

**İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ**

**UÇ DEĞERLER YÖNTEMİ İLE ANİ SU BASKINLARININ İSTATİKSEL  
MODELLENMESİ**

**YÜKSEK LİSANS TEZİ**

**Halil İbrahim ALTUN**

**Matematik Mühendisliği Anabilim Dalı**

**Matematik Mühendisliği Programı**

**MAYIS 2018**



**İSTANBUL TEKNİK ÜNİVERSİTESİ ★ FEN BİLİMLERİ ENSTİTÜSÜ**

**UÇ DEĞERLER YÖNTEMİ İLE ANİ SU BASKINLARININ İSTATİKSEL  
MODELLENMESİ**

**YÜKSEK LİSANS TEZİ**

**Halil İbrahim ALTUN  
509151207**

**Matematik Mühendisliği Anabilim Dalı**

**Matematik Mühendisliği Programı**

**Tez Danışmanı: Prof. Dr. Mustafa NADAR**

**MAYIS 2018**



İTÜ, Fen Bilimleri Enstitüsü'nün 509151207 numaralı Yüksek Lisans Öğrencisi Halil İbrahim ALTUN, ilgili yönetmeliklerin belirlediği gerekli tüm şartları yerine getirdikten sonra hazırladığı “UÇ DEĞERLER YÖNTEMİ İLE ANİ SU BASKINLARININ İSTATİKSEL MODELLENMESİ” başlıklı tezini aşağıda imzaları olan jüri önünde başarı ile sunmuştur.

**Tez Danışmanı :**      **Prof. Dr. Mustafa NADAR** .....  
İstanbul Teknik Üniversitesi

**Jüri Üyeleri :**      **Doç. Dr. Ahmet KIRIŞ** .....  
İstanbul Teknik Üniversitesi

**Dr. Öğr. Üyesi Fatih KIZILASLAN** .....  
Marmara Üniversitesi

**Teslim Tarihi**      :    **04 Mayıs 2018**  
**Savunma Tarihi**    :    **31 Mayıs 2018**





*Aileme,*





## ÖNSÖZ

Yüksek lisansa başladığım ilk günden itibaren, tez konumun belirlenmesinden bitimine kadar benden yardımını, bilgisini ve tecrübesini esirgemeyen Sayın Prof. Dr. Mustafa NADAR'a, bu süreç boyunca bana güvendikleri ve zor zamanlarımda bana destek oldukları için ailem ve arkadaşlarıma teşekkür ediyorum.

Ayrıca yüksek lisans sürecimi 3001 - Başlangıç Ar-Ge Projeleri Destekleme Programı 115Y416 numaralı proje kapsamında destekleyen TÜBİTAK'a teşekkürlerimi sunarım.

Haziran, 2018

Halil İbrahim ALTUN



## İÇİNDEKİLER

### Sayfa

ÖNSÖZ .....	vii
İÇİNDEKİLER .....	ix
KISALTMALAR .....	xi
SEMBOLLER .....	xiii
ÇİZELGE LİSTESİ.....	xv
ŞEKİL LİSTESİ.....	xvii
ÖZET .....	xix
SUMMARY .....	xxi
<b>1. GİRİŞ</b> .....	<b>1</b>
<b>2. TEK DEĞİŞKENLİ UÇ DEĞER YÖNTEMLERİ</b> .....	<b>7</b>
2.1 Tek Değişkenli Blok Maksimum Yöntemi .....	7
2.2 Tek Değişkenli POT (Eşik Seviyesi Aşım) Yöntemi .....	8
2.3 Tek Değişkenli ACER (Ortalama Koşullu Aşım Oranı) Yöntemi.....	10
<b>3. İKİ DEĞİŞKENLİ UÇ DEĞER YÖNTEMLERİ</b> .....	<b>15</b>
3.1 İki Değişkenli Blok Maksimum Yöntemi .....	15
3.2 İki Değişkenli POT Yöntemi.....	21
3.2.1 İki değişkenli uç değerler dağılım fonksiyonu, pickand bağımlılık fonksiyonu ve kopula ilişkisi .....	22
3.2.2 Kopula yaklaşımı .....	23
<b>4. UYGULAMALAR</b> .....	<b>27</b>
4.1 Gereçler .....	27
4.2 Tek Değişkenli Blok Maksimum Yöntemine Göre Uç Değer Yaklaşımları Üzerine Sonuçlar .....	31
4.2.1 Deniz seviyesi değişkeninin blok maksimum yöntemiyle modellemesi ..	31
4.2.2 Basınç şiddeti değişkeninin blok maksimum yöntemiyle modellemesi ..	38
4.3 Tek Değişkenli Eşik Seviyesini Aşan Değerler Yöntemine Göre Uç Değer Yaklaşımları Üzerine Sonuçlar .....	45
4.3.1 Deniz seviyesi değişkeninin pot yöntemiyle modellemesi .....	45
4.3.2 Basınç şiddeti değişkeninin pot yöntemiyle modellemesi .....	48
4.4 Tek Değişkenli ACER Yöntemine Göre Uç Değer Yaklaşımları Üzerine Sonuçlar.....	52
4.4.1 Deniz seviyesi değişkeninin acer yöntemiyle modellemesi.....	52
4.4.2 Basınç şiddeti değişkeninin acer yöntemiyle modellemesi.....	57
4.5 İki Değişkenli Uç Değer Yaklaşımları Üzerine Sonuçlar .....	61
4.5.1 Blok maksimum yöntemi üzerine sonuçlar.....	61
4.5.1.1 Kopula yaklaşımına göre.....	62
4.5.1.2 İki değişkenli uç değer dağılımları kullanarak.....	65
4.5.2 Eşik seviyesini aşan değerler yöntemi üzerine sonuçlar .....	70
<b>5. SONUÇ VE ÖNERİLER.....</b>	<b>73</b>

<b>KAYNAKLAR</b> .....	<b>77</b>
<b>ÖZGEÇMİŞ</b> .....	<b>81</b>



## **KISALTMALAR**

<b>ACER</b>	: Ortalama Koşullu Aşım Oranı
<b>AIC</b>	: Akıke Bilgi Kriteri
<b>ALOG</b>	: Asimetrik Lojistik
<b>ANEGLOG</b>	: Asimetrik Negatif Lojistik
<b>ÇUD</b>	: Çok Değişkenli Uç Değerler
<b>GEV</b>	: Genelleştirilmiş Uç Değerler
<b>GPD</b>	: Genelleştirilmiş Pareto Dağılımı
<b>GUD</b>	: Genellenmiş Uç Değerler
<b>IFM</b>	: Marjinalere İlişkin Çıkarsama
<b>LOG</b>	: Lojistik
<b>MLE</b>	: Maksimum Olabilirlik Tahmini
<b>MSL</b>	: Ortalama Deniz Seviyesi
<b>NEGLOG</b>	: Negatif Lojistik
<b>POT</b>	: Eşik Seviyesi Aşım
<b>PWM</b>	: Olasılık Ağırlık Momenti
<b>TUD</b>	: Tek Değişkenli Uç Değerler



## SEMBOLLER

$u$	: Eşik seviyesi
$\xi$	: Şekil parametresi
$\sigma$	: Ölçek parametresi
$\mu$	: Konum parametresi







## ÇİZELGE LİSTESİ

### Sayfa

<b>Çizelge 3.1</b> : İki değişkenli uç değer dağılımları.....	18
<b>Çizelge 3.2</b> : Uç değer kopula fonksiyonları .....	24
<b>Çizelge 4.1</b> : Deniz seviyesi için oturtulan GEV parametrelerinin %95 lik asimptotik güven aralıkları.....	34
<b>Çizelge 4.2</b> : Deniz seviyesinin GEV dağılımına göre tekrar seviyeleri .....	34
<b>Çizelge 4.3</b> : Deniz seviyesi değişkeni için Gumbel dağılımına göre tekrar seviyeleri.....	37
<b>Çizelge 4.4</b> : Basınç şiddeti için seçilen GEV dağılım parametrelerinin %95 lik asimptotik güven aralıkları.....	41
<b>Çizelge 4.5</b> : Basınç şiddeti için GEV dağılımına göre tekrar seviyesi.....	41
<b>Çizelge 4.6</b> : Basınç şiddeti için Gumbel dağılımına göre tekrar seviyesi.....	44
<b>Çizelge 4.7</b> : Deniz seviyesi GP dağılımına göre tekrar seviyeleri .....	47
<b>Çizelge 4.8</b> : MSL baz alındığında deniz seviyesi GP dağılımına göre tekrar seviyeleri .....	48
<b>Çizelge 4.9</b> : Basınç şiddeti GP dağılımına göre tekrar seviyeleri .....	50
<b>Çizelge 4.10</b> : Deniz seviyesi için k=1 durumundaki ACER yöntemine göre tekrar seviyeleri.....	55
<b>Çizelge 4.11</b> : Deniz seviyesi için k=2 durumundaki ACER yöntemine göre tekrar seviyeleri.....	57
<b>Çizelge 4.12</b> : Basınç şiddeti için k=1 durumundaki ACER yöntemine göre tekrar seviyeleri.....	59
<b>Çizelge 4.13</b> : Basınç şiddeti için k=2 durumundaki ACER yöntemine göre tekrar seviyeleri.....	61
<b>Çizelge 4.14</b> : Blok büyüklüğü aylık aldığında parametrik ve parametrik olmayan korelasyon ilişkileri.....	62
<b>Çizelge 4.15</b> : Kopulalar için üst kuyruk bağımlılık parametresi ve üst sınır... ..	64
<b>Çizelge 4.16</b> : Kopulalar için parametre tahmini $\theta_n$ , uyum iyiliği test istatistiği $S_n$ ve AIC bilgi kriteri.....	64
<b>Çizelge 4.17</b> : İki değişkenli uç değer dağılımları ve AIC bilgi kriteri .....	66
<b>Çizelge 4.18</b> : Tek değişkenli blok maksimum yöntemiyle elde edilen 2, 20, 100 senelik tekrar seviyelerinin iki değişkenli blok maksimum modeline göre eriştiği tekrar seviyeleri.....	68

<b>Çizelge 4.19 :</b> İki deęişkenli blok maksimum için seçilen LOG modeline göre dağılım dilimi kullanarak elde edilen eşik seviyeleri için hesaplanan birleşik tekrar seviyeleri... ..	69
<b>Çizelge 4.20 :</b> Dağılım dilimleri ve buna karşılık gelen deniz seviyesi, basınç şiddeti değerleri ve bunlardan en az birinin tüm verideki aşım oranı.....	70
<b>Çizelge 5.1 :</b> Deniz ve Basınç deęişkenleri için tek deęişkenli ACER yöntemi tekrar seviyesi sonuçları.. ..	74



## ŞEKİL LİSTESİ

### Sayfa

Şekil 4.1 : Eksik verilerin çoklu ‘imputation’ yöntemiyle doldurulup ortalamalarının dağılımları.....	27
Şekil 4.2 : Deniz seviyesi değişkeninin yıllık periyodikliğinin diyagnostiği.....	33
Şekil 4.3 : Deniz seviyesi için seçilen GEV dağılımının model uyumluluk grafikleri .....	35
Şekil 4.4 : Deniz seviyesi için seçilen Gumbel dağılımının model uyumluluk grafikleri.....	36
Şekil 4.5 : Basınç değişkeni için yıllık periyodikliğin diyagnostiği.....	39
Şekil 4.6 : Basınç değişkeni mevsimsellik etkisi diyagnostiği.....	40
Şekil 4.7 : Basınç şiddeti değişkeni için seçilen GEV dağılımının model uyumluluk grafikleri.....	42
Şekil 4.8 : Basınç şiddeti için seçilen Gumbel dağılımının model uyumluluk grafikleri.....	43
Şekil 4.9 : Deniz seviyesi için ortalama aşım değer grafiği.....	46
Şekil 4.10 : Deniz seviyesi için seçilen GP dağılımının model uyumluluk grafikleri.....	47
Şekil 4.11 : Basınç şiddeti için ortalama aşım değer grafiği.. ..	50
Şekil 4.12 : Basınç şiddeti için seçilen GP dağılımının model uyumluluk grafikleri.....	51
Şekil 4.13 : Deniz seviyesi için farklı k değerlerine karşılık ACER fonksiyonu Tahminleri.....	52
Şekil 4.14 : Deniz seviyesi için k=1 ve 2 yıllık periyotta oturtulan ACER fonksiyonu.....	53
Şekil 4.15 : Deniz seviyesi için k=1 ve 20 yıllık periyotta oturtulan ACER fonksiyonu.....	54
Şekil 4.16 : Deniz seviyesi için k=1 ve 100 yıllık periyotta oturtulan ACER fonksiyonu.....	54
Şekil 4.17 : Deniz seviyesi için k=2 ve 2 yıllık periyotta oturtulan ACER fonksiyonu.....	55
Şekil 4.18 : Deniz seviyesi için k=2 ve 20 yıllık periyotta oturtulan ACER fonksiyonu.....	56

<b>Şekil 4.19</b> : Deniz seviyesi için $k=2$ ve 100 yıllık periyotta oturtulan ACER fonksiyonu.....	56
<b>Şekil 4.20</b> : Basınç şiddeti için farklı $k$ değerlerine karşılık ACER fonksiyonu Tahminleri.....	57
<b>Şekil 4.21</b> : Basınç şiddeti için $k=1$ ve 2 yıllık periyotta oturtulan ACER fonksiyonu.....	58
<b>Şekil 4.22</b> : Basınç şiddeti için $k=1$ ve 20 yıllık periyotta oturtulan ACER fonksiyonu.....	58
<b>Şekil 4.23</b> : Basınç şiddeti için $k=1$ ve 100 yıllık periyotta oturtulan ACER fonksiyonu.....	59
<b>Şekil 4.24</b> : Basınç şiddeti için $k=2$ ve 2 yıllık periyotta oturtulan ACER fonksiyonu.....	60
<b>Şekil 4.25</b> : Basınç şiddeti için $k=2$ ve 20 yıllık periyotta oturtulan ACER fonksiyonu.....	60
<b>Şekil 4.26</b> : Basınç şiddeti için $k=2$ ve 100 yıllık periyotta oturtulan ACER fonksiyonu.....	61
<b>Şekil 4.27</b> : Gumbel ve Frank kopulalar için Chi-Plot ve üst kuyruk bağımlılığı.....	63
<b>Şekil 4.28</b> : İki değişkenli dağılımlar için Pickand bağımlılık fonksiyonları.....	67
<b>Şekil 4.29</b> : İki Değişkenli POT yöntemine göre 2, 20 ve 100 tekrar seviyeleri kontor eğrileri.....	71
<b>Şekil 4.30</b> : MSL baz alındığında iki değişkenli POT yöntemine göre tekrar seviyeleri kontor eğrileri.....	72

## UÇ DEĞERLER YÖNTEMİ İLE ANI SU BASKINLARININ İSTATİKSEL MODELLENMESİ

### ÖZET

Ani su baskınları kıyı şehirlerimizin maruz kaldığı en önemli sorunlardan biridir ve baskınlarının yaşandığı şehirlerden birisi de İzmir'dir. İzmir Körfezi gibi kıyı ve haliç bölgelerindeki taşkınlar, uygun yönden esen kuvvetli rüzgar ve hava basıncında belirgin düşüş ve genelde bununla ilişkili yağış nedeni ile deniz seviyesinin yükselmesinden kaynaklanır. Bu tez çalışmasında amaç kıyı yönetim planlarının geliştirilmesinde katkıda bulunacak ani su baskını riskini öngörebilme, belirlenmiş bir deniz seviyesinin tekrarlama periyodu bilgilerini elde edebileceğimiz modeller geliştirmektir. Menteş istasyonundan alınan deniz seviyesi ve atmosfer basıncı ölçümleri verilerini kullanarak elde edilen modeller:

a) Deniz seviyesi ve atmosfer basıncı değerleri için marjinal kuyruk dağılım yapıları her bir değişken için yıllık blok maksimumları alınarak tek değişkenli GEV ve Gumbel dağılımlarıyla oturtulmuştur.

b) Deniz seviyesi ve atmosfer basıncı değişkenleri için uç eşik değerleri saptanarak POT yöntemi ile Genelleştirilmiş Pareto dağılımları oturtulmuştur.

c) Deniz seviyesi ve atmosfer basıncı değişkenleri için belli bir eşik seviyesini aşan değerlerin koşullu saptanmış önceki (k-1) değerinin o seviyeyi aşmaması şartıyla ACER yöntemiyle oturtulmuştur.

d) Deniz seviyesi ve atmosfer basıncı ölçümlerini birlikte değerlendirerek aralarındaki bağımlılık ölçüsünü de dikkate alarak iki değişkenli blok maksimum yöntemine göre blok büyüğü aylık seçilerek hem Gumbel kopula hem de iki değişkenli LOG uç değer dağılımı ile modellenmiştir. Bağımlılık ilişkisi parametrik olmayan yöntemlerden Kendall tau yaklaşımıyla kopula ve Pickand bağımlılık fonksiyonu aracılığı ile tespit edilmiştir.

e) Son olarak LOG modeli baz alınarak yüzdeler dilim yardımıyla saptanan eşik değer için iki değişkenli POT modeli oturtulmuş ve 2, 20 ve 100 yıllık tekrar seviyeleri öngörülerini kontrol eğrileriyle tespit edilmiştir.

