

# Analysis of the Korean peninsula precipitation using inverse statistics methodology

Seungsik Min<sup>a,1</sup>

<sup>a</sup>Department of Natural Science, Korea Naval Academy

(Received December 10, 2015; Revised January 25, 2016; Accepted February 27, 2016)

---

## Abstract

In this paper, we analyze the inverse statistics of rainfall for 12 regions from 1973 to 2014. We obtain a probability density function  $f(x)$  of daily rainfall  $x$ , and  $f(\tau_\rho)$  of the first passage time  $\tau_\rho$  for a given  $\rho$ . Lastly, we derive the relation between  $\rho$  and  $\tau_{\text{mean}}(\rho)$ , i.e., the averaged value of  $\tau_\rho$ . The analyses result in the  $x$  and  $\tau_\rho$  have stretched exponential distributions. Also,  $\tau_{\text{mean}}(\rho)$  has the form of a stretched exponential function. We derive the shape parameter  $\beta$  of the distribution, and analyze the characteristics of 12 regional rainfalls.

Keywords: precipitation, rainfall, inverse statistics, probability density function, stretched exponential distribution

---

## 1. 서론

시계열에 대한 일반적인 분석은 시간을 독립변수로 나타내고, 시계열의 변동성을 종속변수로 나타낸다. 이에 반해 역통계(inverse statistics)는 독립변수와 종속변수가 반대로 된 경우를 다룬다. 가령, 주가지수 시계열의 경우 일반적으로 정해진 시간에 대한 수익률에 관심을 가지고 여러 가지 유의미한 값들을 도출해 낸다. 반면, 역통계 분석에서는 수익률을 독립변수로 두고, 이러한 수익률에 도달하기까지 걸리는 시간을 종속변수로 하여 분석을 실시한다. 따라서 역통계(inverse statistics)에 대한 연구는 첫 통과 시간(first passage time)에 대한 연구와 거의 동일시되며 통계물리학이나 응집물리학계를 중심으로 발전되어 왔다.

Honerkamp와 Baumgärtner (1986)는 임의보행 운동의 첫 통과시간이 분산에 비례한다는 이론이 자기 유사성을 지닌 계층적 구조에서는 일반적으로 성립하지 않는 현상을 해석적으로 보였고, Ding과 Yang (1995)은 브라운운동에서 정해진 목표값에 최초로 도달하는 시간을 분석하였다. 또한 Jensen (1999)과 Rangarajan과 Ding (2000a, 2000b)은 각각 난류와 확산 문제에 대해 첫 통과시간의 분포를 연구하였다. 이어 Rangarajan과 Ding (2003), Ding과 Rangarajan (2004) 그룹은 주로 확산 문제에 대해 해석적인 방법으로 첫 통과시간을 계산하였고, Simonsen 등 (2002)과 Jensen 등 (2003, 2004) 그룹은 주로 주가지수, 환율 등을 중심으로 한 금융 시계열의 첫 통과시간을 실증적으로 분석하였다. 투자 지표에 대한 역통계량 산출이 활발하게 이루어지는 것은 정해진 시간 간격에서의 수익률 못지않게 목표 수익률에

---

<sup>1</sup>Department of Natural Science, Korea Naval Academy, 1 Jungwon-ro, Changwon-si, Gyeongsangnam-do 51704, Korea. E-mail: [fieldsmin@gmail.com](mailto:fieldsmin@gmail.com)



\* 출처 : 국토교통부 국토지리정보원 지도검색서비스, <http://www.ngii.go.kr>

**Figure 2.1.** Investigated regions of rainfall. We partitioned the Korean peninsula into 3 by 3, and then selected a region in each sector. To avoid a bias, we selected mountainous terrain, basin and sea, evenly. Finally, we investigated three islands further.

도달하기까지 걸리는 시간이 투자자들의 주된 관심거리이기 때문이다. 한편, Lee 등 (2008)은 한국의 종합주가지수를 분석하여 역통계량이 멱함수를 이루는 것을 보였다.

본 연구에서는 역통계 방법론을 이용하여 한반도의 강수 특성을 분석하였다. 대한민국은 여름철에는 홍수 피해가, 겨울철에는 가뭄 피해가 빈번하다. 그래서 역통계량 분석을 통해 특정 강수량이 도달하기까지 걸리는 예상 시간을 분석한다면 홍수나 가뭄으로 인한 피해를 예방하는 데에 도움이 될 것으로 사료된다.

