 경우에는 최적 모형이 산출될 때까지 모형 식별 단계부터 다시 반복하였다. 모든 통계 분석은 SAS 9.1(SAS Institute Inc., 2003)을 이용하였다.

결과 및 고찰

1. 1970년부터 2005년 사이의 산불 발생 및 연소면적

1970년부터 2005년까지 36년 동안 산불 발생건수는 15,656건으로 101,785ha의 산림이 산불로 소실되었다. 연평균 435건이 발생하여 2,827ha가 연소되었다. F-test와 Duncan의 다중범위검증 결과 Figure 1 및 Table 1과 같이 월별 특성이 뚜렷하게 구별되는 것을 알 수 있었다. 즉, 3월과 4월 두 달 동안 전체 발생건수의 59%, 연소면적의 72%가 집중되었다. 반면에 7, 8, 9월에는 전체 건수의 0.9%만 발생하였고, 0.2%에 해당되는 면적만이 연소되어 대조되었다. 이 기간 동안 발생건수 및 연소면적 자료에서 통계적으로 유의한 선형추세는 관찰되지 않아 산불의 증가나 감소경향을 파악할 수 없었다. 다만, 일정 기간별로 구분하여 산불의 증감을 살펴보았을 때, 1980년부터 감소하다가 1996년 고성산불을 기점으로 증가하는 양상을 보였다.

2. 산불 발생건수에 대한 시계열 모형

산불 발생건수 시계열에 대하여 정상성인지를 판단하기 위해 백색잡음, 단위근, 계절 단위근을 검정한 결과, 모든 시차에서 유의확률이 0.001보다 작아(Figure 2) 발생건수가 백색잡음이라는 귀무가설을 기각하였다. 이것은 산불 발생건수 시계열에 대한 차분 혹은 계절 차분을 할 필요가 없음을 의미한다(조신섭과 손영숙, 1999).

산불 발생건수에 대한 ARIMA 모형의 차수 p, q는 Figure 3과 같이 ACF와 PACF의 패턴과 가장 작은 수치

Table 1. Monthly sum and mean number of fires and their burned area 1970 through 2005.

Month	No. of fire occurrences		Burned area (ha)	
	Total	Mean	Total	Mean*
Jan.	1,039 (6.6)	28.9 ^{D,E}	2,197 (2.2)	1.44 ^D
Feb.	2,036 (13.0)	56.6 ^C	5,426 (5.3)	1.81 ^C
Mar.	3,908 (25.0)	108.6 ^B	11,207 (11.0)	2.31 ^B
Apr.	5,353 (34.2)	148.7 ^A	61,909 (60.8)	2.73 ^A
May	1,312 (8.4)	36.4 ^D	12,406 (12.2)	1.60 ^{C,D}
June	263 (1.7)	7.3 ^F	624 (0.6)	0.70 ^F
July	36 (0.2)	1.0 ^F	73 (0.1)	0.19 ^G
Aug.	19 (0.1)	0.5 ^F	42 (0.0)	0.11 ^G
Sept.	96 (0.6)	2.7 ^F	84 (0.1)	0.23 ^G
Oct.	266 (1.7)	7.4 ^F	4,376 (4.3)	0.81 ^F
Nov.	550 (3.5)	15.3 ^{E,F}	1,219 (1.2)	1.13 ^E
Dec.	778 (5.0)	21.6 ^{D,E,F}	2,223 (2.2)	1.33 ^{D,E}

^{A-G} Mean values in the column by the same letter are not significantly different ($P < 0.05$).

*Mean of $\log_{10}(\text{burned area}(\text{ha})+1)$.