Tek ve iki değişkenli blok maksimum ve POT yöntemleri ile tek değişkenli ACER yöntemini incelediğimizde en yüksek risk seviyesi öngörüsü tek değişkenli blok maksimum metoduyla diğer yandan en düşük risk seviyesi öngörüsü ise iki değişkenli blok maksimum metoduyla gözlemlenmiştir.

Farklı yöntemler farklı risk seviyeleri öngörmektedir. Ancak iki değişkenli POT yöntemi hem deniz ve basınç değişkenleri arasındaki bağımlılığı dikkate aldığından hem de seçilen bir eşik seviyesine bağlı olarak blok maksimum yöntemleriyle kıyaslandığında daha çok veriden yararlanabildiğinden ve ayrıca mevsimsellik etkisinden arındırıldığından daha ön plana çıkmaktadır.

**Anahtar Kelimeler:** Ani su baskını riski, ACER metodu, tek ve çok deęişkenli genelleştirilmiş uç deęerler yöntemi, kopula, Pickands bağımlılık fonksiyonu, Kendall tau ve Spearman rho istatistikleri.



# STATISTICAL MODELLING OF SUDDEN FLOODS BY EXTREME VALUES METHOD

## SUMMARY

One of the most important problem of coastal cities is the sudden flooding and İzmir is one of the cities that exposure to the sudden flooding. As İzmir bay, flooding in the region of coastal and estuarine is usually caused by strong winds blowing from the appropriate direction, significant decrease in air pressure and rising sea levels due to rainfall associated with it.

The aim of this thesis is to develop a statistical model which may estimate the risk of sudden flooding, and the return levels. This model can contribute to improvement of the coastal management plans. Using the sea level and atmospheric pressure measurements obtained from Mentesh station we obtained the following models:

- a) For both the sea level and atmospheric pressure, we fitted univariate tail distribution, based on annual block maxima, with GEV and Gumbel distributions.
- b) For both the sea level and atmospheric pressure, after obtaining a threshold value Generalized Pareto distribution was fitted using POT method.
- c) The ACER method has been fitted, provided that the conditionally determined previous  $(k-1)$  values of values exceeding a certain threshold level for sea level and atmospheric pressure variables do not exceed that level.
- d) The asymptotic joint distribution of the sea level and atmospheric pressure were fitted by using Gumbel Copula and LOG distribution based on monthly block size based on two parameter block maximum approach. The dependency was studied by copulas, which was expressed in terms of Kendall tau, and Pickand dependency functions.
- e) Finally, using the LOG distribution the joint exceedance probability of the sea level and atmospheric pressure was studied after determining the threshold value (POT) based on quantiles of the data. The return level of 2, 20 and 100 years expressed as contour plots.

When we consider all five methods univariate block max predicts the highest level of risks, on the other hand bivariate block maximum method predicts the least level of risks.

Different methods leads to different level of risks. However, bivariate methods takes dependency between variables which may lead to a better representation of the true problem. Moreover, based on a good choice of a threshold, POT approach has advantage of using a lot more data than block maximum and diminish the seasonality effects.

The thesis consists of five chapters, in the first chapters the topic of the thesis and its purpose are mentioned.

In the second part, Block Maximum and POT methods are described as univariate extreme value methods and these methods are asymptotically representing GEV and GPD distributions. Furthermore, the ACER method is explained in detail in the context of the cascade of conditioning approximations, and this method is explained. Return level is defined to describe univariate states.

In the third chapter, it is mentioned how Block Maximum and POT model generated for univariate extreme value methods are derived for two variable states. Here we talk about the exponential measures, spectral metrics, the functions reflecting stable tail dependency, copula method and the nonparametric methods which are used to represent the dependency structure between two variant states. Models such as LOG, ALOG, NEGLOG, ANEGLOG and their extreme value distributions that are to be used for two variable states throughout the study have been defined. The two variable extreme value distribution functions, relationship between the Pickand dependency function and the copula has been mentioned and the copula approach are discussed in detail. Extreme value copula functions, parameter spaces for Gumbel, Galambos, Husler Reiss, TEV and Tawn, which are extreme value families, are defined and the Pickands dependency functions of these extreme values are given. The joint return levels of the models representing these two variant states are specified.

In the fourth chapter, we give the related definitions, theorems and properties for the univariate and bivariate extreme value methods that we will use as a tool throughout the application. Definition of stationarity and KPSS, Augmented-Dickey Fuller test and Box-Ljung test for stationarity are given. Q-Q drawings are used as diagnostic tools. Chi-Plot is defined for graphical analysis of tail dependency and tail dependency. The Sklar Theorem, the extreme value copula and the Pickand dependency function, which are frequently used in the application, are given short descriptions. Kendall tau and Spearman rho dependent criteria are also defined. The R program are used for statistical analysis throughout the application. Firstly, stationary conditions are provided when annual blocks for sea level and atmospheric pressure data are taken using the univariate Block Maximum method. For models fitted to Gumbel and GEV distributions, a risk prediction for 2, 20 and 100 year return levels are made. With the help of the likelihood ratio test statistic, these models are tested to have no superiority between them and the model can be reduced to the Gumbel distribution with two parameters for two variables. Secondly, the POT method is used for sea level and atmospheric pressure data, and after the stationarity condition was satisfied, the model is fitted to the GPD and the risk prediction for 2, 20 and 100 year return levels is made. Model compatibility graphs for both Block Maximum and POT method are depicted with the help of Q-Q drawings. Finally, the ACER method is applied to univariate extreme value methods. Since the method can also be applied to non-stationary data, no stationarity condition is recognized and with this method, risks for 2, 20, and 100 year return levels are predicted. In the application phase of bivariate extreme value methods, the correlations between the sea level and atmospheric pressure variables are given by Kendal tau, Spearman Rho and Pearson. Dependency relationship between sea level and atmospheric pressure is modeled by using the extreme value copula approach. Since it is natural to expect upper queue dependency between these two variables, this is looked at with Chi-Plot graphs. The goodness of fit tests between the extreme value copulas are calculated by



the Cramer-Von Mises test statistic. Bivariate extreme value distributions are derived with the appropriate selection of copulas and bivariate extreme value distributions are tried to be decided with the aid of Pickands dependency function and AIC information criterion. The LOG model is chosen for the bivariate block maximum method. Likewise, the LOG model is selected for the bivariate POT methods and the threshold values are determined. For both the Block Maximum and the POT methods, the joint return levels are calculated with contour curves in the bivariate extreme value method.

In the fifth section, the joint return levels of 2, 20, and 100 year return levels obtained using univariate extreme value methods are interpreted in bivariate cases. The reasons for preferring the univariate and bivariate extreme value methods have been discussed and the appropriate method has been decided in the modeling of sudden floods in İzmir region.

**Keywords:** Sudden flooding risk, ACER method, univariate and multivariate generalized extreme value method, copula, Pickands dependence function, Kendall tau statistics.



## 1. GİRİŞ

Günümüzde hidrolojiyle alakalı birçok ekstrem olay insanlar için ölümcül sonuçlara ve ciddi mal kayıplarına sebep olmaktadır. Sel baskınlarının analizi bu hidrolojik yapının tasarlanması ve yönetilmesinde özel bir önem teşkil etmektedir. Sel baskınlarının analizinin temel nesnesi ekstrem dilimindeki tahminlerin tespit edilmesi ve ekstrem olayların gerçekleşebilme olasılığıyla ilgili bilgi vermesidir [1]. Sıra dışı olayların ne zaman meydana gelebileceği ve etkilerinin ne olabileceği konularında belirsizlik söz konusu olsa da, bu belirsizliğin en azından belli bir düzeyde tahmin edilebilmesi bazı bilimsel yöntemlerin kullanılmasını gerektirmektedir. Söz konusu belirsizliklerin ortadan kaldırılabilmesi için ise sıra dışı olayların meydana gelme dinamiklerinin anlaşılması gerekmektedir. Bu dinamikler uç değer teoremi olarak bilinen Fisher-Tippet teoremi aracılığıyla asimptotik olarak kuyruk yapısının modellenmesi ile anlaşılabilir [2]. Uç değer teoreminin gelişim serüveni 1928 yılında Fisher ve Tippet ile başlamış [2], ardından 1943 yılında Gnedenko tarafından geliştirilmiş [3] ve son hali ise 1958 yılında Gumbel tarafından ortaya konulmuştur [4]. Daha yakın bir zamanda 1974 yılında Balkema ve De Haan ve 1975 yılında Pickands eşik seviyesine bağlı uç değer metodunun temellerini ortaya attılar [5, 6]. Kendisini ilk başta hidroloji ve klimatoloji alanında ispat etmiş bu teoreme günümüzde finans ve sigortacılık alanlarında da hızla ilgi duyulmaya başlanmıştır ve bununla ilgili uygulamalara Embrecht vd. [7] ve Reiss ve Thomas'ın [8] çalışmalarında rastlanabilir.

Bu çalışmada; mühendislik, tıp ve sigortacılık gibi alanlarda yıllardır başarıyla uygulanan teorik alt yapısı son derece gelişmiş olan bu uç değerler yönteminden yararlanılacaktır. Uç değerler yöntemi ile geçmiş yıllara ait deniz seviyesi ve basınç şiddeti değişkenleri kullanılarak geleceğe yönelik en büyük deniz seviyesi yükselmesi 2, 20, 100 yıllık periyotlar için belli modeller altında öngörülmüştür.

Bu çalışma kapsamında geliştirilen modellerde kullanılan veri, deniz seviyesi ve metalürjik ölçümleridir. Bu çalışmada ihtiyaç duyulan uzun süreli saatlik deniz seviyesi ölçümleri İzmir yakınlarında bulunan Menteş'ten ve saatlik atmosfer basıncı

değerleri de, bu bölgeye en yakın meteoroloji istasyonundan, Harita Genel Komutanlığından ve Meteoroloji Genel Müdürlüğünden temin edilmiştir.

İzmir Menteş istasyonundaki deniz seviyesi ölçümleri ve İzmir Meteoroloji istasyonunda toplanan saatlik basınç değerleri 1985-2015 yılları arasındaki ölçümlere dayandırılmıştır.

Bu çalışmada uç değerler teoremini uygularken tek değişkenli ve iki değişkenli durum için klasik yaklaşımlardan olan Blok Maksimum ve POT yöntemi kullanılmıştır. Ayrıca tek değişkenli durum için daha güncel bir metot olan ACER yöntemi çalışılmıştır.

Blok Maksimum yaklaşımı uç değerleri doğal bir yolla temsil etmesinden ötürü uç değer teoremi içerisinde kullanılan en yaygın yöntemlerden biridir. Bu yaklaşım gözlenen değerlerin oluşturduğu veriyi eşit uzunlukta bloklara böler ve bu blokların her birindeki maksimum veya minimum değeri alıp GEV dağılımlarına oturtarak (fit ederek) bir risk öngörüsünde bulunur [9, 10]. Blok Maksimum yöntemine göre daha modern bir yöntem olan POT yönteminin gelişimi Todorovic ve Zelenhasic'in 1970 yılında öne sürdükleri eşik seviyesini aşan değerler yaklaşımı ile başlamıştır [11]. Belli bir eşik seviyesini aşan değerlerin matematiksel bir metot haline dönüşümü 1975 yılında Pickands öncülüğünde başlamış [6] ve bu aşan değerlerin asimptotik olarak Genelleştirilmiş Pareto dağılımlarına oturtulabileceği 1984 yılında Smith ve Davison tarafından bulunmuştur [12-14].

Blok Maksimum ve POT metodunu kullanmanın pratik sebepleri vardır:

- Sadece mevcut bilgiler blok maksimum yöntemini kullanabilir. Örneğin uzun dönemli geçmiş kayıtlara göre yıllık maksimum verilerinden oluşmuş bir data blok maksimum ile ifade edilebilir [15].
- Blok maksimum yöntemi gözlemlerin tam anlamıyla bağımsız ve özdeş olmadığı zamanlarda tercih edilebilir. Bloklar kendi içerisinde bağımlılık ve/veya sezonsallık etkileri barındırabilir, ancak blokların arası bu etkilerden arındırılmış olmalıdır. Sezonsallık etkilerinden arındırılmış bir data için POT yöntemi kullanılması daha fazla veri kullanımına imkan sağlayacağından tercih edilir [16, 17].
- Blok maksimum metodu doğal olarak bazı durumlarda kendisini açıkça belli ettiğinden uygulanması daha kolaydır. Bu yüzden POT yöntemindeki kritik eşik seviyesini tespit etmek kadar zor olmayabilir. Öte yandan blok büyüklüklerini

değiřtirmek pratikte zor olacađından POT yöntemi de bazı durumlar için daha esnek olacaktır [18-20].

- POT yöntemi yıllık blok maksimum yöntemine nazaran belli bir eşik seviyesini aşan tüm verileri kullandığı için daha etkin analizler yapılmasına imkan tanır.

Blok Maksimum ve POT yöntemleri çalışılırken en çok kullanılan iki tahmin yöntemi MLE ve PWM olmuştur [21, 22]. Bu yöntemler kullanılarak oturtulan modellerin kalitesi uyum yüzdelik dilim grafiđi ve AIC yardımıyla değerlendirilebilir [23].

Çok deđişkenli analizlerin uygulaması tek deđişkenli göre daha karışık olmaktadır. Marjinal dađılımların ve bađımlılık yapısının bilgisini tutan birleşik olasılık fonksiyonlarının matematiksel karmaşıklığından ötürü bu problemle karşılaşılmıştır. İki deđişkenli uç deđer dađılımlarının birleşik dađılım yapısını tanımlamada en popüler yöntem kopula olmuştur. Kopula kavramı ilk kez bilimsel bir topluluk önünde, Sklar tarafından 1959 yılında tanımlanmıştır [24]. Nelsen 2006 yılında, kopulaların kullanımı üzerine teorik bir altyapı ve tanım katmıştır [25]. Kopulalar çok deđişkenli verileri analiz etme de, marjinal dađılımların bilindiđi durumda çok popüler olmuştur [26]. Genest'e göre uç deđer teorem ve kopula uygulamaları hidroloji alanında dünya çapında çok sayıda konferansa konu olmuştur [27]. Fisher'in 1997 yılındaki "Encyclopedia of Statistical Science" notlarında kopulaların iki sebeple istatistikçilerin ilgisini çektiđi belirtilmiştir. Bunlardan ilki bađımlılık ölçülerinin çalışılmasına imkan vermesidir, ikincisi ise iki deđişkenli dađılım ailelerinin üretilmesinde başlangıç noktası olmasıdır [28]. Kopulaların birçok çeşidi mevcuttur ve çevresel çalışmalarda kullanılmaktadır, fakat iki deđişkenli Arşimedyan kopulalar birleştirici, simetrik ve esnek olma özelliğinden dolayı hidroloji ile ilgili çalışmalarda en çok kullanılan kopula çeşidi olmuştur. Kopulaların ilgi çeken bir sınıfı uç deđer kopulalarıdır [29-31]. Uç deđer kopulaları ekstem olaylarla ilgili doğal bir tercih olmasının yanında pozitif bađımlılık sergileyen modellemelere de uygun bir seçim olmaktadır. Arşimedyan kopularına göre avantajı simetrik olmayışıdır. Başlıca uç deđer kopulaları Gumbel, Galambos, Husler Reiss ve TEV olmuştur [29, 32-34]. İki deđişkenli uç deđer dađılımları parametrik, yarı parametrik ve parametrik olmayan tekniklerle tahmin edilebilir. İki deđişkenli dađılım fonksiyonunda bileşenler arasındaki bađımlılık Kendall tau ve Spearman's rho ölçüleri ile dađılımdan bađımsız olarak saptanabilir [35]. İki (çok) deđişkenli uç

değerler (dağılım) metodunun başka bir karakterizasyonu da, Pickands tarafından tanımlanan bağımlılık fonksiyonudur [36]. İki değişkenli uç değerler dağılım fonksiyonunun bağımlılık yapısını ifade etmekte en yaygın olarak karşımıza çıkan Pickands bağımlılık fonksiyonunun asimptotik özellikleri Deheuvels tarafından incelenmiştir [37]. Birçok kaynakta Pickands bağımlılık fonksiyonu tahmin edicisi için marjinal dağılımların bilindiği varsayımı yapılmaktadır. Cap'era`a vd. bu kısıtı kaldırmıştır [38]. Ayrıca bu çalışmada kullanılacak olan bağımlılık modellerinden lojistik model Gumbel tarafından [39], asimetrik lojistik model Tawn tarafından [40], negatif lojistik model Galambos tarafından [41] ve asimetrik negatif lojistik model ise Joe tarafından geliştirilmiştir [42].