## 2. 분석 데이터

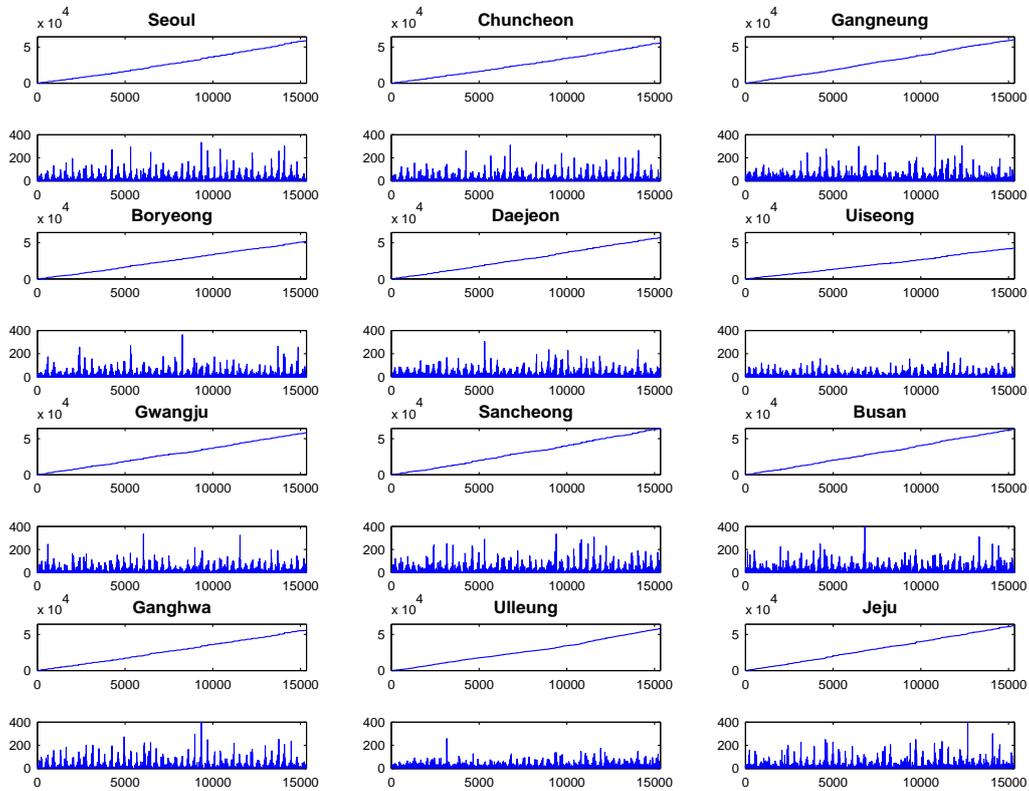
### 2.1. 조사 대상 및 기간

본 연구에서는 한반도의 지형을 9등분하여 북서 지역의 '서울'에서부터 남동 지역의 '부산'까지 총 9개 지역과 '강화도', '울릉도', '제주도'의 3개 섬의 강수량을 대상으로 하였다 (Figure 2.1). Table 2.1에서 보는 바와 같이 산, 분지, 바다, 섬 등 4가지 항목으로 분석하여 보면 선정된 지역은 일부 특성에 편중되지 않고 다양한 지형적 특성을 가지고 있는 것을 확인할 수 있다. 따라서 이들 지역에서 관측된 강수량은 대한민국의 기상 현상을 대표한다고 말할 수는 없으나 비교적 골고루 나타난다고 볼 수 있다. 조사 기간은 12개 지역의 강수량 데이터가 모두 기록되기 시작한 1973년 1월 1일부터 2014년 12월 31일까지 42년 간, 15,340일을 대상으로 하였다.

Figure 2.2는 조사대상 지역 12개의 누적 강수량 시계열과 일일 강수량 시계열을 나타낸 것이다. 일반적으로 시계열은 누적된 값이므로 본 논문에서는 누적된 값을 강수량 시계열이라 명명한다. 7월-9월의

**Table 2.1.** Location and basic terrain types in the surveyed area

지역	연평균 강수량 (mm)	관측시작일	위치	산	분지	바다	섬
서울	1,395	1960년 이전	북서	X	X	X	X
춘천	1,318	1966. 1. 1.	북중	X	O	X	X
강릉	1,429	1960년 이전	북동	X	X	O	X
보령	1,234	1973. 1. 1.	중서	O	X	O	X
대전	1,357	1969. 1. 1.	중중	X	X	X	X
의성	1,012	1973. 1. 1.	중동	O	O	X	X
광주	1,386	1960년 이전	남서	X	X	X	X
산청	1,529	1973. 1. 1.	남중	O	X	X	X
부산	1,507	1960년 이전	남동	X	X	O	X
강화도	1,324	1973. 1. 1.	북서	X	X	O	O
울릉도	1,382	1960년 이전	중동	O	X	O	O
제주도	1,480	1960년 이전	남서	X	X	O	O



**Figure 2.2.** Precipitation time series data of 12 regions from January 1, 1973 to December 31, 2014. The figures above show the cumulative rainfalls, below show the daily rainfalls.

여름철에 강수가 집중되는 대한민국 기후 특성 상 일일 강수량은 계절 주기에 따라 등락이 규칙적으로 반복되고 있으며 ‘산청, 부산, 제주’의 강수량이 비교적 많고, ‘의성’과 ‘보령’의 강수량이 비교적 적은 것을 확인할 수 있다.

### 3. 배경 이론

#### 3.1. Stretched exponential distribution

자연계의 여러 현상에 대한 가장 전통적인 관점이 정규분포(Gaussian distribution)를 이룬다는 것이다. 하지만 근래에는 정규분포보다 두터운 꼬리(heavy-tail)를 갖는 현상들이 많이 알려지고 있다. 무기억성(memoryless property)을 가지고 있는 지수분포(exponential distribution)나, 척도없는(scale-free) 특성을 가지고 있는 멱함수 분포(power-law distribution)가 대표적인 예에 해당한다. 한편, 지수족(exponential family)에 속하는 함수들 중 다음의 특성을 만족하는 분포를 확장된 지수분포(stretched exponential distribution)라고 표현한다 (McCauley, 2004).

$$f_{\beta}(x) \sim e^{-ax^{\beta}}$$

이 때,  $\beta = 2$ 인 경우 정규분포를 의미하고,  $\beta = 1$ 인 경우 지수분포를 의미한다.  $\beta > 1$ 일 때 압축된 지수분포(compressed exponential distribution)로 따로 명명하기도 한다. 확장된 지수분포는  $\beta$ 값이 적을수록 두터운 꼬리를 가지지만, 멱함수 분포보다는 얇은 꼬리를 가진다. 확장된 지수분포는 와이불분포(Weibull distribution)의 누적분포함수(cumulative distribution function)와 관련이 있다. 즉,

$$\begin{aligned} f(x) &= a\beta x^{\beta-1} e^{-ax^{\beta}}, \\ F(x) &= P(X \leq x) = 1 - e^{-ax^{\beta}}, \end{aligned}$$

여기서  $f(x)$ 는 확률밀도함수,  $F(x)$ 는 누적분포함수이다.

강수량에 대한 근래의 연구들은 특정 기간 동안의 강수량 데이터는 확장된 지수분포를 따르는 것으로 인식되고 있다 (Wilson과 Toumi, 2005; Lana 등, 2009). 달리 표현하면 매 순간순간의 강수속도가 와이불분포를 따른다는 것인데, 앞으로 강수가 일어날 확률이 직전 날씨에 영향을 받는 것으로부터 이유를 찾을 수 있다.