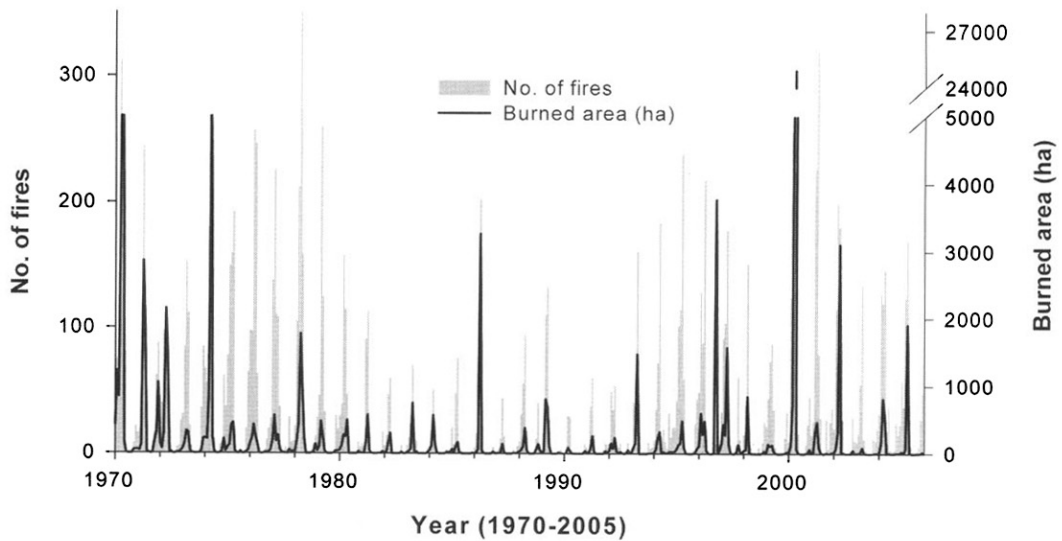


Figure 1. Time series plot of the number of fire occurrences and their burned area 1970 through 2005.

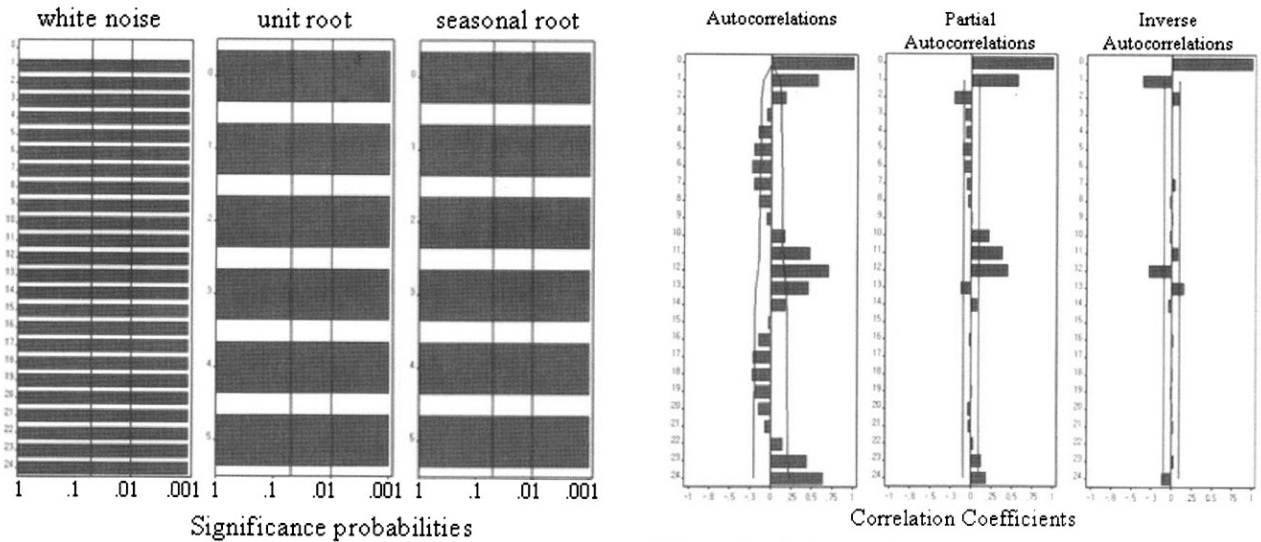


Figure 2. White noise and unit root test of forest fire occurrences.

Figure 3. ACF, PACF and IACF graph of forest fire occurrences.

를 보이는 AIC와 SBC 통계량 결과에 따라 ARIMA(1, 0, 1) 모형으로 결정하였다. ARIMA 모형의 차수 중 차분을 의미하는 d는 발생건수 시계열이 정상성을 가진 자료 분포를 보이므로 0으로 설정하였다(최병선, 1995). ACF와 PACF 분석 결과 발생건수 시계열은 과거 관측값(1, 12번째)과 과거 오차(1, 12번째)에 크게 영향을 받는다는 것을

알 수 있었다. 이는 현재 달의 산불 발생은 1년 전과 바로 전달의 산불 상황에 영향을 받는다는 의미로써 산불의 뚜렷한 월별 특성을 보여주었다.