Klasik uç değer yöntemlerinden Blok Maksimum ve POT yönteminin aksine daha güncel bir metot olan ACER yöntemi Naess ve Gaidai tarafından 2009 yılında tanımlanmıştır [43]. Belli bir seviyeyi aşan değerlerin koşullu saptanmış  $(k-1)$  önceki değerlerinin o seviyeyi aşmaması şartıyla geliştirilmiş kuyruk dağılımını tahmin etmede kullanılan bir yöntemdir. Blok Maksimum ve POT yöntemlerinin aksine verilerin bağımsız ve özdeş olması şartı aranmaz, asimptotik argümanlara ihtiyaç duymaz ve durağan olmayan verilere de uygulanabilmektedir [43]. Bu yöntem başlangıçta sadece Gumbel durumu için önerilmiş olsa da sonradan daha genel durumlar için genişletilmiştir ve daha detaylı bilgi için Naess vd.'nin 2013 yılındaki çalışmaları incelenebilir [44]. Ayrıca ACER metodunun POT yöntemine göre daha kesin uç değer yüzdelik dilimleri tahmin ettiği gösterilmiştir [43].

Tez beş bölümden oluşmakta olup, birinci bölümde tezin konusu ve amacından söz edilmiştir.

İkinci bölümde tek değişkenli uç değer yöntemlerinden Blok Maksimum ve POT yöntemi tanımlanmış ve bu yöntemlerin asimptotik olarak GEV ve GPD dağılımlarını temsil ettiğinden söz edilmiştir. Ayrıca ACER yönteminin kademeli şartlı yakınsama mantığından ayrıntılı bir şekilde bahsedilip, bu yöntem açıklanmıştır. Tek değişkenli durumları tarif eden tekrar seviyesi tanımlanmıştır.

Üçüncü bölümde tek değişkenli uç değer yöntemleri için Blok Maksimum ve POT yöntemiyle oluşturulan modellerin iki değişkenli durumlar için nasıl türetildiğinden söz edilmiştir. Burada iki değişkenli durumlar arasındaki bağımlılık yapısını temsil etmek için kullanılan üstel ölçülerden, spektral ölçülerden, istikrarlı kuyruk

bağımlılığını yansıtan fonksiyonlardan, kopula metodundan ve parametrik olmayan yöntemlerden bahsedilmiştir. Çalışma boyunca iki değişkenli durumlar için kullanılacak olan LOG, ALOG, NEGLOG, ANEGLOG gibi modeller ve bunların uç değer dağılımları tanımlanmıştır. İki değişkenli uç değer dağılım fonksiyonu, Pickand bağımlılık fonksiyonu ve kopula arasındaki ilişkiden söz edilip, kopula yaklaşımından ayrıntılı bir şekilde bahsedilmiştir. Uç değer kopula ailelerinden olan Gumbel, Galambos, Husler Reiss, TEV ve Tawn için uç değer kopula fonksiyonları, parametre uzayları tanımlanmıştır ve bu uç değer kopyuların Pickands bağımlılık fonksiyonları verilmiştir. Bu iki değişkenli durumları temsil eden modellerin bileşik tekrar seviyeleri belirtilmiştir.

Dördüncü bölümde öncelikle uygulama boyunca gereç olarak kullanacağımız tek ve iki değişkenli uç değer yöntemleri için ilgili tanım, teorem ve özellikler verilmiştir. Durağanlık tanımı ve durağanlık için kullanacağımız KPSS, Augmented\_Dickey Fuller test ve Box-Ljung test verilmiştir. Diyagnostik araçlar olarak Q-Q çizimler kullanılmıştır. Kuyruk bağımlılığı ve kuyruk bağımlılığının grafiksel analizi için Chi-Plot tanımlanmıştır. Uygulama içerisinde sıkça kullanıldığından Sklar Teoremi, uç değer kopula ve Pickand bağımlılık fonksiyonu için kısa tanımlar verilmiştir. Ayrıca Kendal tau ve Spearman rho bağımlılık ölçütleri tanımlanmıştır. Uygulama boyunca istatistiksel analizler için R programı kullanılmıştır. Öncelikle tek değişkenli Blok Maksimum yöntemi kullanılarak deniz seviyesi ve basınç şiddeti verileri için yıllık bloklar alındığında durağanlık şartları sağlanmıştır. Gumbel ve GEV dağılımlarına oturtulan modeller için 2, 20 ve 100 yıllık tekrar seviyeleri için risk öngörülerinde bulunulmuştur. Olabilirlik oran test istatistiği yardımıyla bu modellerin aralarında bir üstünlük olmadığı test edilmiş ve model iki değişken için de iki parametrelili Gumbel dağılımına indirgenebilmiştir. Aynı şekilde deniz seviyesi ve basınç şiddeti verileri için POT yöntemi kullanılıp, durağanlık şartı sağlandıktan sonra model GPD'ye oturtulmuştur ve 2, 20 ve 100 yıllık tekrar seviyeleri için risk öngörülerinde bulunulmuştur. Hem Blok Maksimum hem de POT yöntemi için model uyumluluk grafikleri Q-Q çizimleri yardımıyla tasvir edilmiştir. Tek değişkenli uç değer yöntemlerinden son olarak ACER yöntemi uygulanıp, yöntemin durağan olmayan verilere de uygulanabilmesinden ötürü durağanlık şartı aranmamıştır ve 2, 20 ve 100 yıllık tekrar seviyeleri için risk öngörülerinde bulunulmuştur. İki değişkenli uç değer yöntemleriyle ilgili uygulama aşamasında öncelikle deniz seviyesi ve basınç şiddeti

değişkenleri arasındaki korelasyon ilişkileri Kendal tau, Spearman rho ve Pearson ile verilmiştir. Deniz seviyesi ve basınç şiddeti arasındaki bağımlılık ilişkisi uç değer kopula yaklaşımı hesaba katılarak modellenmiştir. Bu iki değişken arasında üst kuyruk bağımlılığı beklemek doğal olduğundan buna Chi-Plot grafikleriyle bakılmıştır. Uç değer kopulaları arasındaki uyum iyiliği testleri Cramer-Von Mises test istatistiğiyle hesaplanmıştır. Uygun kopula seçimleriyle iki değişkenli uç değer dağılımları türetilmiştir ve iki değişkenli uç değer dağılımlarına Pickands bağımlılık fonksiyonu ve AIC bilgi kriteri yardımıyla karar verilmeye çalışılmıştır. İki değişkenli Blok Maksimum yöntemi için LOG modeli seçilmiştir. Aynı şekilde iki değişkenli POT yöntemi için de LOG modeli seçilmiş ve eşik değerleri belirlenmiştir. İki değişkenli uç değer yönteminde hem Blok Maksimum hem de POT yöntemi için bileşik tekrar seviyeleri kontor eğrileri yardımıyla hesaplanmıştır.

Beşinci bölümde tek değişkenli uç değer yöntemleri kullanılarak elde edilen 2, 20 ve 100 yıllık tekrar seviyelerinin iki değişkenli durumda karşılık geldiği bileşik tekrar seviyeleri yorumlanmıştır. Tek değişkenli ve iki değişkenli uç değer yöntemlerinin tercih edilme sebepleri tartışılmıştır ve İzmir bölgesindeki ani su baskınlarının modellenmesinde uygun yöntem karar verilmiştir.



## 2. TEK DEĞİŞKENLİ UÇ DEĞER YÖNTEMLERİ

### 2.1. Tek Değişkenli Blok Maksimum Yöntemi

GEV teorisi dağılımların kuyruk davranışı ile ilgilenir ve oldukça güçlü matematiksel temellere dayanmaktadır. GEV bir sürecin uç olaylarına dair çıkarsama amaçlı kullanılır. Uç değerler yönteminden genellikle değişkenlerin gözlemlenen daha anormal düzeylerinin ekstrapolasyonunu elde etmek amaçlı yararlanır. Öncelikle tek değişkenli uç değerler teorisinden kısaca bahsedelim.

Varsayalım  $X_1, X_2, \dots, X_n$  bağımsız rasgele değişkenleri bilinmeyen bir  $F(x)$  dağılım fonksiyonundan gelsinler ve  $M_n = \max(X_1, X_2, \dots, X_n)$  olarak alınsın. Eğer aşağıdaki eşitliği sağlayacak öyle  $\{a_n\} > 0$  ve  $\{b_n\}$  dizileri için dejenere olmayan bir  $H$  dağılım fonksiyonu mevcut ise

$$\lim_{n \rightarrow \infty} P \left\{ \frac{M_n - b_n}{a_n} \leq x \right\} = \lim_{n \rightarrow \infty} F^n(a_n x + b_n) = H_{\xi, \mu, \sigma} \quad (2.1)$$

$F, H$ 'nin (maksimum) çekim-bölgesindedir denir ve denklem (2.1)'i sağlayan dağılım fonksiyonlar kümesi  $H_{\xi, \mu, \sigma}$  genelleştirilmiş uç değerler ailesi olarak ifade edilir [10, 45]. Dağılım fonksiyonu  $H_{\xi, \mu, \sigma}$  aşağıdaki biçimde ifade edilir.

$$H_{\xi, \mu, \sigma} = \begin{cases} \exp \left[ - \left( 1 + \frac{\xi(x-\mu)}{\sigma} \right)^{-\frac{1}{\xi}} \right], & \xi \neq 0 \\ \exp \left[ - \exp \left( - \frac{x-\mu}{\sigma} \right) \right], & \xi = 0 \end{cases} \quad (2.2)$$

$1 + \xi(x - \mu)/\sigma > 0$  ve  $\sigma > 0, \xi, \mu \in \mathbb{R}$  dir.  $\xi, \mu, ve \sigma$  sırasıyla şekil, lokasyon ve ölçek parametreleri olarak adlandırılır.  $H_{\xi, \mu, \sigma}$  dağılım fonksiyonu şekil parametresi  $\xi > 0$  ise Frechet,  $\xi < 0$  ise Weibull, ve  $\xi = 0$  (limit durumunda) ise Gumbel dağılımına yakınsar [2]. GUD ailesi çokça çalışılmış olup, Kotz ve Nadarajah detaylı bir şekilde bu çalışmalardan bahsetmektedir [46].

## 2.2. Tek Değişkenli POT (Eşik Seviyesi Aşım) Yöntemi

POT yöntemi, dalga yükseklikleri, sel, rüzgar hızları, sigorta talepleri gibi uç değer olaylarını tanımlamak için birçok alanda kullanılmıştır. Bu yaklaşım, yüksek bir eşik seviyesi seçimini aşan bağımsız uç gözlemlere ilişkin bir model sağlar. Bu yöntem için bir eşik seviyesi belirtilmesine ihtiyaç vardır. Ancak bu eşik seviyesi çok yüksek alındığında modelleme yapabilmek için yeteri kadar çoklukta veri elde edilemeyecek veya çok düşük alındığında ise uç olmayan değerler hesaba katılmak durumunda kalınacaktır. Aşağıda eşik değeri  $u$  olarak belirtilecektir.

Rasgele değişkenlerin belirlenen bir uç eşik değerini aşması koşullu olasılığınının asimptotik yapısı literatürde önemli bir yer tutmakta olup POT yöntemi olarak adlandırılmaktadır. Bu yöntemi de iki farklı şekilde incelemek mümkündür.

### Genelleştirilmiş Pareto (GP) Dağılımı

Klasik yaklaşımı temsil eden modellere alternatif olarak Smith [47], Davison ve Smith [14] ve Leadbetter tarafından [48] geliştirilen POT yaklaşımında, eşik seviyesini aşan  $(y_1, y_2, \dots, y_m)$  verilerimizin kümülatif olasılık dağılımı şu şekilde ifade edilir.

$$F_u(y) = \Pr\{X \leq u + y \mid X > u\} = \frac{F(u+y) - F(u)}{1 - F(u)}, \quad y > 0 \quad (2.3)$$

ve asimptotik olarak ilk defa Pickands tarafından [6] ortaya konulan teoriye göre eşik seviyesi dağılımın sağındaki en son noktaya giderken, eşik seviyesini aşan değer dağılım fonksiyonu aşağıdaki gibi yakınsar.

$$F_u(y) \rightarrow G_{\xi, \sigma}(y) \quad (2.4)$$

Bu dağılım 2 parametrelili GPD dağılımı olarak adlandırılır ve aşağıdaki gibi ifade edilir.

$$G_{\xi, \sigma} = \begin{cases} 1 - \left(1 + \frac{\xi}{\sigma} y\right)^{-\frac{1}{\xi}}, & \xi \neq 0 \\ 1 - e^{-\frac{y}{\sigma}}, & \xi = 0 \end{cases} \quad (2.5)$$

$\xi > 0$  ise Pareto dağılımına yakınsar.  $\xi < 0$  ise Beta dağılımına yakınsar.  $\xi = 0$  ise üssel (eksponansiyel) dağılım olarak adlandırılır.

### Nokta Süreci (Point-Process)

Eğer  $X$  rasgele değişkeni  $(\mu, \tau, \sigma)$  parametrelili GEV dağılım fonksiyonunun çekim-bölgesinde ise koşullu dağılım fonksiyonunun bir eşik değerini aşması aşağıdaki özelliği sağlamaktadır:

$$P(X > u + x | X > u) \sim \left(1 + \tau \frac{x - \mu}{\sigma + \tau(u - \mu)}\right)_+^{-1/\tau}, u \rightarrow \infty \quad (2.6)$$

Bu uygulamalarda çok kullanılan bir özellik olup çokça çalışılmıştır [5, 6]. Bu yaklaşımın bir avantajı da  $X$  rasgele değişkeninin bir eşik değerini aşmasına karşılık gelen nokta-sürecinin (point-process) yoğunluk ölçüsü  $\left(1 + \tau \frac{x - \mu}{\sigma}\right)_+^{-1/\tau}$  homojen olmayan Poisson süreçle yakınsanabilir [47, 49]. Yine bu yaklaşım pratikte parametrik olmayan yöntemlerin kullanımına imkan tanımaktadır. TUD teorisinden yararlanıp gerçek hayat problemlerine uygulaması çok da kolay olmamakta ve özellikle uygun model tercihi ve buna karşılık gelen  $\{a_n\} > 0$  ve  $\{b_n\}$  dizilerinin seçimi ve dağılım parametrelerinin tahmin edicilerinin bulunması problemin önemli bir parçasını oluşturmaktadır.

### Tekrarlama seviyesi (return level)

Eldeki verileri kullanarak oturtulmuş modellerin, GPD veya GEV, temel pratik kullanımı,  $N$  yıllık tekraralama seviyeleri hesaplamalarıdır. Böyle bir seviyeyi aşan bir olayın her  $N$  yılda bir oluşması beklenir.

Yeteri kadar büyük bir  $u$  eşik seviyesi için  $N$  yıllık tekraralama seviyesi  $z_N$  aşağıdaki gibi ifade edilir:

$$z_N = \begin{cases} u + \left(\frac{\sigma}{\xi}\right) \left[ (\phi_u N n_y)^\xi - 1 \right], & \xi \neq 0 \\ u + \sigma \log(N \phi_u N n_y) & , \xi = 0 \end{cases} \quad (2.7)$$

Burada,  $\phi_u = P(X > u)$  ve  $n_y$  her bloktaki gözlem sayısıdır (örneğin yıllık gözlemler için bu değer 365). Eldeki probleme GPD modeli oturtulduktan sonra parametrelerin en çok olabilirlik tahmin edicileri kullanılarak ve  $\phi_u$ 'nun tahmin edicisi olarak da eşik değeri aşan değerlerin tüm veriye olan oranı alınarak tekraralama seviyesi tahmin edilir.