#### 3.2. First passage time of rainfall

본 연구에서 사용하는 데이터는 12개 지역의 일일 강수량이다. 시간  $t$ 에서의 강수량을  $S(t)$ 라고 할 때, 시간  $t_1, t_2$ 에서 강수량이 문턱값  $\rho$ 를 넘어섰다면 우리는 첫 통과시간(first passage time)을 두 사건 사이의 시간 간격으로 정의한다. 즉,

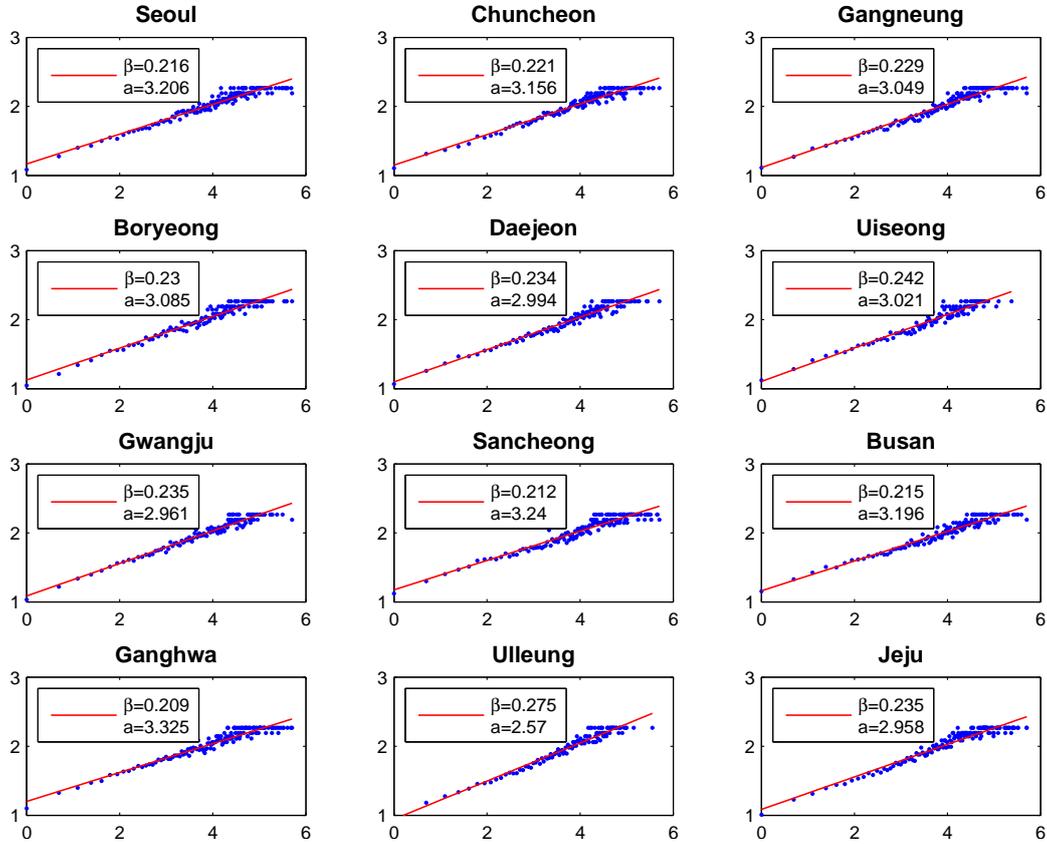
$$\tau_{\rho} = t_2 - t_1, \quad \text{where } S(t_1) \geq \rho, S(t_2) \geq \rho, \text{ and } \max_{t_1 < t < t_2} \{S(t)\} < \rho.$$

만약 원 시계열이 지수분포를 이룬다면 무기억성을 가지고 있으므로, 첫 통과시간의 분포 역시 지수분포를 이루게 된다. 하지만 일반적인 분포의 경우 첫 통과시간에 대한 분포는 자명하지 않은 분포를 이룬다.

주가 수익률에 대한  $\tau_{\rho}$ 의 분포는  $0 \leq \rho \leq 0.08$ 인 경우 멱함수 분포를 따르는 것으로 알려져 있다. 이 경우 시계열은 브라운 운동(Brownian motion)을 따르는 것에 기인한다 (Ding과 Rangarajan, 2000b).

$$p(\tau_{\rho}) \propto \tau_{\rho}^{-\alpha}.$$

이 때  $\alpha$ 는 멱함수 분포의 지수로  $\rho$  값에 관계된다. 만약  $\rho$  값이 증가한다면  $\tau_{\rho}$ 의 분포는 주봉(mode)를 갖는 감마 분포(gamma distribution)를 따르는 것으로 알려져 있다 (Simonsen 등, 2002; Ding과 Ran-



**Figure 4.1.** Probability density functions of cumulative rainfall for 12 regions. The  $x$ -axis means  $\ln x$  and  $y$ -axis means  $\ln\{-\ln[f(x)]\}$ , where  $x$  is a daily rainfall and  $f(x)$  is the probability density function of  $x$ . They show the features of linearity, i.e., their distributions are stretched exponential(The best fit line  $\ln\{-\ln[f(x)]\} = c_1 + c_2 \ln x$  means  $\ln[f(x)] = -e^{c_1+c_2 \ln x} = -e^{c_1}x^{c_2}$ , so  $f(x) = e^{-e^{c_1}x^{c_2}} = e^{-ax^\beta}$ . Therefore,  $a = e^{c_1}$  and  $\beta = c_2$  for stretched exponential distribution). The coefficients of determination for linear regressions are between 0.936 and 0.958.