ARIMA(1, 0, 1) 모형을 이용하여 모수를 추정한 결과 AR 부분의 첫 번째 모수가 유의하지 않음($Pr > t=0.3613$)에 따라 제외시킨 다음 모형을 재추정하였다. 이러한 과

Table 2. ARIMA model parameter estimates of forest fire occurrences and the statistics of AIC and SBC.

Parameter	Estimate	Standard error	t-value	Pr > t
Constant	47.99082	9.61190	4.99	<.0001
MA1,1 (lag=1)	-0.32767	0.04057	-8.08	<.0001
MA1,2 (lag=12)	0.45772	0.04950	9.25	<.0001
AR1,1 (lag=12)	0.89317	0.02945	30.33	<.0001
AIC	4,340.125			
SBC	4,356.390			

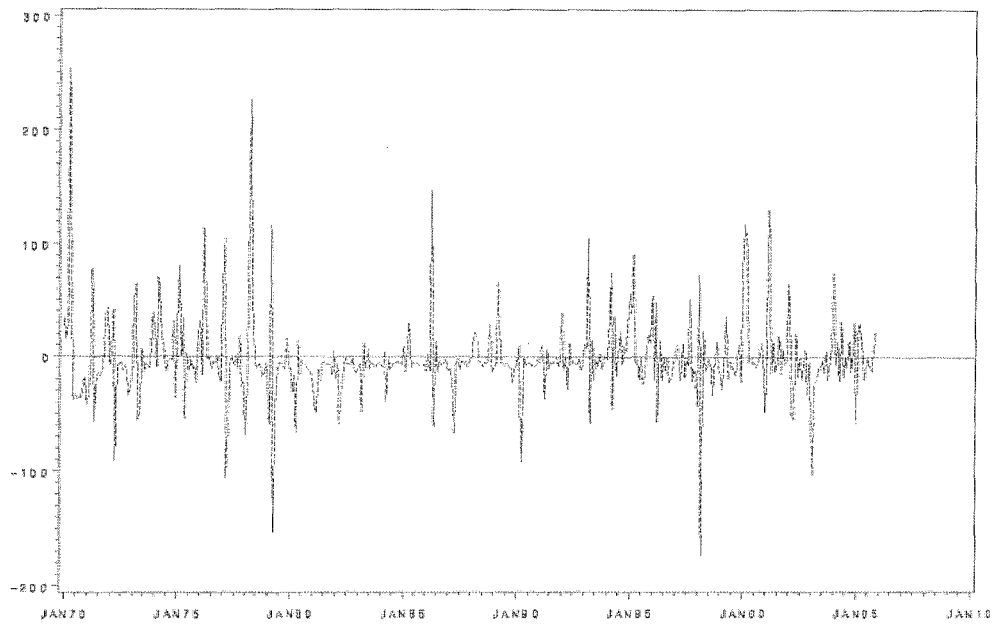


Figure 4. Residual plot for the time series model of forest fire occurrences.

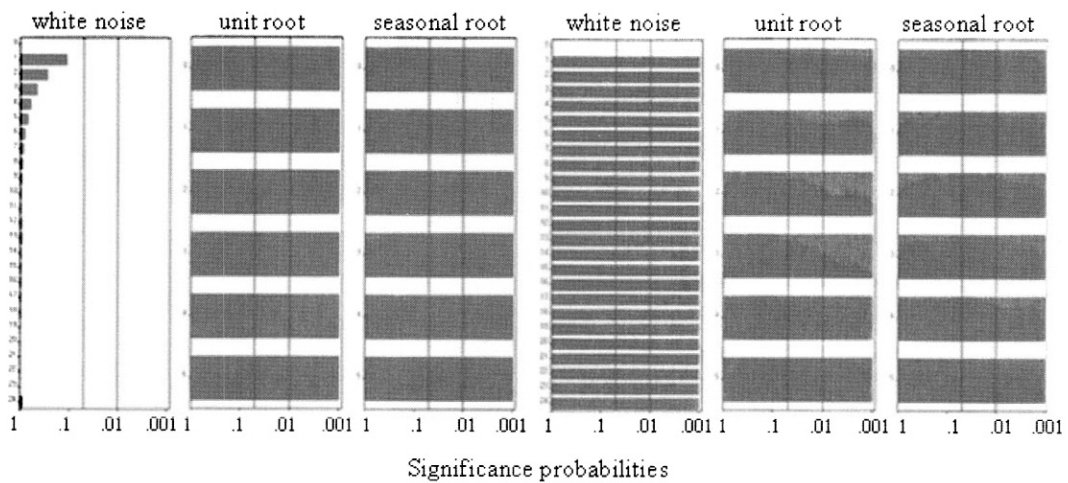


Figure 5. White noise and unit root test, (left) burned area, (right) $\log_{10}(\text{burned area} + 1)$.