Diğer yandan modelleme GEV ile yapılmışsa:  $N = \frac{1}{p}$  lik tekraralama seviyesi  $z_p$ , bu dağılımdan elde edilen parametre tahminleri kullanarak aşağıdaki gibi ifade edilir.

$$z_p = \begin{cases} \mu - \sigma/\xi[1 - \{-\log(1 - p)\}^{-\xi}], & \xi \neq 0 \\ \mu - \sigma \log\{-\log(1 - p)\} & , \xi = 0 \end{cases} \quad (2.8)$$

### 2.3. Tek Değişkenli ACER (Ortalama Koşullu Aşım Oranı) Yöntemi

ACER yöntemi belli bir seviyeyi aşan değerlerin koşullu saptanmış (k-1) önceki değerlerinin o seviyeyi aşmaması şartıyla Arvid Naess tarafından geliştirilmiş kuyruk dağılımını tahmin etmede kullanılan yeni bir yöntemdir [43]. ACER yönteminin Blok maksimum ve POT yöntemleriyle karşılaştırıldığında, datayı temsil eden uç değer dağılımını uç değer dağılım parametresine bağlı olmadan ifade edebilmesi en önemli özelliğidir. Gözlemlerin bağımsız ve özdeş dağılıma sahip olması şartı aranmamaktadır ve metodun türetilmesinde asimptotik argümanlara ihtiyaç duyulmamaktadır. Ayrıca durağan olmayan veri kümelerine de uygulanabilmektedir. Bu yüzden sezonsallık etkilerini hesaba katmaya gerek yoktur.

ACER metodu tek değişkenli zaman serilerinin uç değer dağılımlarının tahminine imkan sağlar. Bunu da kademeli şartlı yakınsama mantığını göz önüne alarak yapar. Bu kademeli şartlı yakınsama mantığından aşağıda bahsedilmiştir.

#### Kademeli şartlı yakınsama mantığı

Z(t) stokastik sürecimiz için, (0,T) zaman aralığında  $t_1, t_2, \dots, t_N$  anlarında gözlenen  $X_1, X_2, \dots, X_N$  değerlerinin olduğunu varsayalım. Uç değer olarak tanımlanan  $M_N = \max \{ X_j; j = 1,2,\dots,N \}$  rasgele değişkeninin dağılım fonksiyonu belirlenecektir. Özellikle, yeterince büyük  $\eta$  değerleri için  $P(\eta) = \Pr (M_N \leq \eta)$  olasılığı tahmin edilmek istenmektedir. Saatlik, günlük veya yıllık gözlenebilecek deniz seviyesi, basınç şiddeti gibi değerleri baz alarak  $P(\eta)$  tanımını aşağıdaki gibi ifade edebiliriz.

$$\begin{aligned} P(\eta) &= \Pr(M_N \leq \eta) = \Pr\{X_N \leq \eta, \dots, X_1 \leq \eta\} \\ &= \Pr\{X_N \leq \eta \mid X_{N-1} \leq \eta, \dots, X_1 \leq \eta\}. \\ &\quad \Pr\{X_{N-1} \leq \eta, \dots, X_1 \leq \eta\} \quad (2.9) \\ &= \prod_{j=2}^N \Pr\{X_j \leq \eta \mid X_{j-1} \leq \eta, \dots, X_1 \leq \eta\} \cdot \Pr(X_1 \leq \eta) \end{aligned}$$

Genel olarak  $X_j$  değerleri istatistiksel olarak birbirine bağımlıdır. Bunun yerine bütün  $X_j$  değerlerinin bağımsız olması varsayımıyla ki bu durumda klasik yaklaşıma imkan sağlayan formu aşağıdaki gibi elde edilir.

$$P(\eta) \approx P_1(\eta) = \prod_{j=1}^N Pr(X_j \leq \eta) \quad (2.10)$$

$X_j$ 'lerin,  $2 \leq j \leq N$  iken, arasındaki bağımlılık bir adım yakınsama yaklaşımıyla (one-step memory approximation),

$$Pr\{X_j \leq \eta | X_{j-1} \leq \eta, \dots, X_1 \leq \eta\} \approx Pr\{X_j \leq \eta | X_{j-1} \leq \eta\} \quad (2.11)$$

ifade edilip,  $P(\eta)$  aşağıdaki gibi elde edilir.

$$P(\eta) \approx P_2(\eta) = \prod_{j=2}^N Pr\{X_j \leq \eta | X_{j-1} \leq \eta\} \cdot Pr(X_1 \leq \eta) \quad (2.12)$$

Bir data daha fazla kullanılarak yapılan şartlı bir adım yakınsama (one-step memory approximation),

$$Pr\{X_j \leq \eta | X_{j-1} \leq \eta, \dots, X_1 \leq \eta\} \approx Pr\{X_j \leq \eta | X_{j-1} \leq \eta, X_{j-2} \leq \eta\} \quad (2.13)$$

$3 \leq j \leq N$  için gibi genelleştirilip,  $P(\eta)$  aşağıdaki gibi ifade edilir.

$$P(\eta) \approx P_3(\eta) = \prod_{j=3}^N Pr\{X_j \leq \eta | X_{j-1} \leq \eta, X_{j-2} \leq \eta\} \cdot Pr\{X_2 \leq \eta | X_1 \leq \eta\} \cdot Pr(X_1 \leq \eta) \quad (2.14)$$

Her  $k$  değeri için daha genel bir ifadeyle  $2 \leq k \leq N$  için aşağıdaki sonuç elde edilir.

$$P(\eta) \approx P_k(\eta) = \prod_{j=k}^N Pr\{X_j \leq \eta | X_{j-1} \leq \eta, X_{j-k+1} \leq \eta\} \cdot \prod_{j=2}^{k-1} Pr\{X_j \leq \eta | X_{j-1} \leq \eta, \dots, X_1 \leq \eta\} \cdot Pr(X_1 \leq \eta) \quad (2.15)$$

$\alpha_{1j}(\eta) = Pr(X_j > \eta)$  ve  $j = 1, 2, \dots, N$  olması şartıyla Denklem (2.10) aşağıdaki gibi

$$P(\eta) \approx P_1(\eta) = \prod_{j=1}^N (1 - \alpha_{1j}(\eta)) \quad (2.16)$$

yazılıp, verilerin bağımsızlık varsayımı altındaki yakınsamasıyla aşağıdaki sonuç elde edilir.

$$P(\eta) \approx F_1(\eta) = \exp\left(-\sum_{j=1}^N \alpha_{1j}(\eta)\right) \quad (2.17)$$

Tabiki verilerin bağımlı olduğu durumda  $j \geq k \geq 2$  için  $\alpha_{kj}(\eta) = Pr\{X_j > \eta | X_{j-1} \leq \eta, X_{j-k+1} \leq \eta\}$  olması şartıyla denklem (2.15) aşağıdaki gibi düzenlenir.

$$P(\eta) \approx P_k(\eta) = \prod_{j=k}^N (1 - \alpha_{kj}(\eta)) \cdot \prod_{j=1}^{k-1} (1 - \alpha_{jj}(\eta)) \quad (2.18)$$

(k - 1) tane aşılmayan adım üzerindeki şartlı aşım olasılığı denklem (2.18) kullanılarak  $\eta$  değerleri sonsuza giderken aşağıdaki gibi özetlenir.

$$P(\eta) \approx F_k(\eta) = \exp\left(-\sum_{j=k}^N \alpha_{kj}(\eta) - \sum_{j=1}^{k-1} \alpha_{jj}(\eta)\right) \quad (2.19)$$

k değerleri için yeterince büyük bir N sayısı belirlendiğinde  $F_N(\eta) = P(\eta)$  iken,  $F_k(\eta) \rightarrow P(\eta)$  ve  $\sum_{j=1}^{k-1} \alpha_{jj}(\eta)$  ifadesi  $\sum_{j=k}^N \alpha_{kj}(\eta)$ 'ye göre göz ardı edilecek şekilde küçük kalacağından  $\sum_{j=1}^{k-1} \alpha_{jj}(\eta)$  ihmal edilebilir. O halde denklem (2.19) durağan ve durağan olmayan veriler için daha sade biçimde aşağıdaki gibi yazılabilir.

$$F_k(\eta) = \exp\left(-\sum_{j=k}^N \alpha_{kj}(\eta)\right) \quad , k \geq 1 \quad (2.20)$$

### Ortalama koşullu aşım oranı (ACER) metodu

ACER metodu kullanılırken ilk olarak aşım oran fonksiyonu üretilecektir. Bu fonksiyon verilen bir seviyeyi aşma ihtimali için koşullu saptanmış (k-1) adım önceki aşamayan değerlerin sayısına bağlı olarak belirlenecektir. Bu seviye ise, ortalama aşım fonksiyonlarını temsil ettiğimiz grafikte (k-1) adım önceki aşamayan değerlerin sayısına bağlı olarak hesaplanır. Tabi bu belli bir seviyeyi aşamayan değerler aşım oran fonksiyonunun daha yüksek (k-1) adım mertebeleri için kuyruk bölgesinde özdeş bir hal alacaktır. Sonuç olarak seçilen gözlemsel ACER fonksiyonunun kuyruk yapısı aşağıdaki gibi bir formda olacaktır.

$$\varepsilon(\eta) = -q(\eta)(1 + a(\eta - b)^c)^{-1/\xi} \quad (2.21)$$

Bu kuyruk yapısı formu için,  $q(\eta)$  ve  $c$  asimptotik parametreler olarak kabul edildiğinde asimptotik GEV dağılımının kuyruk yapısı temsil edilebilecektir.

$q(\eta) = c = 1$  olması durumu basitçe GEV dağılımını temsil eder.  $q(\eta) = c = 1$  özel hali için,  $\xi \rightarrow 0$  değerine yakınsarken ACER methodunun kuyruk yapısı Gumbel dağılımında olması durumu için geliştirildi. Ağır kuyruk yapılı verilerle ilgilenildiğinde bu durum geçerli değildir çünkü veriyi temsil eden uç değer dağılımı Frechet dağılımı olacaktır. O halde  $\xi > 0$  halinde aşım oran seviyesi  $\eta$  için aşağıdaki gibi form yeniden inşa edilebilir.

$$\varepsilon(\eta) = q(\eta)(1 + a(\eta - b)^c)^{-1/\xi} \approx q(1 + a(\eta - b)^c)^{-1/\xi} \quad (2.22)$$

Yukarıdaki denklemde  $q(\eta)$  değeri denklemin geri kalan ifadesine göre çok etkisiz bir şekilde değiştiği için  $q$  sabiti olarak kabul edildi. Bu varsayım altında denklem (2.22)'nin parametreleri kare hata fonksiyonunun (square error function) minimize edilmesiyle tahmin edilebilir. Bu işlemler aşağıdaki denklemlerle ifade edilir.

$$F(a,b,c,q, \xi) = \sum_{i=1}^n w_i [\log(\hat{\varepsilon}(\eta_i)) - \log(q) + \xi^{-1} \log(1 + a(\eta_i - b)^c)]^2 \quad (2.23)$$

Ağırlıklı doğrusal regresyon (weighted linear regression)  $w_i$  kullanılarak, her  $\xi, q$  değeri için  $\xi^{-1}$  ve  $q$  parametresi aşağıdaki denklemlerle hesaplanır.

$$\xi^{-1} = \frac{\sum_{i=1}^n w_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n w_i (x_i - \bar{x})^2} \quad (2.24)$$

$$\log(q) = \bar{y} - \hat{\xi}^{-1} \bar{x} \quad (2.25)$$

Tabi burada  $x_i = \log(1 + a(\eta_i - b)^c)$ ,  $\bar{x} = \frac{\sum_{i=1}^n w_i x_i}{\sum_{i=1}^n w_i}$ ,  $y_i = \log(\hat{\varepsilon}(\eta_i))$  ve  $\bar{y} = \frac{\sum_{i=1}^n w_i y_i}{\sum_{i=1}^n w_i}$  şeklinde verilmiştir. Böylece minimizasyon problemi 3 parametreye kadar indirgenmiş oldu. Daha sonrasında  $a$ ,  $b$  ve  $c$  parametreleri denklem (2.23) kullanılarak minimizasyon yöntemiyle hesaplanır.





### 3. İKİ DEĞİŞKENLİ UÇ DEĞER YÖNTEMLERİ

#### 3.1 İki Değişkenli Blok Maksimum Yöntemi

Tek değişkenli uç değerler teorisini çok değişkenli uç değerler teorisine genellemek çok değişkenli gözlemleri sıralamanın doğal bir yolu olmadığından açık değildir. En yaygın izlenen yol gözlemlerin bileşenlere göre sıralaması üzerinedir ve uygun istatistik metodolojileri nispeten yakın zamanlarda geliştirilmiştir [50-52]. Bu çalışmada iki değişkenli uç değer yaklaşımları üzerinde duracağımızdan, bunun teorisinden kısaca bahsedelim.  $\mathbf{X} = (X_1, X_2)$ , bilinmeyen bir 2 boyutlu  $\mathbf{F}(\mathbf{x})$  dağılım fonksiyonuna sahip rasgele vektör olmak üzere eğer aşağıdaki eşitliği sağlayacak öyle  $\{a_{n,j}\} \in R^n$  ve  $\{b_{n,j}\} \in R^n$ ,  $a_{n,j} > 0$  ve  $b_{n,j} \in R$ , her  $j = 1, 2$  dizileri ve dejenere olmayan bir  $\mathbf{G}$  dağılım fonksiyonu mevcut ise

$$P \left\{ \frac{\max_{i=1, \dots, n} X_{i,j} - b_{n,j}}{a_{n,j}} \leq x \right\} = F^n(a_{n,j}x + b_{n,j}) \rightarrow G(x), \quad n \rightarrow \infty \quad (3.1)$$

Bu durumda  $\mathbf{F}$ ,  $\mathbf{G}$ 'nin (maksimum) çekim bölgesindedir denir ve  $\mathbf{G} = (G_1, G_2)$  ise iki değişkenli uç değerler dağılım fonksiyonu olarak adlandırılır.

Gözlemlerin sıralanması probleminin yanı sıra ÇUD teorisindeki bir diğer zorluk ise uç değerler arasındaki ilişkiyi (bağımlılığı) tespit etmektir. TUD'nin aksine ÇUD dağılımları sonlu boyutlu parametre vektörüyle indekslenmiş bir parametrik aile ile ifade edilemez. Bunun yerine ÇUD ailesi, örneğin konveks fonksiyonlar sınıfı veya sonlu ölçüler sınıfı gibi yapılarla indekslenebilir. Değişkenler arasındaki korelasyon ilişkisi dağılımın kuyruğunun yapısını modellemede invariyanlık sorunundan ötürü kullanılamaz. Bu bağımlılık yapısını tanımlayacak birçok farklı yollar mevcuttur. Bunlardan bazıları: a) üstel ölçüler, b) spektral ölçüler, c) istikrarlı kuyruk bağımlılığını yansıtan bir fonksiyon d) kopula metodu e) parametrik olmayan yöntemler gibi. Bunlardan kısaca bahsedelim.

a) Denklem (3.1) deki 2-boyutlu  $G$  dağılım fonksiyonu  $G(\mathbf{x}) = \exp(-\mu[-\infty, \infty] \setminus [-\infty, \mathbf{x}])$  olacak şekilde  $[-\infty, \infty)$  aralığında bir  $\mu$  ölçüsü vardır ve  $G$  üstel ölçü olarak da adlandırılır.

b) Üstel ölçünün homojenlik özelliği ( $\mu(s) = \mu(\cdot)/s, 0 < s < \infty$ )  $G$  dağılım fonksiyonunun kutupsal koordinatlar cinsinden farklı ifadelerini mümkün kılmaktadır. Bunlardan bir tanesi spektral ölçüdür. Marjinal dağılımları birleştirilmiş Frechet olan 2-boyutlu dağılım fonksiyonu  $G$ 'nin bir ÇUD dağılım fonksiyonu olması için gerekli ve yeterli koşul bir birim simpleks  $S_p = \{\mathbf{w} = (w_1, w_2): w_i > 0, w_1 + w_2 = 1\}$  üzerinde

$$G(\mathbf{x}) = \exp \left\{ -2 \int_{S_p} \max\left(\frac{w}{x_1}, \frac{1-w}{x_2}\right) H(dw) \right\} \quad (3.2)$$

olacak şekilde  $\int_{S_p} w_j H(dw) = \frac{1}{2}, j = 1, 2$  kısıtını sağlayan  $H$  sonlu ölçüsünün mevcut olmasıdır [50,51]. Buradaki  $H$  ölçüsü spektral ölçü olarak adlandırılmaktadır.

$$G^n(\mathbf{x}) = G^n(x_1, x_2) = G\left(\frac{x_1}{n}, \frac{x_2}{n}\right) \quad (3.3)$$

Yani  $\mathbf{X} = (X_1, X_2)$   $G$  dağılım fonksiyonuna sahipse  $M_n = (\max_{i=1, \dots, n} X_{i,1}, \max_{i=1, \dots, n} X_{i,2})$  de aynı dağılıma sahiptir. Yukarıda bahsedilen özellikler iki değişkenli uç değer dağılımları sınıflarını tanımlamaktadır.

c) İki değişkenli uç değerler (dağılım) metodunun başka bir karakterizasyonu da, 1981 yılında Pickands tarafından tanımlanan bağımlılık fonksiyonudur [36]. İki değişkenli uç değerler dağılım fonksiyonunun bağımlılık yapısını ifade etmekte en yaygın olarak karşımıza çıkan  $A(t)$  Pickands bağımlılık fonksiyonunun asimptotik özellikleri 1991 yılında Deheuvels tarafından incelenmiştir [37]. Birçok kaynakta Pickands bağımlılık fonksiyonu tahmin edicisi için marjinal dağılımların bilindiği varsayımı yapılmaktadır. Cap'era'a vd. 1997 yılında bu kısıtı kaldırmışlardır [38].

d) Genelde, marjinal dağılımları  $F_1, F_2, \dots, F_d$  olan çok değişkenli bir dağılım fonksiyonu  $F$ 'e karşılık gelen ve marjinal dağılımları  $(0,1)$  aralığında düzgün dağılıma sahip olan bir  $C_F$  dağılım fonksiyonu vardır öyle ki  $F(\mathbf{x}) = C_F \{F_1(x_1), F_2(x_2), \dots, F_d(x_d)\}, \mathbf{x} \in R^d$  [24]. Buradaki  $C_F$ ' e bir kopula denir. Farklı kopula aileleri [25, 53] ve bunlardan Archimedean kopula (Gumbel) ailesi bu çalışmada da kullanılabilecek hidrolojik analizler için önerilmiştir [54].

e) İki deęişkenli daęılım fonksiyonunda bileşenler arasındaki baęımlılık Kendall tau ve Spearman's rho ölçüleri ile daęılımdan baęımsız olarak saptanabilir ve Pickands baęımlılık fonksiyonu cinsinden ifade olunabilir [35].