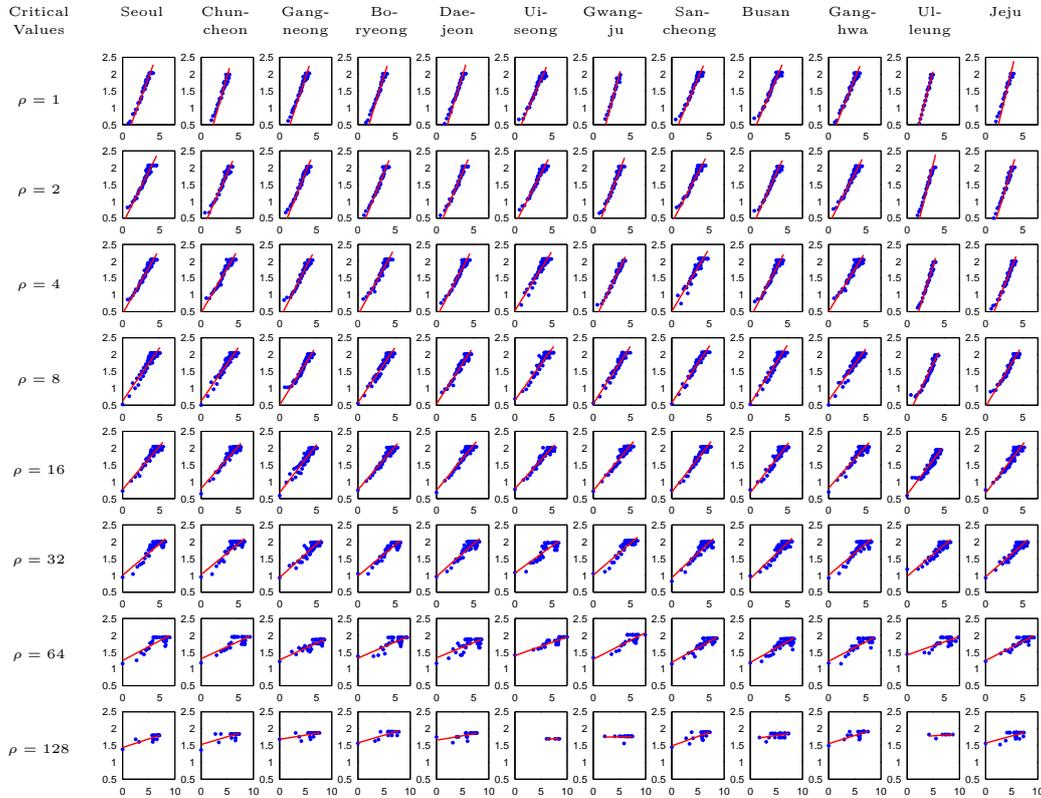
garajan, 2000a).

$$p(\tau_\rho) = \frac{\beta}{\sqrt{\pi}} \tau_\rho^{-\alpha} e^{-\frac{\beta^2}{\tau_\rho}}$$

한편, 강수량에 대한  $\tau_\rho$ 의 분포는 크게 알려진 바가 없다. 브라운 운동에서 변동률  $\rho$  및 첫 통과시간  $\tau_\rho$ 가 멱함수 분포를 이룬다고 앞에서 언급하였다. 마찬가지로 일일 강수량이 확장된 지수분포를 따르는 사실로부터 착안하여 이번 연구에서 첫 통과시간  $\tau_\rho$ 의 분포 역시 확장된 지수분포를 이룬다고 가정하여 분석을 실시하였다.

#### 4. 분석 결과

누적 강수 시계열의 변화율, 즉 일일 강수량이 특정 값  $\rho$ 를 넘어서기까지의 시간(first passage time)



**Figure 4.2.** Probability density functions of the first passage time for 12 regions. The  $x$ -axis means  $\ln \tau_\rho$  and  $y$ -axis means  $\ln\{-\ln[f(\tau_\rho)]\}$ , where  $x$  is a daily rainfall and  $f(x)$  is the probability density function of  $x$ . They show the features of linearity for  $\rho \leq 16\text{mm}$  because the smallest coefficient of determination for  $\rho \leq 16\text{mm}$  is 0.878, i.e., their distributions are stretched exponential on the range. However, the linearity begins to violate for  $\rho \geq 32\text{mm}$ . Nevertheless, the linearity property is still powerful on  $\rho = 32\text{mm}$  because the coefficients of determination for linear regressions are between 0.777 and 0.871. We may consider that there are phase transitions for the probability density functions around  $\rho = 32\text{mm}$ .

$\tau_\rho$ 를 구하는 것이 역통계 분석의 목적이다. 이 때  $\tau_\rho$ 는 특정 값이 아니라 확률밀도함수 형태의 분포를 이루게 된다. 따라서 우리는 몇몇  $\rho$  값들에 대해  $\tau_\rho$ 의 확률밀도함수  $f(\tau_\rho)$ 를 구하고 분포 특성을 분석하였다 (Figure 4.2). 이후 각각의 경우  $\tau_\rho$ 의 대푯값인  $\tau_{\text{mean}}(\rho)$ 를 구하여  $\rho$ 와의 관계를 도출하였다 (Figure 4.3).  $f(\tau_\rho)$ 의 분포 특성을 뒷받침하기 위해 기존의 방법으로 일일 강수량  $x$ 의 확률밀도함수  $f(x)$ 를 구하고 분포 특성을 분석하였다 (Figure 4.1).

Figure 4.1은 조사 대상 지역 일일 강수량의 확률밀도함수  $f(x)$ 의 그래프이다. 본 논문에 표현하지는 않았으나, 일일 강수량  $x$ 가 증가할수록  $f(x)$ 는 감소하며 그 감소 정도는 멱함수보다는 크고, 지수함수보다는 작다. 따라서 우리는  $f(x) \sim e^{-ax^\beta}$ 의 확장된 지수분포(stretched exponential distribution)를 가정하여 모수인  $a$ 와  $\beta$  값을 추정하였다. Figure 4.1에서 가로축은  $\ln x$ 를, 세로축은  $\ln\{-\ln[f(x)]\}$ 를 의미하며, 모든 지역에서 그래프가 직선형을 띠는 것을 확인할 수 있다. 따라서 이들 지역의 적합선(fitted line) 기울기와  $y$ 절편을 구하면  $\beta$ 와  $a$  값을 추정할 수 있다. 여기서  $\beta$ 는 형태 모수(shape