정을 거쳐 생성된 발생건수 시계열 모형은 식 (1)과 같다. 이 모형을 통해 현재의 발생건수(Z_t)는 시차 1과 시차 12에서의 발생건수와 밀접한 연관을 갖고 있는 것으로 분석되었다. Table 2는 추정된 모형의 인수를 보여주고 있는데, 상수항을 포함한 모든 인수들이 유의함을 알 수 있다.

$$(1 - 0.89317B^{12})Z_t = 47.99082 + (1 + 0.32767B - 0.45772B^{12})\epsilon_t \quad (1)$$

잔차에 대한 포트맨토 검정 결과에서 유의확률이 모두 0.05보다 큰 것으로 나타났다. 따라서 모든 시차에서 자기상관계수들이 동시에 '0'이라는 가설을 기각할 수 없으므로 잔차들을 백색잡음으로 판정할 수 있다(Schwartz, 1978). 또한 잔차의 시계열도(Figure 4)와 ACF와 PACF 분석 결과 역시 잔차들이 백색잡음인 것으로 확인되어 추

정된 모형이 적합한 것으로 나타났다.

3. 연소면적에 대한 시계열 모형

산불의 연소면적에 관한 시계열은 Figure 5의 왼쪽에서 보는 바와 같이 백색잡음 검정에서 유의확률이 높아 정상성을 가정할 수 없었다. 따라서 $\log(\text{연소면적} + 1)$ 변환을 통해 분산의 안정화를 꾀했다(최병선, 1995). 로그 변환 결과 Figure 5의 오른쪽 그림에서와 같이 정상성 조건을 만족하였다.

연소면적에 대한 ARIMA 모형의 차수 p, q 는 발생건수 시계열모형에서의 마찬가지로 ACF와 PACF의 패턴(Figure 6)과 가장 작은 수치를 보이는 AIC와 SBC 통계량 결과에 따라 ARIMA(1, 0, 1) 모형으로 결정하였다.

ACF 그래프를 분석한 결과 계절특성이 파악되어 연소면적 시계열에 대하여 계절차분과 1차, 2차 차분을 조합

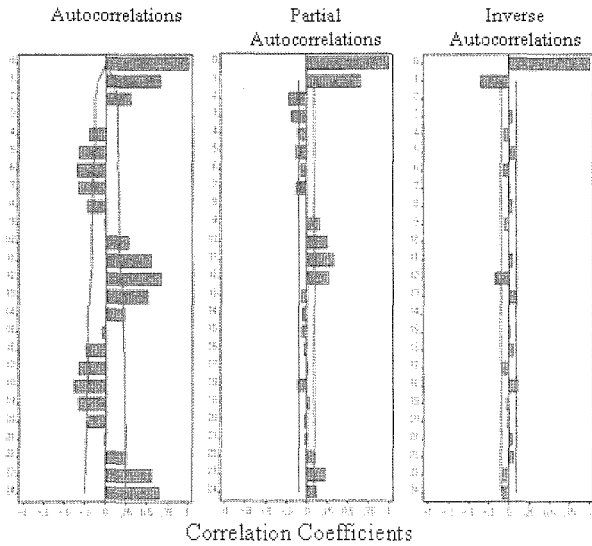


Figure 6. ACF, PACF and IACF graphs of $\log_{10}(\text{burned area}(ha)+1)$.

하여 잠정 모형을 고려해 봤으나 모형 판정단계에서 유의하지 않은 것으로 분석되었다. 아울러 시차 2, 3, 10, 11 모수가 유의하지 않음에 따라 연소면적에 대한 최종 시계열 모형은 발생건수 시계열 모형과 동일한 형태로 선정되

었다. 즉 현재 달의 로그 변환된 연소면적(Z_t)은 AR부분에서 시차 12에서의 연소면적과 MA 부분에서 시차 1, 12의 연소면적에 영향을 받아 산출되며 상수항은 3.72로 나타났다. 추정된 시계열 모형의 인수가 Table 3에 제시되어 있는데 모든 인수들이 유의했으며 모형의 AIC, SBC 통계량은 각각 1,527 및 1,543으로 분석되었다.