Bu çalışmada yukarıda belirtilen c) istikrarlı kuyruk baęımlılıęını yansıtan bir fonksiyon cinsinden, d) kopula metodu ve e) parametrik olmayan yöntemlerden yararlanılacaktır.

Denklem (3.2)'de tanımlanan  $H$  daęılım fonksiyonu için süreklilik koşulu gerekli deęildir. Örneęin,  $P(w = 0) = P(w = 1) = 0.5$  alındığında iki deęişkenli uç deęer daęılımını  $G(x_1, x_2) = \exp\{-\frac{1}{x_1}, \frac{1}{x_2}\}$  dır ve bu  $X_1, X_2$  nin baęımsız olduęunu söyler.

Eęer  $P(w = 0.5) = 1$  alırsak  $G(x_1, x_2) = \exp\{-\max(\frac{1}{x_1}, \frac{1}{x_2})\}$  olur ve bu da  $X_1, X_2$  nin tam baęımlı olduęunu söyler.

Birçok iki deęişkenli (ve çok deęişkenli) uç deęer daęılımları Tawn [40], Coles ve Tawn [55], Joe [56], Tawn [57] çalışmalarında önerilmiştir. Kotz ve Nadatajah'ın 2000 yılındaki çalışmalarında daha kapsamlı literatür bulunabilir [46]. Bunlardan en yaygın olarak bilinen iki deęişkenli uç deęer modeli için geliştirilen limit daęılımlarını lojistik (LOG), asimetrik lojistik (ALOG), negatif lojistik (NEGLOG) ve asimetrik negatif lojistik (ANEGLOG) şeklinde sayabiliriz. Bu daęılımlarla birlikte bu çalışmada deęerlendirdiğimiz dięer iki deęişkenli uç deęer daęılımları da Çizelge 3.1'de belirtilmiştir. Baęımlılık parametresi  $a$  ve asimetri parametreleri  $t_1, t_2$  ile gösterilmiştir.

**Çizelge 3.1:** İki değişkenli uç değer dağılımları.

Model	Yoğunluk Fonksiyonları	Tanımlama
LOG	$G(x, y) = \exp\{-(x^{1/a} + y^{1/a})^a\}$	<ul style="list-style-type: none"> <li>• <math>0 &lt; a \leq 1</math>.</li> <li>• <math>a \rightarrow 1</math>, bağımsız değişkenler.</li> <li>• <math>a \rightarrow 0</math>, tamamen bağımlı.</li> </ul>
ALOG	$G(x, y) = \exp\{-(1 - t_1)x - (1 - t_2)y - [(t_1x)^{\frac{1}{a}} + (t_2y)^{\frac{1}{a}}]^a\}$	<ul style="list-style-type: none"> <li>• <math>0 &lt; a \leq 1</math> ve <math>0 \leq t_1, t_2 \leq 1</math>.</li> <li>• <math>t_1 = t_2</math> durumunda LOG modeline indirgenir.</li> <li>• <math>a = 1, t_1 = 0</math> veya <math>t_2 = 0</math> da bağımlı.</li> <li>• <math>t_1 = t_2 = 1</math> ve <math>a</math> sifıra yakınsadığında tam bağımlılık oluşur.</li> </ul>
HR	$G(x, y) = \exp\left[-x \varphi\left\{\frac{1}{a} + \frac{a}{2} \ln\left(\frac{x}{y}\right)\right\} - y \varphi\left\{\frac{1}{a} + \frac{a}{2} \ln\left(\frac{y}{x}\right)\right\}\right]$	<ul style="list-style-type: none"> <li>• <math>\varphi</math> standart normal dağılımı temsil eder ve <math>a &gt; 0</math> olması şartı vardır.</li> <li>• <math>a \rightarrow 0</math>, bağımsız değişkenler.</li> <li>• <math>a \rightarrow \infty</math>, tam bağımlı.</li> </ul>

NEGLOG	$G(x, y) = \exp\{-x - y + (x^{-1/a} + y^{-1/a})^{-a}\}$	<ul style="list-style-type: none"> <li>• <math>a &gt; 0</math>.</li> <li>• <math>a \rightarrow 0</math>, bağımsız değişkenler.</li> <li>• <math>a \rightarrow \infty</math>, tam bağımlı.</li> </ul>
ANEGLOG	$G(x, y) = \exp\left\{-x - y - \left[(t_1 x)^{-\frac{1}{a}} + (t_2 y)^{-\frac{1}{a}}\right]^{-a}\right\}$	<ul style="list-style-type: none"> <li>• <math>a &gt; 0</math> ve <math>0 \leq t_1, t_2 \leq 1</math>.</li> <li>• <math>t_1 = t_2 = 1</math> durumunda NEGLOG modeline indirgenir.</li> <li>• <math>a, t_1</math> veya <math>t_2</math> sıfıra yakınsadığında bağımsızlık elde edilir.</li> <li>• <math>t_1 = t_2 = 1</math> ve <math>a</math> sonsuza gittiğinde tam bağımlılık hali oluşur.</li> </ul>
BILOG	$G(x, y) = \exp\{-x(q)^{1-\alpha} - y(1-q)^{1-\beta}\}$	<ul style="list-style-type: none"> <li>• <math>(1 - \alpha)(x)(1 - q)^{\beta} - (1 - \beta)(y)(q)^{\alpha} = 0</math> denkleminin kökü <math>q=q(x, y, \alpha, \beta)</math>, <math>0 &lt; \alpha, \beta &lt; 1</math> parametre şartları için verilmiştir.</li> <li>• <math>\alpha = \beta</math> durumunda LOG modeline denktir.</li> <li>• <math>\alpha = \beta \rightarrow 0</math>, tam bağımlı.</li> <li>• <math>\alpha = \beta \rightarrow 1</math>,</li> </ul>

		<p>bağımsızlık.</p> <ul style="list-style-type: none"> <li>Alpha ve beta parametrelerinden biri sabit iken diğeri bir değerine yaklaşırken yine bağımsız olur.</li> </ul>
NEGBILOG	$G(x,y)=\exp\{-x-y+xq^{1+\alpha}-y(1-q)^{1+\beta}\}$	<ul style="list-style-type: none"> <li><math>(1+\alpha)(x)(q)^{\alpha} - (1+\beta)(y)(1-q)^{\beta}=0</math> denkleminin kökü  <math>q=q(x,y,\alpha,\beta)</math>,  <math>0&lt;\alpha</math> ve <math>0&lt;\beta</math> parametre şartları için verilmiştir.</li> <li>Alpha=beta durumunda NEGLOG modeline eşittir.</li> <li>Alpha=beta<math>\rightarrow 0</math>, tam bağımlılık.</li> <li>Alpha=beta <math>\rightarrow \infty</math>, bağımsızlık.</li> <li>Alpha ve beta parametrelerinden biri sabit iken diğeri sonsuza yaklaşırken yine bağımsız olur.</li> </ul>
CT	$G(x,y)=\exp\{-x[1-\text{Be}(q;\alpha+1,\beta)]-y\text{Be}(q;\alpha,\beta+1)\}$	<ul style="list-style-type: none"> <li><math>0&lt;\alpha</math> ve <math>0&lt;\beta</math> parametre şartları için  <math>q = \frac{\alpha(y)}{\alpha(y)+\beta(x)}</math> şeklinde verilir.</li> </ul>

		<ul style="list-style-type: none"> <li>• <math>Be(q; \alpha, \beta)</math> <math>q</math>'daki beta dağılım fonksiyonudur.</li> <li>• <math>\alpha = \beta \rightarrow \infty</math>, tam bağımlılık.</li> <li>• <math>\alpha = \beta \rightarrow 0</math>, bağımsızlık.</li> <li>• <math>\alpha</math> ve <math>\beta</math> parametrelerinden biri sabit iken diğeri sıfır değerine yaklaşırken yine bağımsız olur.</li> </ul>
AMIX	$A(t) = 1 - (\alpha + \beta)t + (\alpha)t^2 + (\beta)t^3$	<ul style="list-style-type: none"> <li>• <math>\alpha \in [0, 1.5]</math> ve <math>\beta \in [-0.5, 0.5]</math> olması şartıyla elde edilir.</li> <li>• <math>\alpha = \beta = 0</math>, bağımsızlık.</li> <li>• Tam bağımlılık elde edilemez.</li> </ul>

### 3.2. İki Değişkenli POT Yöntemi

$(Z_1, Z_2)$  gözlenen rasgele vektör olmak üzere,  $(u_1, u_2)$  seçilen eşik seviyesi olsun. Öyleyse,  $(X_1, X_2) = (Z_1 - u_1, Z_2 - u_2)$  eşik değerini aşan rasgele vektörü ifade eder. Eşik aşım değerleri için genelleştirilmiş iki değişkenli Pareto dağılımının farklı gösterimleri mevcuttur, bunlardan Rootzén and Tajvidi'nin 2006 yılındaki çalışmalarında kullandıkları gösterim aşağıdaki gibidir [58].

$$H(x_1, x_2) = \frac{-1}{\log G(0,0)} \log \frac{G(x_1, x_2)}{G(\max(x_1, 0), \max(x_2, 0))} \quad (3.4)$$

Burada,  $0 < G(0,0) < 1$ . Denklem (3.4) en az bir bileşen için verilen eşik seviyesini aşan gözlemlerin limit dağılımını ifade eder. Aşağıdaki dönüşümle marjinaleri birleştirilmiş Frechet dağılımı elde edilir.

$$t_i = t_i(x) = \left(1 + \frac{(x_i - \mu_i)}{\sigma_i}\right)^{1/\xi_i} \quad (3.5)$$

Burada,  $1 + \frac{(x_i - \mu_i)}{\sigma_i} > 0$  ve  $\sigma_i > 0$ ,  $i = 1, 2$ . Genelleştirilmiş iki değişkenli Pareto dağılımının mevcut olması halinde,

$$h(x) = \frac{\partial H}{\partial x_1 \partial x_2}(x) = \frac{\partial}{\partial x_1 \partial x_2} \left(1 - \frac{\log G(x)}{\log G(0)}\right) = \frac{t_1'(x_1)t_2'(x_2)}{v(t_1(0), t_2(0))} \times \frac{\partial v}{\partial t_1 \partial t_2}(t_1(x_1), t_2(x_2))'$$

dir.

Not:  $H_1(x_1) = H(x_1, \infty)$ , denklem (2.1)'de belirtilen tek boyutlu Genelleştirilmiş Pareto dağılımına sahip olduğu söylenemez. Sadece  $X_1 > 0$  verildiğinde  $X_1$ 'in koşullu dağılımı Genelleştirilmiş Pareto dağılımına sahiptir.

### 3.2.1 İki değişkenli uç değer dağılım fonksiyonu, pickand bağımlılık fonksiyonu ve kopula ilişkisi

İki değişkenli uç değerler dağılım fonksiyonunu Pickand bağımlılık  $A(t)$  cinsinden aşağıdaki şekilde ifade edebiliriz.

$$G(t_1, t_2) = \left(\frac{1}{t_1} + \frac{1}{t_2}\right)A\left(\frac{t_1}{t_1+t_2}\right) \quad (3.6)$$

burada  $t_i$  Denklem (3.5)'de belirtilmiştir. Ayrıca Pickand bağımlılık cinsinden kopula  $C(u, v)$ ' nin ifadesi aşağıdadır.

$$C(u, v) = \exp \left\{ \log(uv) * A\left(\frac{\log(v)}{\log(uv)}\right) \right\} \quad \forall u, v \in (0,1) \quad (3.7)$$

Uç değer kopularının Pickands bağımlılık fonksiyonu cinsinden ifadeleri:

- Gumbel kopula:  $\theta \in [1, \infty)$  için,

$$A(t) = (t^\theta + (1-t)^\theta)^{1/\theta}$$

- Galambos kopula:  $\theta \in [0, \infty)$  için,

$$A(t; \theta) = 1 - (t^{-\theta} + (1-t)^{-\theta})^{-1/\theta}$$

- Hüsler-Reiss kopula:  $\theta \in [0, \infty)$  için,



$$A(t; \theta) = (1 - t)\phi\left(\theta + \frac{1}{2\theta} \log\left(\frac{1-t}{t}\right)\right) + t\phi\left(\theta + \frac{1}{2\theta} \log\left(\frac{t}{1-t}\right)\right)$$

- Tawn kopula:  $\theta \in [0, 1]$  için,

$$A(t; \theta) = 1 - \theta t + \theta t^2$$

- t-EV kopula:  $\theta \in (0, \infty) \times (-1, 1)$  için, burada,  $t_\theta$ , t-dağılım fonksiyonudur.

$$A(t; \theta) = t t_{\theta_1+1} \left( \frac{\left(\frac{t}{1-t}\right)^{\frac{1}{\theta_1}-\theta_2} (\theta_1 + 1)^{\frac{1}{2}}}{(1-\theta_2^2)^{\frac{1}{2}}} \right) + (-t) t_{\theta_1+1} \left( \frac{\left(\frac{1-t}{t}\right)^{\frac{1}{\theta_1}-\theta_2} (\theta_1 + 1)^{\frac{1}{2}}}{(1-\theta_2^2)^{\frac{1}{2}}} \right)$$

### 3.2.2 Kopula yaklaşımı

Aşağıda temel kopula ailelerinden bahsedilecektir ve ilerleyen aşamalarda bunların arasından uygun olanı seçilecektir.

Arşimedyan Kopula Aileleri: Arşimedyan kopulalar, kopula ailelerinin en önemli sınıfıdır. İlk defa 1965'te Ling'in makalesinde "arşimedyan" terimi kullanılmıştır. Birçok kopula ailesi Arşimedyan'dır ve Arşimedyan kopulalar çok çeşitli ve farklı bağımlılık yapılarına sahiptir. Çoğu kopula fonksiyonunun aksine bu kopulalar Sklar teoremi kullanılarak türetilmezler. Oluşturulmalarının kolay olması, bu sınıfa ait çok çeşitli kopula ailelerinin bulunması ve bu sınıfa ait kopulaların birçok faydalı cebirsel özelliğinin olması sebebiyle Arşimedyan kopulalarının uygulama alanları oldukça geniştir [59]. Daha çok finans uygulamalarında kullanılmasının yanında hidroloji alanında da kullanılmaktadır. Clayton, Frank, Normal ve Gumbel kopulaları hidroloji ve istatistikte kullanılan başlıca Arşimedyan kopulalarıdır. Bunun dışında eliptik kopulalardan Gaussian ve Student-t kopulaları yaygın olarak kullanılmaktadır. Boyutu iki olan kopulaların bir tane pickand bağımlılık fonksiyonu vardır.