parameter),  $a$ 는 크기 모수(scale parameter)라고 부른다. 상대적으로 꼬리가 두터울수록(heavy-tail),  $\beta$  값은 작아지고 따라서  $a$  값은 커지게 된다. 그림에서 보면 조사 지역 중 강수량이 가장 많은 '산청, 부산'의  $\beta$  값이 작은 편에 속하는 것을 알 수 있다. 이것은 어느 정도 추정 가능한데 강수량이 많은 지역일수록 폭우가 많이 내리고, 따라서 확률밀도함수의 꼬리가 두터워져서  $\beta$  값이 낮게 나온 것으로 추정된다. 그러나 세 번째로 비가 많이 오는 지역인 '제주도'의  $\beta$  값은 보통 수준인데 이는 섬 지역의 특성상 비나 눈이 자주 오면서도 폭우의 비율이 상대적으로 낮은 것으로부터 유래했다고 추정된다. 반대로 강수량이 가장 낮은 지역인 '의성'의  $\beta$  값은 예상대로 높은 편에 속해 강수량이 낮은 것은 물론 폭우의 비율도 낮은 것을 알 수 있다. 그러나 대구 지역과 소우 지역의  $\beta$  값이 최소와 최대값을 나타내지는 않는 것을 알 수 있다. '강화도'의  $\beta$  값이 전 지역에서 가장 낮고, '서울' 역시 낮은 편에 속하는 것은 중부 지방의 여름철 강수 집중률과 관련되어 있는 것으로 추정할 수 있다. 따라서 이들 지역은 잠재적인 홍수 피해에 대비해야 할 것이다. 흥미로운 것은 '울릉도'의  $\beta$  값이 0.275로 타 지역에 비해 특히 높은 점이다. '울릉도'의 강수량이 조사 대상 지역 중 중간 정도에 속하는 데 반해  $\beta$  값이 매우 크게 나타나는 것은 울릉도의 겨울철 강수 집중률이 타 지역에 비해 월등히 높기 때문인 것으로 추정된다. 즉, 대표적인 다설 지역인 울릉도는 여름과 겨울의 강수 편차가 크지 않기 때문에  $\beta$  값이 크게 나타나는 것으로 여겨진다.

Figure 4.2는 임계값  $\rho$ 에 따른 첫 통과시간의 확률밀도함수  $f(\tau_\rho)$ 의 그래프이다. 마찬가지로 가로축은  $\ln \tau_\rho$ , 세로축은  $\ln\{-\ln[f(\tau_\rho)]\}$ 을 나타낸다. Figure 4.1에서 나타나는 것처럼 확실한 형태는 아니지만  $\rho = 16\text{mm}$ 일 때까지는 선형성을 띠는 것을 알 수 있다.  $\rho$ 가 32mm를 넘어서면서부터는 선형성이 파괴되고 일일 강수량이  $\rho$ 를 넘어서는 시간  $\tau_\rho$ 의 분포는 가늠하기가 힘들어진다. 같은 이유로  $\rho$  값이 증가할수록 형태 모수(shape parameter)  $\beta$  값이 점점 작아지고 매우 두터운 꼬리를 가지게 된다. 즉,  $\tau_\rho$ 는 난수에 가까울 정도로 예측하기가 힘들어진다.

Table 4.1은 임계값  $\rho$ 에 따른  $\tau_\rho$ 의 확률밀도함수  $f(\tau_\rho)$ 를 고려했을 때, 형태 모수  $\beta$ 와  $\tau_\rho$ 의 평균값인  $\tau_{\text{mean}}(\rho)$  및 평균값의 표준오차  $\text{s.e.}\{\tau_{\text{mean}}(\rho)\}$  값을 나타낸 것이다. Figure 4.2에서 선형성이 강하게 나타났던  $\rho = 16\text{mm}$ 일 때까지는  $\tau_{\text{mean}}(\rho)$ 의 표준오차가 평균값에 비해 6% 이내로 안정되게 나타난다. 이후  $\rho = 64\text{mm}$ 일 때까지 10% 내외를 유지하다가  $\rho = 128\text{mm}$ 일 때는 40% 대까지 늘어나 그 값을 가늠할 수 없게 된다. 따라서 우리는  $\rho = 64\text{mm}$  정도를  $\tau_{\text{mean}}(\rho)$ 를 신뢰할 수 있는 한계로 여길 수 있다. 지난 42년간의 연간 강수량은 '산청, 부산, 제주도, 강릉' 순으로 많지만 일일 강수량이  $\rho = 64\text{mm}$ 를 넘어서는 기대 시간은 '부산, 산청, 강화도, 서울' 순으로 짧아 차이가 있는 것을 확인할 수 있다. 이는 Figure 4.1에서 형태 모수  $\beta$  값이 '강화도, 산청, 부산, 서울' 순으로 작게 나타나는 것으로부터도 어느 정도 추정할 수 있다. 한편, 연간 강수량은 '의성, 보령, 춘천, 강화도' 순으로 적지만 일일 강수량이  $\rho$ 를 넘어서는 기대 시간은 '의성, 울릉도, 보령, 제주도' 순으로 길어 차이가 있는 것을 알 수 있다. 마찬가지로 Figure 4.1에서 형태 모수  $\beta$  값이 '울릉도, 의성, 광주, 제주도' 순으로 커서 어느 정도 추정할 수 있다.

최종적으로 Figure 4.3은 강수량  $\rho$ 에 따른 첫 통과시간(first passage time)의 기대값  $\tau_{\text{mean}}(\rho)$ 의 그래프를 나타낸 것이다.  $\rho$ 의 확률밀도함수 모형 및  $\tau_\rho$ 의 확률밀도함수 모형과 비슷하게  $\tau_{\text{mean}}(\rho) \sim e^{a\rho^\beta}$ 의 확장된 지수함수를 가정하여 적합선(fitted line)을 찾으면 Figure 4.3과 같다. 그림에서 확인하듯이 모든 지역에서 선형성을 띠는 것을 알 수 있다. 흥미롭게도 제주도를 제외하면 거의 완벽한 선형성을 띠고 있어 확장된 지수함수를 가지는 것을 확신할 수 있다.  $\rho$ 와  $\tau_\rho$ 의 확률밀도함수와 다르게  $\tau_{\text{mean}}(\rho)$ 은 지수항이  $a\rho^\beta$ 로 양수이므로 증가함수이다. Table 4.1에서  $\rho = 64\text{mm}$ 일 때를 보면  $\tau_{\text{mean}}(\rho)$  값은 '의성, 울릉도, 보령, 제주도' 순으로 크지만 형태 모수  $\beta$  값은 '의성, 부산, 강릉, 울릉도' 순으로 나타난다. 이로 보아 '부산'과 '강릉'은  $\rho$  값이 64mm를 크게 넘어섰을 때  $\tau_{\text{mean}}(\rho)$  값이 급증할 것으로 예상된다. 본