$$(1 - 0.9506B^{12})Z_t = 3.72189 + (1 + 0.2683B^1 - 0.63483B^{12})\epsilon_t \quad (2)$$

Where, $Z_t = \log_{10}(\text{burned area}_t(ha)+1)$

연소면적에 대해 추정된 모형이 적절한지를 판단하기 위해 포트맨토 검정과 잔차의 시계열도(Figure 7), ACF와 PACF를 분석한 결과 유의확률이 큰 것으로 나타나 잔차들이 백색잡음이라는 가설을 만족할 수 있었다. 따라서 추정된 모형이 통계적으로 적합하다는 결론을 내릴 수 있었다. 추정된 시계열 모형을 살펴보면 산불발생 모형과 마찬가지로 1년 전과 바로 전 달의 수치가 현재 달의 연소면적과 밀접한 관련이 있음을 알 수 있었다. 이는 월별 산불발생 특성이 뚜렷한 우리나라의 산불 환경(Lee et al., 2006)을 반영한다고 볼 수 있다.

Table 3. ARIMA model parameter estimates of $\log_{10}(\text{burned area}(ha)+1)$ and the statistics of AIC and SBC.

Parameter	Estimate	Standard error	t-value	Pr > t
Constant	3.72189	0.38796	9.59	<.0001
MA1,1 (lag=1)	-0.26830	0.03601	-7.45	<.0001
MA1,2 (lag=12)	0.63483	0.04190	15.15	<.0001
AR1,1 (lag=12)	0.95060	0.02166	43.89	<.0001
AIC	1,527.031			
SBC	1,543.296			

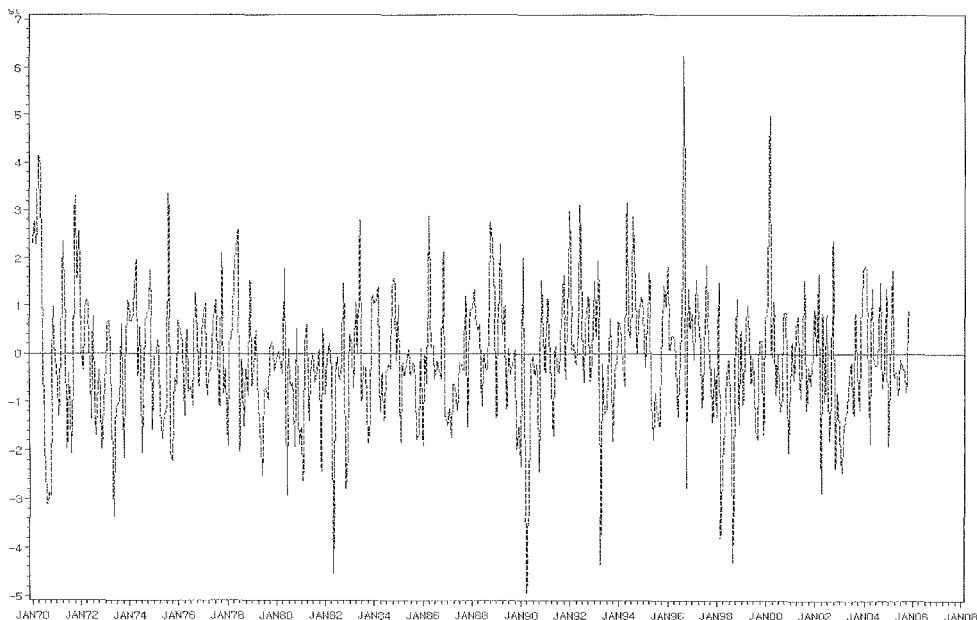


Figure 7. Residual plot for the time series model of $\log_{10}(\text{burned area}(ha)+1)$.