Rasgele değişkenlerimiz arasındaki bağımlılık ilişkisi  $\pm 1$  değerini aldığı durumlar hariç bu ilişkinin pozitif veya negatif olması bize kullanacağımız kopulalarla ilgili bir takım tahmini önbilgiler verecektir. Arşimedyan kopulalardan Clayton ve Ali-Mikhail-Haq kopula aileleri genellikle pozitif bağımlılık için uygulanabilir. Uç Değer Kopulalardan (Extreme Value Copula) Gumbel, Galambos, Kimeldorf-Sampson ve Hüsler Reiss kopulaları negatif ilişkiler için tercih edilmez. Bunun dışında Frank kopula ailesi hem negatif hem de pozitif ilişki için kullanılabilir.

Kopular kullanılarak yapılan istatistiksel çıkarımlarda kullanılan yöntemleri parametrik ve parametrik olmayan yöntemler olarak ikiye ayırabiliriz. MLE ve IFM yöntemi parametrik yöntemlerdir. Kendall Tau ve Spearman Rho ölçülerine dayanan yöntemler ise parametrik olmayan yöntemlerdir. MLE yöntemi çok boyutlu durumlarda yoğun hesaplamalar gerektirmektedir. Çünkü hem marjinal dağılımların parametrelerini hem de kopula ile gösterilen bağımlılık yapısının parametrelerini hesaplamak gerekmektedir. Bu açıdan alternatif bir yöntem olan IFM yöntemi kullanılmaktadır. Bu yöntem ile ilk aşamada tek boyutlu marjinal dağılımların parametresi hesaplanıp, sonraki aşamada bu hesaplanan parametre yardımıyla kopula parametresi hesaplanır.

Bu çalışmada kullanılan başlıca uç değer kopula fonksiyonları ve bu kopula fonksiyonlarının  $\theta$  parametre uzayı Çizelge 3.2’de verilmiştir.

**Çizelge 3.2:** Uç değer kopula fonksiyonları.

Copula	$C_{\theta}(u, v)$	$\theta$ uzayı
Gumbel	$\exp[-(\ln u^{\theta} + \ln v^{\theta})^{1/\theta}]$	$[1, \infty)$
Galambos	$uv \exp[(\ln u^{-\theta} + \ln v^{-\theta})^{-1/\theta}]$	$[0, \infty)$
Husler-Reiss	$\exp[-(\ln u)\varphi\{1/\theta + \frac{\theta}{2} \ln(\frac{\ln u}{\ln v})\} - (\ln v)\varphi\{1/\theta + \frac{\theta}{2} \ln(\frac{\ln v}{\ln u})\}]$	$[0, \infty)$
TEV	$\exp[-h(-\log u, -\log v)]$	$(0, \infty)$
Tawn	$\exp[(\log u + \log v)A[(\log v)/\log(uv)]]$	$[0, 1]$

Bu tabloda  $\varphi$  standart normal dağılımı ifade etmektedir.  $h(x,y) = (x+y) h(\frac{x}{x+y}, \frac{y}{x+y})$  ve A fonksiyonundaki asimetrik  $t_1$  ve  $t_2$  parametreleri için  $A(x,y) = \exp\{-(1-t_1)x - (1-t_2)y - [(t_1x)^{\frac{1}{\theta}} + (t_2y)^{\frac{1}{\theta}}]^{\theta}\}$  şeklinde tanımlıdır.

#### **Birleşik tekrar seviyesi:**

$X_i > x_i$ , olayı için tekrar seviyesi aşağıdaki şekilde tanımlandığından daha önce bahsetmiştik.

$$T_{x_i} = \frac{1}{1-F(x_i)}, \quad i = 1, 2. \quad (3.8)$$

Burada,  $F(x_i) = P(X_i \leq x_i)$  dir. Diğer yandan  $X_1, X_2$ 'nin bileşik tekrar seviyesi  $T_{x_1, x_2}$  için farklı şekilde genellemeler yapılmıştır, bu çalışmada da kullanacağımız tanım,  $X_1$  veya  $X_2$  den en az bir tanesinin aşılması ( $X_1 > x_1$  veya  $X_2 > x_2$ ,  $X_1 > x_1$  ve  $X_2 > x_2$ ) olayına göre Yue ve Wang tarafından [60] verilen tanım,

$$T_{x_1, x_2} = \frac{1}{1-F(x_1, x_2)} \quad (3.9)$$

Eğer her blok büyüklüğü (örneğin yıl) için birden fazla uç değerden söz edilebilirse,

$$T_{x_1, x_2} = \frac{1}{k(1-F(x_1, x_2))} \quad (3.10)$$

Burada,  $k$  ortalamada seçilen blok büyüklüğünde bulunan uç değer gözlem sayısıdır.

$X_1$  ve  $X_2$  değişkenleri için bağımlılık parametresi tahmin edilebilirse bu ifadeyi kopula cinsinden de yazmak mümkündür. Örneğin, bağımlılık parametresi  $0 < \theta \leq 1$  olan iki değişkenli LOG) dağılım için  $F(x_1, x_2) = \exp\left\{-\left(y_1^{-\frac{1}{\theta}} + y_2^{-\frac{1}{\theta}}\right)^\theta\right\}$ ,  $x_1$  ve  $x_2$  için Frechet dönüşümü uygulanarak  $y_1$  ve  $y_2$  elde edilir.

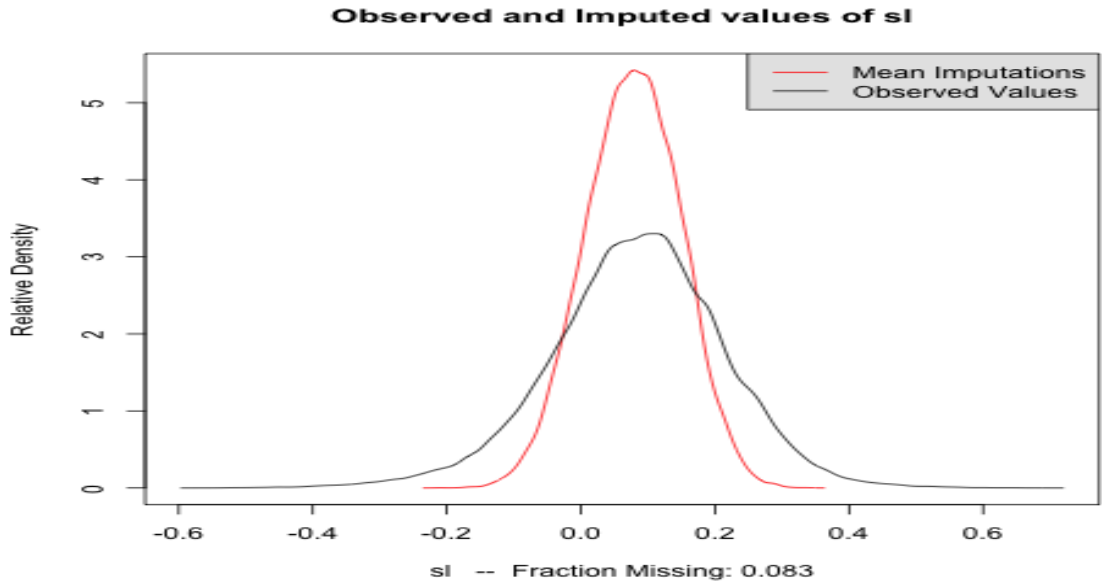


## 4. UYGULAMALAR

### 4.1 Gereçler

Uygulamaya geçmeden önce uygulama süresince gereç olarak kullanacağımız tek ve iki değişkenli uç değerler yöntemleri için ilgili tanımlar, teoremler ve özellikler aşağıda verilecektir.

Öncelikle çalışmada belirtilmeyen ancak verilerin tedarik edilmesi sonucunda teknik nedenlerden ötürü zaman serilerinde meydana gelen eksiklikler istatistiksel metotlarla doldurulmuştur. Bu eksik verilerin gözlenmiş ve gözlenememiş bütün verilerin çoklu normal dağılımdan geldiği varsayımı altında R istatistik programı kullanılarak tamamlanması gerçekleştirilmiştir. Bu yaklaşımın çok genel olmakla beraber diğer bilinen yöntemler kadar iyi çalıştığı söylenebilir [61]. Şekil 4.1’de bu varsayım resmedilmiştir.



**Şekil 4.1:** Eksik verilerin çoklu ‘imputation’ yöntemiyle doldurulup ortalamalarının dağılımları.

Parametre tahmin edicisi olarak en çok olabilirlik tahmin edicileri ve parametrik olmayan yöntemler kullanılmıştır.

Kullanılan testler olabirlik oran testi, uyum iyiliđi testi, durađanlık testleri ise KPSS ve Augmented Dickey-Fuller olmuřtur.

Diagnostik arařlarımız olarak Q-Q çizimler ve Chi-Plot kullanılmıřtır.

**Tanım** (Durađanlık): Rasgele bir süreç  $X_1, X_2, \dots$  verilen  $\{i_1, \dots, i_k\}$  keyfi tam sayılar kümesi ve herhangi bir  $m$  tamsayı deđeri için  $(X_{i_1}, X_{i_2}, \dots, X_{i_k})$  ve  $(X_{i_1+m}, X_{i_2+m}, \dots, X_{i_k+m})$  rasgele deđişkenlerinin birleşik olasılık dağılım fonksiyonları özdeş ise durađandır denir.

Veriler tamamlandıktan sonra uç deđerler yöntemi ile modelleme yapabilmek için temel varsayımlardan olan durađanlığın sağlanması ve mevsimsellik etkilerinin olmaması gerekmektedir. Bu çalışmada seçilen blok büyüklükleri için durađanlık koşulunun sağlanıp sağlanmadığını KPSS ve/veya Augmented Dickey-Fuller test yöntemi kullanarak incelenmiştir.

**Teorem** (Sklar):  $H(x, y)$ , marjinal olasılık dağılım fonksiyonları  $F(x)$  ve  $G(y)$  olan  $(X, Y)$  nin birleşik dağılım fonksiyonu olarak tanımlansın. Öyleyse reel düzlemdeki her  $x$  ve  $y$  rastsal deđerleri için  $H(x, y) = C_H(F(x), G(y))$  şeklinde bağımlılığı ifade eden bir  $C_H$  kopulası vardır. Aynı şekilde bunun aksini de ifade edersek; eđer  $C_H$  bir kopula ve  $F(x)$  ve  $G(y)$  dağılım fonksiyonları ise  $C_H(F(x), G(y)) = H(x, y)$  şeklinde tanımlı birleşik dağılım fonksiyonudur. Ayrıca önemli bir sonuç da eđer  $F(x)$  ve  $G(y)$  dağılım fonksiyonları sürekli ise kopula fonksiyonumuz tektir.

**Tanım** (Kopula): Eđer  $C_F(0, u) = C_F(u, 0) = 0$ ,  $C_F(1, u) = C_F(u, 1) = u$  ve  $C_F(u_2, v_2) - C_F(u_2, v_1) - C_F(u_1, v_2) + C_F(u_1, v_1) \geq 0$  şartları tüm  $0 \leq u_1 \leq u_2 \leq 1$  ve  $0 \leq v_1 \leq v_2 \leq 1$  deđerleri için sağlanıyorsa  $C_F: [0,1] \times [0,1] \rightarrow [0,1]$  iki deđerkenli kopula olarak adlandırılır.

**Tanım** (Uç deđer kopula): Her  $t > 0$  için, iki deđerkenli  $C$  kopulası  $C^t(u, v) = C(u^t, v^t)$  ilişkisini sağlıyor ise bu  $C$  kopulası iki deđerkenli uç deđer kopula olarak adlandırılır.

Pickands bağımlılık fonksiyonu iki deđerkenli uç deđer kopularının kullanımına büyük ölçüde imkan tanımıştır.

**Tanım** (Pickand bağımlılık fonksiyonu):  $F_X(\xi)$  ve  $G_Y(\eta) \in [0,1]$  marjinal dağılım fonksiyonları için,  $C$  uç deđer kopulası  $[0,1]$  aralığında en az bir reel deđerli  $D$  fonksiyonu için

$C(F_X(\xi), G_Y(\eta)) = \exp\{ \log(F_X(\xi) G_Y(\eta)) * D(\log(G_Y(\eta)) / \log(F_X(\xi) G_Y(\eta))) \}$   
 şartını sağlıyorsa  $D$  fonksiyonu Pickand bağımlılık fonksiyonu olarak adlandırılır.

Özel halde Pickand bağımlılık fonksiyonunun “0” değerini alması değişkenlerin bağımsız ve “1” değerini alması hali ise değişkenlerin tam bağımlı olduğunun göstergesidir.

Kopulaların seçiminde kullanılan Chi-Plot, kuyruk bağımlılığı ve parametrik olmayan korelasyon tanımları aşağıda verilmiştir.

**Tanım** (Uyumluluk):  $(x_i, y_i)$  ve  $(x_j, y_j)$  sürekli rastgele değişkenler vektörü olan  $(X, Y)$ 'den alınmış gözlemler olsun. Eğer  $x_i < x_j$  ve  $y_i < y_j$  ya da  $x_i > x_j$  ve  $y_i > y_j$  ise,  $(x_i, y_i)$  ve  $(x_j, y_j)$  uyumludur denir.

**Tanım** (Kendal tau ve Spearman rho bağımlılık):  $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ , iki değişkenli bir dağılımdan alınmış  $n$  tane gözlem olsun. Bu kümede  $\binom{n}{2}$  tane farklı sayıda uyumlu ya da uyumsuz gözlem çifti vardır. Bu gözlem çiftlerinden  $c$  tanesinin uyumlu ve  $d$  tanesinin uyumsuz çift olduğunu düşünürsek Kendall Tau aşağıdaki gibi ifade edilir.

$$\tau_n = \frac{(c - d)}{\binom{n}{2}}$$

Aynı şekilde  $\tau_n$  örneklemden rasgele seçilmiş  $(x_i, y_i)$  ve  $(x_j, y_j)$  gözlem çiftleri için uyum olasılığından uyumsuzluk olasılığının çıkarılmasıyla elde edilebilir. Kendall Tau'nun popülasyon versiyonu,  $X$  ve  $Y$  rasgele değişkenleri için benzer şekilde ifade edilebilir.  $(X_1, Y_1)$  ve  $(X_2, Y_2)$ ,  $(X, Y)$  ile aynı dağılımdan gelen bağımsız rasgele vektörler olsun. Bu durumda popülasyon için Kendall Tau aşağıdaki gibi tanımlanır.

$$\tau = P\{(X_1 - X_2)(Y_1 - Y_2) > 0\} - P\{(X_1 - X_2)(Y_1 - Y_2) < 0\}$$

$X$  ve  $Y$  sürekli rasgele değişken olduğunda

$$\tau = 4 \int_0^1 \int_0^1 C_{X,Y}(u, v) dC_{X,Y}(u, v) - 1$$

olarak ifade edilebilir. Burada  $C_{X,Y}(u, v)$   $X$  ve  $Y$  rasgele değişkenlerinin kopulasıdır.

Spearman Rho da Kendall Tau gibi uyum ve uyumsuzluk olasılığı cinsinden ifade edilebilir ve kopula ilişkisi aşağıdaki gibidir.

$$\rho = 12 \iint_0^1 C_{X,Y}(u, v) du dv - 3$$

**Tanım** (Kuyruk bağımlılığı): Kuyruk bağımlılığı tek değişkenli dağılımların kuyruk bölgelerinde değer alan veriler arasındaki bağımlılık ölçüsü ile ilgilidir. Bu yüzden çok değişkenli yapılarda ekstrem değerlerin bağımlılık yapısının analizinde önemli sayılmıştır. İki değişken arasında, üst limitinin var olması şartıyla,  $z$  ekstrem eşik değeri için üst kuyruk bağımlılık katsayısı aşağıdaki şekilde tanımlanır.

$$\lambda_U = \lim_{z \rightarrow 1^-} Pr \{ F(X) > z \mid G(Y) > z \}$$

Bu eşitlik yardımıyla bir değişkende uç değer gözlenmesi olayı verildiğinde diğer değişkende de uç değer gözlenmesi koşullu olasılığının tahmini yapılabilir. Bu çalışmada, değişkenlerimiz deniz seviyesi ve basınç şiddeti alındığında üst kuyruk bağımlılığı  $\lambda_U \in (0,1]$  olacaktır. Özel halde  $\lambda_U = 0$  ise üst kuyruk bölgesinde değişkenler arası asimptotik bağımlılıktan söz edilemez.