**Table 4.1.** Exponent  $\beta$  of the stretched exponential distribution from  $\rho = 1, \dots, 128$ 

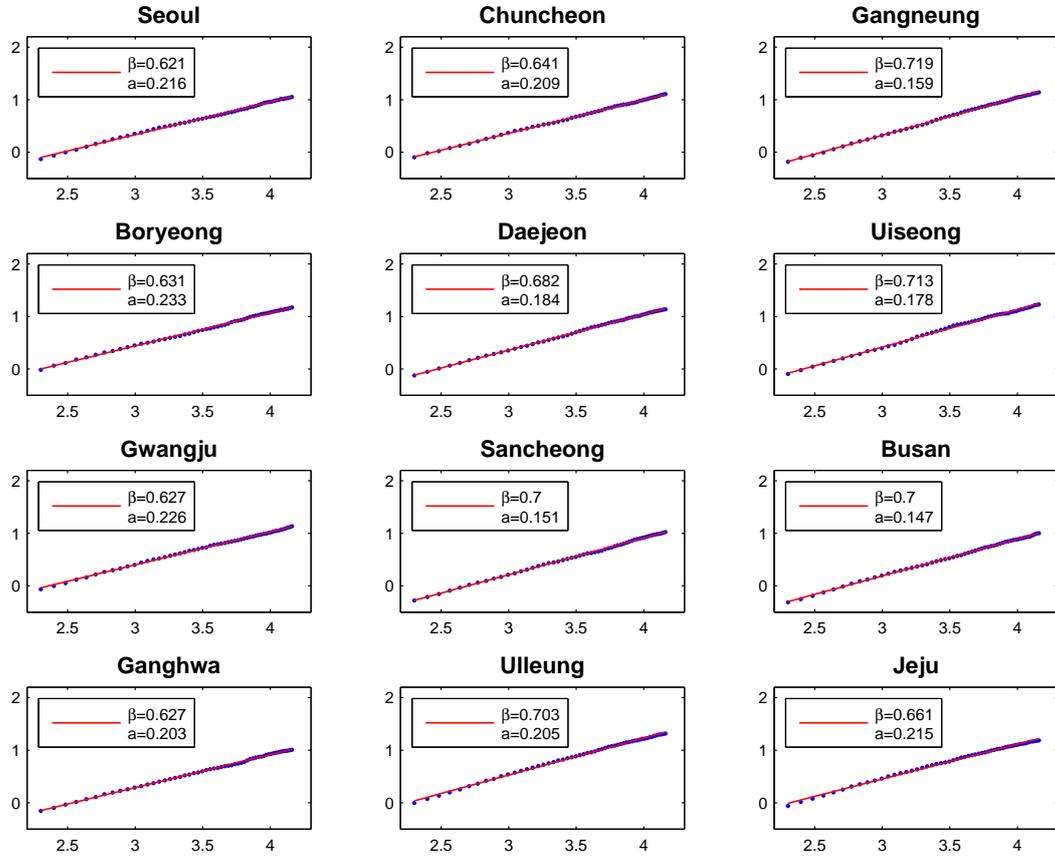
Threshold $\rho$		서울	춘천	강릉	보령	대전	의성	광주	산청	부산	강화도	울릉도	제주도
	$\beta$	<b>0.600</b>	<b>0.773</b>	<b>0.673</b>	<b>0.718</b>	<b>0.721</b>	<b>0.504</b>	<b>0.811</b>	<b>0.521</b>	<b>0.515</b>	<b>0.547</b>	<b>0.821</b>	<b>0.910</b>
1mm	$\tau_{\text{mean}}(\rho)$	4.57	4.60	4.32	4.39	4.29	5.09	4.05	4.68	4.78	5.09	3.35	3.73
	s.e. $\{\tau_{\text{mean}}(\rho)\}$	0.09	0.09	0.09	0.09	0.08	0.12	0.07	0.11	0.11	0.12	0.05	0.07
	$\beta$	<b>0.445</b>	<b>0.501</b>	<b>0.549</b>	<b>0.548</b>	<b>0.515</b>	<b>0.431</b>	<b>0.571</b>	<b>0.424</b>	<b>0.488</b>	<b>0.433</b>	<b>0.890</b>	<b>0.712</b>
2mm	$\tau_{\text{mean}}(\rho)$	5.45	5.43	5.06	5.38	5.17	6.06	4.92	5.50	5.54	6.03	4.00	4.44
	s.e. $\{\tau_{\text{mean}}(\rho)\}$	0.13	0.13	0.12	0.12	0.12	0.17	0.11	0.15	0.14	0.16	0.07	0.09
	$\beta$	<b>0.390</b>	<b>0.380</b>	<b>0.435</b>	<b>0.400</b>	<b>0.425</b>	<b>0.355</b>	<b>0.458</b>	<b>0.373</b>	<b>0.424</b>	<b>0.359</b>	<b>0.704</b>	<b>0.555</b>
4mm	$\tau_{\text{mean}}(\rho)$	6.91	6.94	6.29	7.17	6.45	7.78	6.33	6.84	6.75	7.67	5.15	5.71
	s.e. $\{\tau_{\text{mean}}(\rho)\}$	0.20	0.21	0.17	0.20	0.18	0.26	0.16	0.22	0.19	0.25	0.10	0.13
	$\beta$	<b>0.316</b>	<b>0.319</b>	<b>0.354</b>	<b>0.329</b>	<b>0.343</b>	<b>0.307</b>	<b>0.342</b>	<b>0.334</b>	<b>0.348</b>	<b>0.307</b>	<b>0.454</b>	<b>0.369</b>
8mm	$\tau_{\text{mean}}(\rho)$	9.60	9.84	8.76	10.20	9.05	10.91	8.98	8.87	8.86	10.55	7.64	8.21
	s.e. $\{\tau_{\text{mean}}(\rho)\}$	0.39	0.40	0.28	0.37	0.32	0.46	0.29	0.33	0.31	0.44	0.18	0.25
	$\beta$	<b>0.260</b>	<b>0.251</b>	<b>0.287</b>	<b>0.259</b>	<b>0.271</b>	<b>0.243</b>	<b>0.270</b>	<b>0.285</b>	<b>0.288</b>	<b>0.251</b>	<b>0.294</b>	<b>0.289</b>
16mm	$\tau_{\text{mean}}(\rho)$	15.56	15.74	14.09	17.26	14.83	18.33	14.90	13.55	13.57	16.55	14.17	14.59
	s.e. $\{\tau_{\text{mean}}(\rho)\}$	0.89	0.94	0.59	0.92	0.77	1.05	0.76	0.65	0.59	0.99	0.47	0.62
	$\beta$	<b>0.189</b>	<b>0.184</b>	<b>0.201</b>	<b>0.185</b>	<b>0.196</b>	<b>0.162</b>	<b>0.191</b>	<b>0.209</b>	<b>0.214</b>	<b>0.188</b>	<b>0.187</b>	<b>0.194</b>
32mm	$\tau_{\text{mean}}(\rho)$	29.61	31.38	30.39	34.70	30.38	43.97	30.36	25.85	25.24	30.20	36.20	32.20
	s.e. $\{\tau_{\text{mean}}(\rho)\}$	2.61	2.85	1.99	2.87	2.34	4.16	2.60	1.91	1.67	2.59	2.40	2.28
	$\beta$	<b>0.118</b>	<b>0.102</b>	<b>0.100</b>	<b>0.095</b>	<b>0.091</b>	<b>0.079</b>	<b>0.113</b>	<b>0.123</b>	<b>0.124</b>	<b>0.116</b>	<b>0.078</b>	<b>0.114</b>
64mm	$\tau_{\text{mean}}(\rho)$	80.07	93.96	98.20	110.45	96.64	155.53	90.55	75.83	72.78	78.87	139.83	98.77
	s.e. $\{\tau_{\text{mean}}(\rho)\}$	9.11	10.73	8.98	13.08	9.93	19.65	10.51	7.53	6.96	8.67	17.42	11.14
	$\beta$	<b>0.051</b>	<b>0.046</b>	<b>0.025</b>	<b>0.046</b>	<b>0.026</b>	<b>0.000</b>	<b>0.002</b>	<b>0.057</b>	<b>0.023</b>	<b>0.048</b>	<b>0.011</b>	<b>0.043</b>
128mm	$\tau_{\text{mean}}(\rho)$	389.58	389.63	467.90	450.52	485.17	2452.00	739.80	286.60	400.13	307.91	1757.00	329.61
	s.e. $\{\tau_{\text{mean}}(\rho)\}$	59.13	70.15	87.10	113.56	116.69	895.27	125.65	44.81	63.32	50.68	729.37	57.90