결 론

본 연구에서는 1970년부터 2005년까지 36년 동안 월단위로 수집된 산불 발생건수와 연소면적 자료를 바탕으로 산불 발생 및 연소면적 특성을 분석하고, Box-Jenkins ARIMA 모형을 이용하여 시계열 예측 모형을 추정하였다. 분석 결과 발생건수 및 연소면적에 있어서 월별로 유의한 차이가 있었으며 4월과 3월에 가장 많이 발생했고 연소면적 또한 컸다. 발생건수와 연소면적 모두에 있어서 ARIMA(1, 0, 1) 모형이 적합한 예측 모델로 선정되었으며, 현재의 발생건수 및 연소면적은 1개월과 12개월 전의 수치를 밀접한 관련이 있음을 알 수 있었다. 이러한 연구 결과는 단기보다는 장기 산불정책을 수립하는데 있어서 요구되는 산불특성에 관한 이해와 산불 발생 및 피해면적 예측을 위해 활용될 수 있을 것이다. 차후의 연구에서는 횡단면적 분석을 통해 산불 발생과 확산에 영향을 미치는 것으로 알려진 기상(Schoenberg *et al.*, 2003; Crimmins and Comrie, 2004), 지형, 임상 요인(Mouillot *et al.*, 2003; Schoenberg *et al.*, 2003; 이병두, 2005) 및 사회, 경제적 요인(Martell *et al.*, 1987; Pew and Larsen, 2001)들과의 유기적 연계성을 이해하기 위한 분석이 필요할 것으로 판단된다.

인용문헌

1. 산림청. 2005. 2005년 산불통계연보. pp. 187.
2. 이병두. 2005. GIS와 RS를 이용한 2000년 삼척산불 행동 특성 분석 및 산불확산예측모델 개발. 서울대학교 박사학위 논문. pp. 117.
3. 조신섭, 손영숙. 1999. SAS/ETS를 이용한 시계열분석. 울곡출판사. pp. 430.
4. 최병선. 1995. 다변량 시계열분석. 세경사. pp. 1051.
5. Akaike, H. 1971. Information theory and an extension of the maximum likelihood principle. pp. 267-281. In : Petrov, B.N. and Csaki, F., ed. 2nd international symposium on information theory. Institute of Statistical Mathematics. Research Memorandum No. 46.
6. Akaike, H. 1974. A new look at the statistical model identification. IEEE Transactions on Automatic Control 79: 716-723.
7. Barney, R.J. and Stocks, B.J. 1983. Fire frequencies during the suppression period. pp. 45-62. In : Wein, R.W. and MacLean, D.A., The role of fire in northern circum-polar ecosystems. John Wiley and Sons. New York.
8. Box, G.E.P. and Jenkins, G.M. 1976. Time series analysis: Forecasting and control. Holden-Day, San Francisco.
9. Crimmins, M.A. and Comrie, A.C. 2004. Interactions between antecedent climate and wildfire variability across southeastern Arizona. International Journal of Wildland Fire 13: 455-466.
10. Duncan, D.B. 1955. Multiple range and multiple F tests. Biometrics 11: 1-42.
11. Gill, A.M. 1975. Fire and the Australian flora : a review. Australian Forestry 38: 4-25.
12. Kasischke, E.S., Williams, D. and Barry, D. 2002. Analysis of the patterns of large fires in the boreal forest region of Alaska. International Journal of Wildland Fire 11: 131-144.
13. Lee, B., Park, P.S. and Chung, J. 2006. Temporal and spatial characteristics of forest fires in South Korea between 1970 and 2003. International Journal of Wildland Fire 15(3): 389-396.
14. Martell, D.L., Otukol, S. and Stocks, B.J. 1987. A logistic model for predicting daily peoplecaused forest fire occurrence in Ontario. Canadian Journal of Forestry Research 17: 394-401.
15. Mouillot, F., Ratte, J., Joffre, R., Moreno, J.M. and Rambal, S. 2003. Some determinants of the spatiotemporal fire cycle in a Mediterranean landscape (Corsica, France). Landscape Ecology 18: 665-674.
16. Pew, K.L. and Larsen, C.P.S. 2001. GIS analysis of spatial and temporal patterns of humancaused wildfires in the temperature rain forest of Vancouver Island, Canada. Forest Ecology and Management 140: 1-18.
17. SAS Institute Inc. 2003. SAS 9.1 Help and Documentation. SAS Institute Inc. Cary, N.C., USA.
18. Schoenberg, F.P., Peng, R., Huang, Z. and Rundel, P. 2003. Detection of nonlinearities in the dependence of burn area on fuel age and climate variables. International Journal of Wildland Fire 12: 1-6.
19. Schwartz, G. 1978. Estimating the dimension of a model. Annals of Statistics 6: 461-464.
20. Stephens, S.L. 2005. Forest fire causes and extent on United States forest service lands. International Journal of Wildland Fire 14(3): 213-222.
21. Van Wagner, C.E. 1988. The historical pattern of annual area burned in Canada. The Forestry Chronicle 64: 182-188.