Gözlenen verilerin kuyruk bağımlılığının grafiksel analizi Chi-Plot'a bağlı olarak yapılır [62]. İki değişken arasındaki bağımlılık yapısını tanımlamada farklı ölçütler mevcuttur. Günümüzde yaygın olarak kullanılan Pearson korelasyon katsayısı gibi global bağımlılık ölçüleri, iki veri arasındaki kompleks bağımlılık yapısını tam olarak yansıtamamaktadır. Bu nedenle iki veri arasındaki bağımlılık yapısını sadece global bağımlılık ölçüleri ile değil aynı zamanda lokal bağımlılık ölçüleri ile de incelemek gerekir. Bu açıdan lokal bağımlılık ölçüsünü değerlendirmede ve bu yapının belirlenmesinde oldukça kullanışlı, pratik bir yöntem olan Chi-Plot'u kullanacağız. Fisher ve Switzer iki veri arasındaki ilişkiyi tanımlamada Chi-Plot kavramından bahsetmiş ve tartışmışlardır [63, 64]. Chi-Plot veriler arasındaki bağımsızlık durumuna, bazı ilişki derecelerine ya da daha kompleks bağımlılık yapılarına göre karakteristik şekillere sahiptir. Chi-Plot, ham verilerin değerine bağlı değildir, fakat veri üzerindeki değerlerin derecelerine bağlıdır.

**Tanım** (Chi-Plot):  $(X, Y)$  rasgele vektörlerinden birleşik dağılım fonksiyonu  $H$  ve  $(x_i, y_i)$ ,  $i = 1, 2, \dots, n$  için bu dağılımdan gelen rastgele örneklem olsun. Her  $(x_i, y_i)$  örneklem noktaları,  $X \leq X_i$  ve  $Y \leq Y_i$  bölgelerinin arakesiti alındığında  $(X, Y)$  düzlemini çeyrek düzlemlere ayıracaktır.  $(x_i, y_i)$  nokta rastgele örneklemimizin  $i$ . elemanıdır. Öyle ki geriye  $n - 1$  tane örneklem noktaları kalır. Bu noktalar dörtlük düzlemler üzerinde dağılır. Bu noktaların dörtlük düzlemlerdeki sıklıklarını kullanarak ampirik



bileşik dağılım fonksiyonu  $H_n$  ve ampirik marjinal dağılım fonksiyonları  $F_n$  ve  $G_n$  aşağıda verildiği gibi üretilir.  $I(Z)$ ,  $Z$  olayının gösterge fonksiyonu olmak üzere her  $(x_i, y_i)$  gözlemleri ve  $i = 1, 2, \dots, n$ , için verilmiştir.

$$H(x_i, y_i) = H_{n_i} = \sum_{i=j} I(x_j \leq x_i, y_j \leq y_i) / (n - 1)$$

$$F_n(x_i) = F_{n_i} = \sum_{i \neq j} I(x_j \leq x_i) / (n - 1)$$

$$G_n(x_i) = G_{n_i} = \sum_{i \neq j} I(y_j \leq y_i) / (n - 1)$$

Eğer  $X$  ve  $Y$  rasgele değişkenleri istatistiksel olarak bağımsız ise bunların ampirik birleşik dağılım fonksiyonları  $H_n$ , ampirik marjinal dağılım fonksiyonları olan  $F_n$  ve  $G_n$  çarpımına eşit olması beklenir.

$$S_i = \text{sign}\{(F_i - 1/2)(G_i - 1/2)\}$$

$$\chi_i = (H_i - F_i G_i) / \{F_i(1 - F_i)G_i(1 - G_i)\}^{1/2}$$

olarak tanımlansın. Veri kümesinin merkezinden  $(x_i, y_i)$  noktalarına olan uzaklık ölçüsü  $\lambda_i$

$$\lambda_i = 4S_i \max\{(F_i - \frac{1}{2})^2 (G_i - \frac{1}{2})^2\}$$

alındığında Chi-Plot,  $(\lambda_i, X_i)$  çiftlerinin dağılım grafiğidir.

## 4.2 Tek Değişkenli Blok Maksimum Yöntemine Göre Uç Değer Yaklaşımları Üzerine Sonuçlar

### 4.2.1 Deniz seviyesi değişkeninin blok maksimum yöntemiyle modellenmesi

Öncelikle elimizdeki deniz seviyesi değerleri bazı günler 10 dakikalık periyotlarla, bazı günler 15 dakikalık periyotlarla ve diğer günlerde boş veriler olmasına rağmen 1 saatlik zaman farkları şeklinde sıralanmıştır. Uç değerler yöntemi ile modelleme yapabilmek için temel varsayımlardan olan durağanlığın sağlanması ve mevsimsellik etkilerinin olmaması gerekmektedir. Bu amaçla her bir saatlik zaman periyodunda dakikaların maksimum değerlerini kullanarak elde ettiğimiz veri kümesi için yapılan

testler ařađıda verilmiřtir.

Durađanlık (Stationary):

Durađanlıđın kontrolü iin Augmented Dickey-Fuller Test ve KPSS test yntemi kullanılmıřtır.

### 1. Augmented Dickey-Fuller Test.

Null hipotezi: Veri durađan deđildir, Alternatif hipotezi: Veri durađan.

řeklinde kurulmuřtur. Dickey-Fuler test istatistiđi = -23.434, Lag order = 63 ve bu test istatistiđine karřılık gelen p-deđeri = 0.01 bulunmuřtur.

Sonu: Durađanlıktan sapma yok.

### 2. KPSS Test

Null hipotezi: Veri durađan, Alternatif hipotezi: Veri durađan deđildir.

řeklinde oluřturulmuřtur. KPSS seviye = 57.335, Truncation lag parameter = 115, ve p-deđeri= 0.01 bulunmuřtur.

Sonu: Durađanlıktan sapma var.

neri: 4 ve 12 aylık farklar alınıp durađanlıđın yeniden incelenmesi.

#### A1) ncelikle yıllık farklara bakalım:

Null hipotezi: Veri durađan, Alternatif hipotezi: Veri durađan deđildir.

řeklinde oluřturulmuřtur. KPSS seviye = 0.001391, Truncation lag parameter = 115, ve p-deđeri = 0.1 bulunmuřtur. p-deđerimiz byk ıkmıřtır.

Sonu: Yıllık farklar incelendiđinde durađanlık var.

#### A2) řimdi 4 aylık (mevsimsel) farklara bakalım:

Null hipotezi: Veri durađan, Alternatif hipotezi: Veri durađan deđildir.

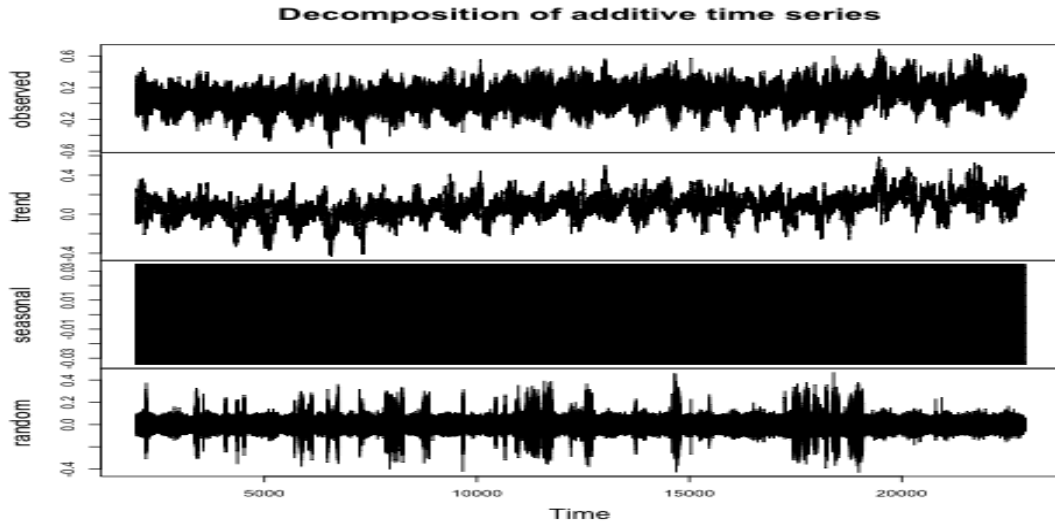
řeklinde oluřturulmuřtur. KPSS seviye = 0.00075591, Truncation lag parameter = 115 ve p-deđeri = 0.1 řeklinde elde edilmiřtir. p-deđerimiz byk ıkmıřtır.

Sonu: 4 aylık (mevsimsel) farklar incelendiđinde durađanlık var.

eřitli periyotlara gre sezonsallık etkisi:

**B1)** Yıllık bakıldıđında sezonsallık etkisi: R paket programı kullanılarak yıllık olarak periodikliđin testi yapılmıřtır.

Sonuç: Yıllık etki mevcuttur. Şekil 4.2’de periyodiklik belirtilmiştir.



**Şekil 4.2:** Deniz seviyesi değişkeninin yıllık periyodikliğinin diyagnostiği.

**B2)** 4 aylık (mevsimsel) bakıldığında sezonsallık etkisi: Yapılan testler sonucu mevsimsellik etkisi yoktur.

Buradan çıkan sonuç mevsimler arası değil yıllar arasında periyodikliğin var olduğudur. Her ne kadar 4 aylık (mevsimsel) sezonsallık etki olmasa da durağanlık ve periyodiklik yıllık olarak gözlemlenmektedir. Bu etkilerden kaçınmak için blok maksimumları yıllık alınacaktır. Ancak bu yaklaşımın olumsuz tarafı veri sayısını oldukça azaltmasıdır. Diğer yandan bu yaklaşımın olumlu tarafı ise GEV yaklaşımıyla kuyruk dağılımını modellemeye imkan sunması ve sezonsallık etkisinden varsa durağanlık probleminden kurtarmasıdır.

Ayrıca Coles’ a göre güvenilir sonuçlar alabilmek için yeteri kadar büyük veri (blok maksimumları alındıktan sonra en az 20 adet veri kalması hali) kümeleri için blok maksimumların yıllık alınması önerilmektedir [65]. Bizim bu çalışmamızda blok maksimumları alındıktan sonra veri kümesinin büyüklüğü 30 olmuştur.

Varsayımları sağlayacak şekilde blok maksimumları yıllık olarak alındığında saatlik deniz seviyesi yüksekliklerinin maksimumlarından oluşan veriler üzerinden blok maksimumlar alındığında oturtulan asimptotik dağılım 3 parametrelili GEV modelidir.

MLE kullanıldığında negatif log olabilirlik değeri (Negative Log-Likelihood Value) -29.46346’dır. Tahmin edilen konum (location), ölçek (scale) ve şekil (shape) parametre değerleri sırasıyla 0.43710418, 0.07206448 ve 0.11280306 hesaplanmıştır.

Aynı şekilde standart hata tahminleri konum, ölçek ve şekil parametreleri için sırasıyla 0.01708802, 0.01383857 ve 0.25764096 değerlerine sahiptir. Bu tahmin edilen parametreler için Kovaryans Matrisi (Covariance Matrix) aşağıdaki değeri alır.

$$A = \begin{pmatrix} 0.0002920003 & 0.0001498744 & -0.002489875 \\ 0.0001498744 & 0.0001915061 & -0.0019171926 \\ -0.002489875 & -0.0019171926 & 0.066378866 \end{pmatrix}$$

AIC = -52.92692 ve BIC = -48.72333 şeklinde hesaplanmıştır. AIC modelin kalitesi hakkında bilgi verir ve düşük olması tercih edilir. Çizelge 4.1’de deniz seviyesi için GEV parametrelerinin %95 lik asimptotik güven aralıkları verilmiştir.

**Çizelge 4.1:** Deniz seviyesi için oturtulan GEV parametrelerinin %95 lik asimptotik güven aralıkları.

%95 güven aralığı	Alt Sınır	Parametre Tahmini	Üst Sınır
Konum Parametresi	0.40361228	0.43710418	0.47059608
Ölçek Parametresi	0.04494137	0.07206448	0.09918758
Şekil Parametresi	-0.39216395	0.11280306	0.61777007

Verilerin durağanlığının sağlanması ve GEV dağılımıyla oturtulmasından sonra 2, 20 ve 100 yıllık tekrar seviyeleri sonuçları çizelge 4.2’de verilmiştir.

**Çizelge 4.2:** Deniz seviyesinin GEV dağılımına göre tekrar seviyeleri.

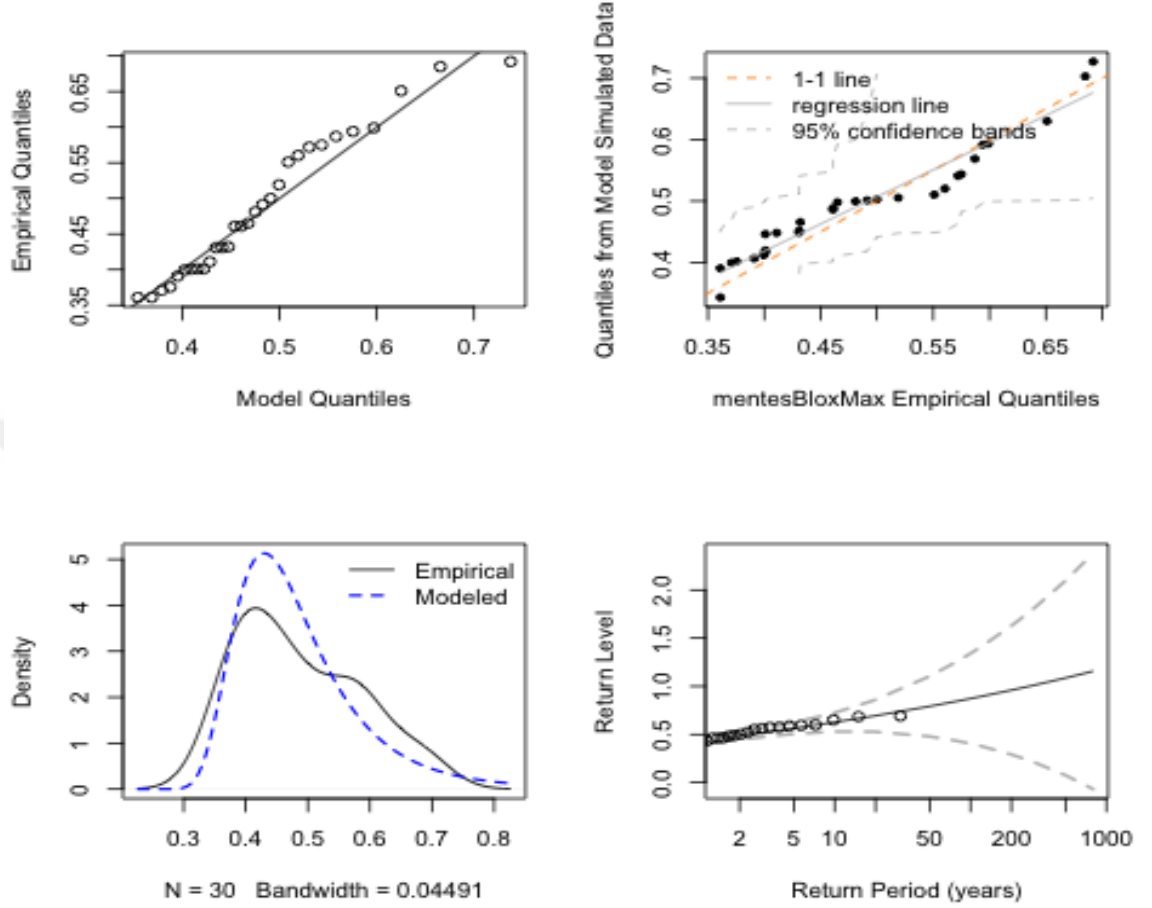
Tekrar Periyodu	2 yıllık	20 yıllık	100 yıllık
Tekrar Seviyesi	0.4640703 metre	0.6913713 metre	0.8716493 metre

Yorum: (i) deniz seviyesinin 0.46’yı geçme olasılığı %50’dir. Bir başka ifadeyle 2 sene içerisinde deniz seviyesi 0.46 m. düzeyini görecektir. (ii) deniz seviyesinin 0.69’u geçme olasılığı %5’dir. Bir başka şekilde 20 sene içerisinde deniz seviyesi 0.69 m düzeyini görecektir. (iii) deniz seviyesinin 0.87’yi geçme olasılığı %1’dir. Bir başka deyişle 100 sene içerisinde deniz seviyesi 0.87 m düzeyini görecektir.