연구에서는  $\rho$ 값이 128mm를 넘어섰을 때 표준오차가 커져 값을 크게 신뢰할 수 없으나, 부산과 강릉지방의  $\tau_{\text{mean}}(\rho)$ 가 급증한 것을 Table 4.1에서 확인할 수 있다.

## 5. 결론 및 논의

우리는 1973년 1월 1일부터 2014년 12월 31일까지 42년(15,340일) 간, ‘서울, 춘천, 강릉, 보령, 대전, 의성, 광주, 산청, 부산, 강화도, 울릉도, 제주도’ 등 대한민국 12개 지역의 일일 강수량 데이터를 분석하여 역통계량을 산출하였다. 분석은 다음과 같이 크게 3가지로 분류하여 이루어졌다. 첫째, 일일 강수량  $x$ 의 확률밀도함수  $f(x)$ 를 도출하였다. 둘째, 특정 강수량  $\rho$ 를 처음으로 넘어서는 기간  $\tau_\rho$ 의 확률밀도함수  $f(\tau_\rho)$ 를 도출하였다. 셋째, 특정 강수량  $\rho$ 에 대해  $\tau_\rho$ 의 대푯값을 그들 분포의 평균으로 두고  $\tau_{\text{mean}}(\rho)$ 를 계산하여  $\rho$ 에 대한  $\tau_{\text{mean}}(\rho)$ 의 관계식을 도출하였다.  $f(x)$ 와  $f(\tau_\rho)$  모두에서 형태 모수(shape parameter)  $\beta$ 의 값이 미만인 확장된 지수분포(stretched exponential distribution)를 이루는 것을 확인하였다. 또한  $\rho$ 와  $\tau_{\text{mean}}(\rho)$ 의 관계 역시 확장된 지수 함수를 이루는 것을 알 수 있었다. 이는 브라운 운동에서  $f(x)$ 와  $f(\tau_\rho)$ 가 멱함수 분포(power-law distribution)를 이루고,  $\rho$ 와  $\tau_{\text{mean}}(\rho)$ 의 관계 역시 멱함수 관계를 가지는 것과 뚜렷이 대조된다.

여러 가지 분석 결과 다우 지역과 소우 지역의 특성 및 강수 집중도 등이 분포 특성으로부터 유도되었다. 강수량이 가장 많은 ‘산청’과 ‘부산’은 물론 ‘강화도’와 ‘서울’의 일일 강수량 분포  $f(x)$ 의 꼬리가 두터운(heavy tail) 것을 확인하였다. 즉, 이들은 폭우의 가능성이 높은 지역으로 분류될 수 있다. 반대로 울릉도의 일일 강수량 분포가 두터운 꼬리를 가지는데 강수량의 통계 집중률이 상대적으로 높은 지역의 특성을 나타내는 것으로 풀이된다. 비슷하게  $\rho = 64\text{mm}$ 일 때  $f(\tau_\rho)$ 의 분포에서 형태 모수  $\beta$  값이 ‘강화도, 산청, 부산, 서울’ 순으로 작게 나타나는 것을 알 수 있었다. 다시 말해서 강수량이 12개 지



**Figure 4.3.** Plot of the inverse statistics  $\tau_{\text{mean}}(\rho)$  versus  $\rho$  on  $\rho = 10\text{mm}$  to  $100\text{mm}$  for 12 regions. The x-axis means  $\ln \rho$  and y-axis means  $\ln\{\ln[\tau_{\text{mean}}(\rho)]\}$ . They show the features of strong linearity, i.e., their distributions are stretched exponential on the range. The coefficients of determination for linear regressions are between 0.9972 and 0.9996.

역 중 9번째인 ‘강화도’나 5번째인 ‘서울’이 폭우가 내릴 기대 시간은 짧다는 것을 알 수 있다. 한편,  $\rho$ 와  $\tau_{\text{mean}}(\rho)$ 의 관계에서 형태 모수  $\beta$ 의 값은 ‘의성, 부산, 강릉, 울릉도’ 순으로 나타난다. 이번 연구에서 기준으로 삼은  $\rho = 64\text{mm}$ 에서의 기대 시간은 부산이 가장 짧게 나타났으나 이를 크게 넘어서는(예를 들어 일일 강수량이  $\rho = 200\text{mm}$ 을 넘어서는) 기대 시간은 급증할 것이라는 것을 추정할 수 있다.

한편, 첫 통과시간(first passage time) 추정 시 계절적 요인을 제거하고 계산을 하면 보다 더 정확한 결과를 얻을 수 있을 것이다. 하지만 우리의 주된 관심사는 어떤 한계값을 초과하여 폭우가 내리기까지의 기대 시간을 구하는 것이다. 따라서 추세나 노이즈를 제거하지 않은 실제적인 데이터를 이용하여 분석하는 데에 치중하였다. 그러나 봄철에 폭우나 폭설이 내릴 것으로 기대하기 힘들 듯이 계절별로 강수량의 특성들을 구분하여 고려한다면 보다 자세하게 첫 통과시간 값을 얻을 수 있을 것이다. 가령, 월별로 강수 측정치를 구분한다면 각 지역마다 1월부터 12월까지 12개의 강수량 확률밀도함수가 구해진다 (Figure 4.1). 또한 첫 통과시간의 확률밀도함수도 월별로 구할 수 있고 (Figure 4.2), 이로부터 강수량에 따른 첫 통과시간의 기댓값 함수도 월별로 표현될 것이다 (Figure 4.3). 하지만 유의할 점은 특정 기

간 별로 구분하여 강수량 특성을 분석한다면 전년도 마지막 날과 차년도 첫 날이 불연속적 특성을 발생 시킨다는 점이 있다. 극단적으로 3월 1일부터 6월 30일까지 4개월 단위로 구분하여 강수량 특성을 분석한다면 전년도 6월 30일은 강수량이 가장 많은 계절인 반면 3월 1일은 대표적인 건기로 전년도 6월 30일과 차년도 3월 1일 데이터를 연속된 시계열로 해석하기 힘들다. 이러한 문제점을 해결하기 위해 시간을 세분화할수록 개별 시계열의 데이터 수는 줄어들어 통계학적 신뢰성을 상실하는 문제점이 발생한다. 따라서 타당성과 신뢰성이 동시에 만족되도록 시간을 구분하여 강수량 특성을 분석한다면 본 연구에서 수행한 방법보다 발전된 결과를 얻을 수 있을 것으로 기대된다.

## References

- Ding, M. and Rangarajan, G. (2004). *New Directions in Statistical Physics*, Springer.
- Ding, M. and Yang, W. (1995). Distribution of first return time in fractional Brownian motion and its application to the study of on-off intermittency, *Physical Review E*, **52**, 207.
- Honerkamp, J. and Baumgärtner, A. (1986). Anomalous diffusion on a fractals: a first passage time problem, *Zeitschrift für Physik B: Condensed Matter*, **64**, 253–255.
- Jensen, M. H. (1999). Multiscaling and structure functions in turbulence: an alternative approach, *Physical Review Letter*, **83**, 76–79.
- Jensen, M. H., Johansen, A., Petroni, F., and Simonsen, I. (2004). Inverse statistics in the foreign exchange market, *Physica A*, **340**, 678.
- Jensen, M. H., Johansen, A., and Simonsen, I. (2003). Inverse statistics in economics: the gain-loss asymmetry, *Physica A*, **324**, 338.
- Lana, X., Burgueno, A., Martinez, M. D., and Serra, C. (2009). A review of statistical analyses on monthly and daily rainfall in Catalonia, *Journal of Weather & Climate of the Western Mediterranean*, **6**, 15–29.
- Lee, C.-Y., Kim, J., and Hwang, I. (2008). Inverse statistics of the Korea composite stock price index, *Journal of the Korean Physical Society*, **52**, 517–523.
- McCauley, J. L. (2004). *Dynamics of Markets: Econophysics and Finance*, Cambridge University Press.
- Rangarajan, G. and Ding, M. (2000a). First passage time distribution for anomalous diffusion, *Physics Letters A*, **273**, 322–330.
- Rangarajan, G. and Ding, M. (2000b). Anomalous diffusion and the first passage time problem, *Physical Review E*, **62**, 120–133.
- Rangarajan, G. and Ding, M. (2003). First passage distributions for long memory processes, *LNP*, **621**, 167–180.
- Simonsen, I., Jensen, M. H., and Johansen, A. (2002). Optimal investment horizon, *The European Physical Journal B*, **27**, 583–586.
- Wilson, P. S. and Toumi, R. (2005). A fundamental probability distribution for heavy rainfall, *Geophysical Research Letters*, **32**, 812.

# 역통계 방법론을 이용한 한반도의 강수 특성 분석

민승식<sup>a,1</sup>

<sup>a</sup>해군사관학교 이학과

(2015년 12월 10일 접수, 2016년 1월 25일 수정, 2016년 2월 27일 채택)

---

## 요약

본 논문에서는 1973년부터 2014년까지 42년 간, 12개 지역의 일일 강수량 데이터의 역통계(inverse statistics) 분석을 실시하였다. 구체적으로 일일 강수량  $x$ 의 확률밀도함수  $f(x)$ 를 도출하였고, 특정 강수량  $\rho$ 를 처음으로 넘어서는 기간  $\tau_\rho$ 의 분포  $f(\tau_\rho)$ 를 도출하였다. 최종적으로  $\tau_\rho$ 의 대푯값인  $\tau_{\text{mean}}(\rho)$ 와  $\rho$ 의 관계를 도출하였다. 분석 결과  $x$ 와  $\tau_\rho$ 는 확장된 지수분포(stretched exponential distribution)를 이루는 것을 확인하였다. 더불어  $\rho$ 와  $\tau_{\text{mean}}(\rho)$ 도 확장된 지수함수의 관계를 이루는 것을 알 수 있었다. 이들 분포를 바탕으로 형태 모수(shape parameter)  $\beta$  값을 도출하고, 12개 지역의 강수 특성을 분석하였다.

주요용어: 강수량, 역통계량, 확률밀도함수, 확장된 지수분포

---

<sup>1</sup>교신저자: (51704) 경상남도 창원시 진해구 중원로 1, 해군사관학교 이학과. E-mail: fieldsmi@gmail.com