Burada uygun olduğuna karar verilen modelin uyumluluğu şekil 4.3’deki grafik

yöntemleriyle de tasdik edilebilir.

fevd(x = mentesBloxMax)



**Şekil 4.3:** Deniz seviyesi için seçilen GEV dağılımının model uyumluluk grafikleri.

Şekil 4.3’de verilen Q-Q grafiğinde çizilen noktalar kümesinin yaklaşık olarak lineer olduğunun gözlemlenmesi, oturtulan genelleştirilmiş uç değer modelinin desteklendiğini söyler.

Bununla birlikte iki değişkenli GEV modeli çalışılırken marjinal dağılımlara ihtiyaç duyulacaktır. Dolayısıyla bunun içinde uç değer teoreminin gerektirdiği şekilde olasılık dağılımı Frechet, Weibull ya da Gumbel ailelerinden birine oturtulmaya çalışılmıştır.

Uç değer dağılımı için 0.11280306 olarak tahmin edilen şekil parametresi -0.5 ‘ten büyük olduğundan maksimum olasılık tahmin edicilerinin genel asimptotik özelliklere uygun ve dağılımın sınırlı olduğunu gösterir. Diğer taraftan şekil parametresi için yukarıda verilen güven aralığı 0’ ı içerdiğinden verilerin Gumbel dağılımına da uyabileceği söylenebilir. Böylelikle oturttuğumuz dağılımı 2

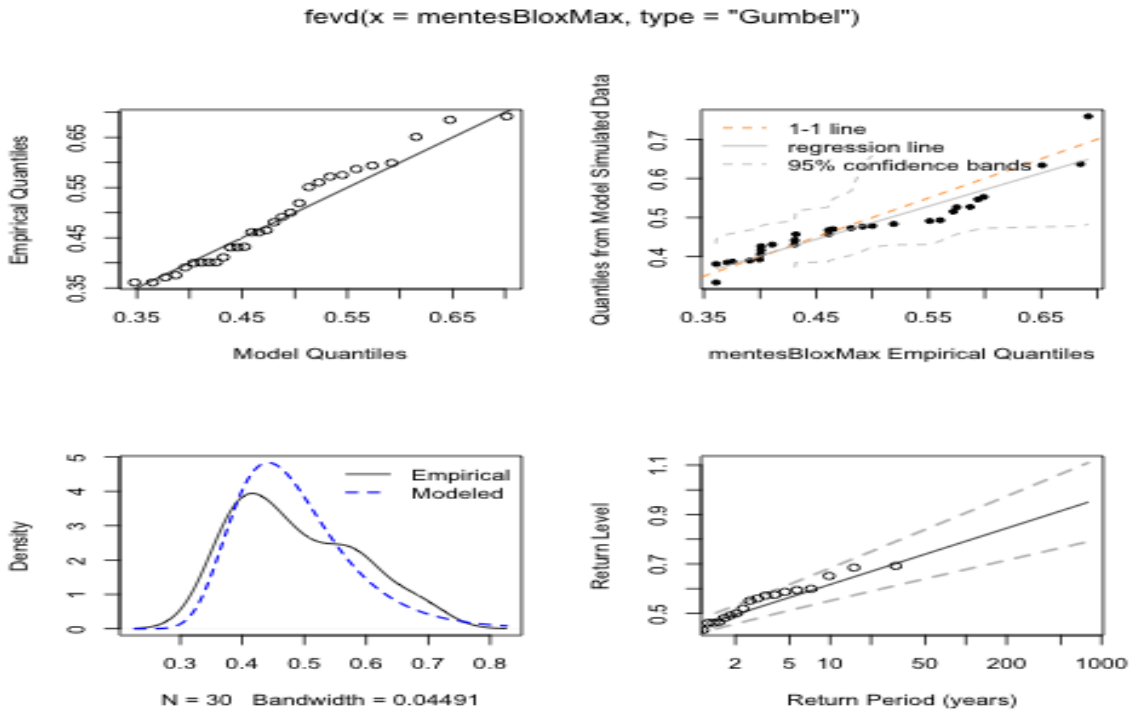
parametreyle ifade edebiliriz ki bu da ilerde iki deęişkenli GEV daęılımları oturturken kolaylık sağlayacaktır. Yine bir istatistiki kıyaslamayla 3 parametrelili GEV daęılımını ile 2 parametrelili Gumbel daęılımını arasında model oturtma bakımından anlamlı bir farklılık var mı sorusuna oran testi yaparak cevap bulacağız. Öncelikle 2 parametrelili Gumbel daęılımını oturtalım.

Maksimum olabilirlik estimatörü kullanıldığında negatif log olabilirlik deęeri - 29.36227'dir. Tahmin edilen konum ve ölçek parametre deęerleri sırasıyla 0.44160427 ve 0.07607385 olarak hesaplanmıştır. Aynı şekilde standart hata tahminleri konum ve ölçek parametreleri için sırasıyla 0.01462151 ve 0.01134305 deęerlerine sahiptir. Bu tahmin edilen parametreler için kovaryans matrisi aşığıdaki deęeri alır.

$$A = \begin{pmatrix} 2.137886 * 10^{-4} & 5.183824 * 10^{-5} \\ 5.183824 * 10^{-5} & 1.286648 * 10^{-4} \end{pmatrix}$$

AIC = -54.72455 ve BIC = -51.92215 olarak hesaplanmıştır.

Burada bahsedilen modelin uyumluluęu Őekil 4.4'deki grafik yöntemleriyle de tasdik edilebilir.



**Őekil 4.4:** Deniz seviyesi için seęilen Gumbel daęılımının model uyumluluk grafikleri.

Şekil 4.4’de verilen Q-Q grafiğinde çizilen noktalar kümesinin yaklaşık olarak lineer olduğunun gözlemlenmesi oturtulmuş genelleştirilmiş uç değer modelinin desteklendiğini söyler. Yapılan analizler sonucunda 3 parametrelili GEV ve 2 parametrelili Gumbel gibi iki farklı dağılıma verilerin uygun bir şekilde oturtulabileceği görülmüştür. Bu iki dağılımdan hangisinin daha iyi sonuç verdiğini test etmek için olabilirlik oran testi kullanılmıştır.

$D=2(-29.46346-(-29.36227)) = 0.20237$  olabilirlik oran test istatistiği ve 0.6528 p-değeri ile bulunan bu iki dağılımın birbirlerine üstünlükleri olmadığı ve dolayısıyla verilerin oturtulmuş 2 parametrelili Gumbel dağılımına indirgenebileceği sonucuna varabiliriz. Buradaki sonuçları AIC yardımıyla da ölçebiliriz. AIC bir veri kümesi için elde edilen istatistiksel modellerin göreceli kalitesini ölçer. AIC değeri düşük olan model tercih edilebilecek model olduğundan Gumbel dağılımı için ölçülen -54.724 değeri GEV dağılımı için ölçülen -52.92692 değerinden küçük olduğundan Gumbel dağılımının tercih edilebileceği bir kez daha belirtilmiştir.

Yukarıda bahsedilen oran testinin sonuçları: Olabilirlik oran testi (Likelihood-ratio test) = 0.20237, Ki-kare kritik değeri (chi-square critical value) = 3.8415, alfa = 0.0500, serbestlik derecesi (Degrees of Freedom) = 1.0000 ve p değeri= 0.6528 olarak hesaplanmıştır.

Sonuç: İki model arasında anlamlı bir fark yoktur.

Sonuç olarak deniz seviyesi değişkeni için 2 parametrelili Gumbel dağılımının uygun olduğuna karar verilmiştir. Bu dağılımın tahmin edilen konum parametresi 0.01462151 ve ölçek parametresi ise 0.01134305 değerini almıştır.

Saatlik maksimum deniz seviyesi verileri için durağanlık blok büyüklüğü yıllık alındığında sağlanmaktadır. Bu veriye Gumbel modeli oturtulduktan sonra 2, 20 ve 100 yıllık tekrar seviyeleri için çizelge 4.3’de sonuçlar elde edilmiştir.

**Çizelge 4.3:** Deniz seviyesi değişkeni için Gumbel dağılımına göre tekrar seviyeleri.

Tekrarlama Periyodu	2 yıllık	20 yıllık	100 yıllık
Tekrarlama Seviyesi	0.4694863 metre	0.6675585 metre	0.791555 metre

Yorum: (i) deniz seviyesinin 0.469'u geme olasılıđı %50'dir. Bir bařka deđiřle 2 sene ierisinde deniz seviyesi 0.469 m dzeyini grecektir. (ii) deniz seviyesinin 0.668'i geme olasılıđı %5'dir. Bir bařka řekilde ifade edilirse 20 sene ierisinde deniz seviyesi 0.668 m dzeyini grecektir. (iii) deniz seviyesinin 0.79'u geme olasılıđı %1'dir. Diđer bir ifadeyle 100 sene ierisinde deniz seviyesi 0.79 m dzeyini grecektir.

Gumbel dađılımla oturtulan modelde orta ve uzun dnem iin deniz seviye tahmini GEV dađılımına gre biraz daha dřk ıkmıřtır.

Buraya kadar saatlik maksimum deniz seviyeleri iin bazı testler yapıldı ve modeller ileriye srld. řimdi saatlik basın deđerleri iin aynı řekilde testler uygulanacak ve modeller ileri srlecektir.

#### **4.2.2 Basın řiddeti deđiřkeninin blok maksimum yntemiyle modellemesi**

Durađanlık: Durađanlıđın kontrol iin Augmented Dickey-Fuller Test ve KPSS test yntemi kullanılmıřtır.

##### **1. Augmented Dickey-Fuller Test**

Null hipotezi: Veri durađan deđildir, Alternatif hipotezi: Veri durađan.

řeklinde kurulmuřtur. Dickey-Fuler test istatistiđi = -28.051, Lag order = 63 ve bu test istatistiđine karřılık gelen p-deđerisi = 0.01 bulunmuřtur.

Sonu: Durađanlıktan sapma yok.

##### **2. KPSS Test**

Null hipotezi: Veri durađan, Alternatif hipotezi: Veri durađan deđildir.

řeklinde oluřturulmuřtur. KPSS seviye = 0.99486, Truncation lag parameter = 116, ve p-deđerisi= 0.01 bulunmuřtur.

Sonu: Durađanlıktan sapma var.

nerisi: 4 ve 12 aylık farklar alınıp durađanlıđın yeniden incelenmesi.

##### **A1) ncelikle yıllık farklara bakalım:**

Null hipotezi: Veri durađan, Alternatif hipotezi: Veri durađan deđildir.

řeklinde oluřturulmuřtur. KPSS seviye = 0.00073091, Truncation lag parameter = 116 ve p-deđerisi=0.1 bulunmuřtur. p-deđerimiz byk ıkmıřtır.



Sonuç: Yıllık farklar incelendiğinde durağanlık var.

**A2)** Şimdi 4 aylık (mevsimsel) farklara bakalım:

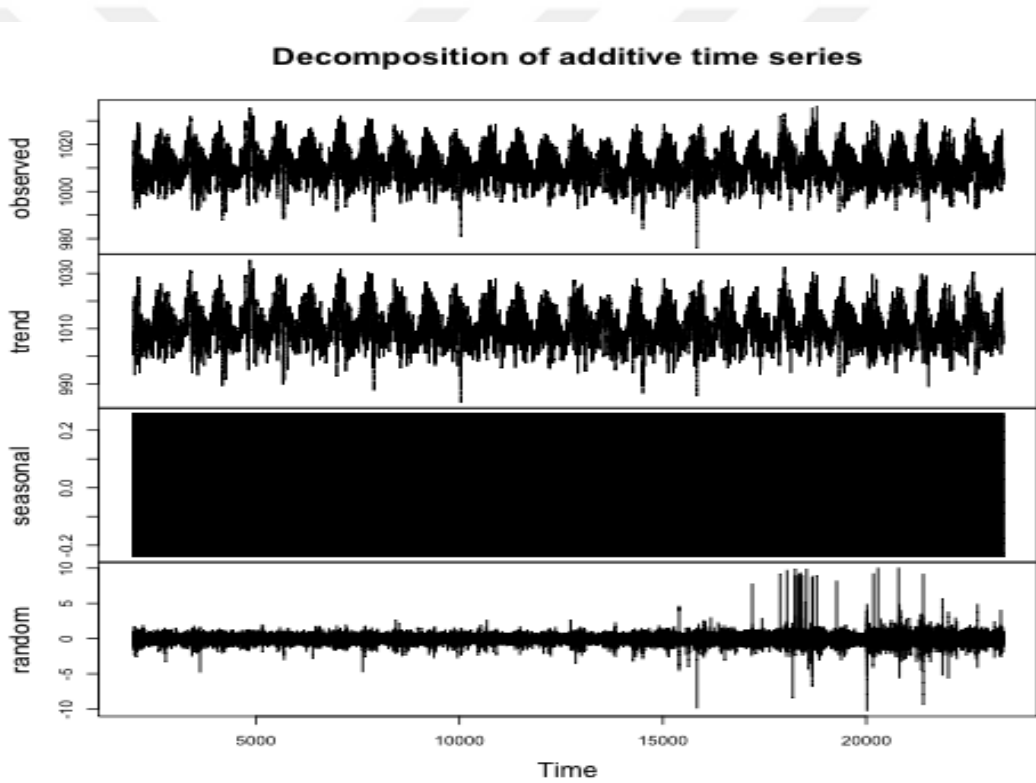
Null hipotezi: Veri durağan, Alternatif hipotezi: Veri durağan değildir.

Şeklinde oluşturulmuştur. KPSS seviye = 0.00060288, Truncation lag parameter = 116 ve p-değeri=0.1 şeklinde elde edilmiştir. p-değerimiz büyük çıkmıştır.

Sonuç: 4 aylık (mevsimsel) farklar incelendiğinde durağanlık var.

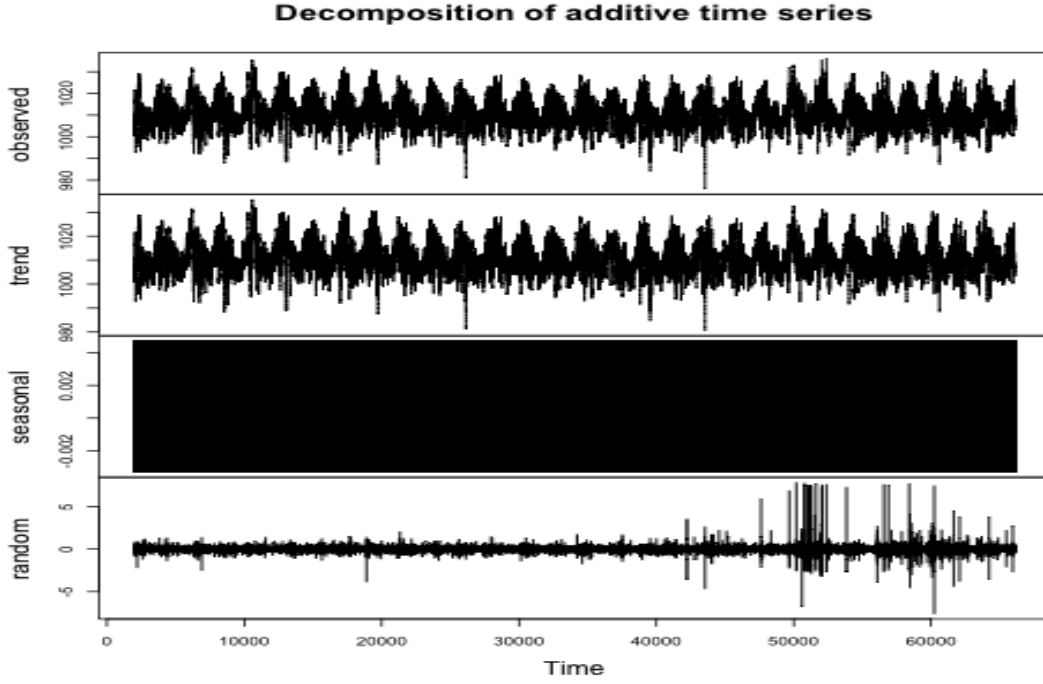
Çeşitli Periyotlara Göre Sezonallık Etkisi:

**B1)** Yıllık bakıldığında sezonallık etkisi: Yapılan testler sonucunda yıllık periyodikliğin mevcut olduğu şekil 4.5’de görülmüştür.



**Şekil 4.5:** Basınç değişkeni için yıllık periyodikliğin diyagnostiği.

**B2)** 4 aylık (mevsimsel) bakıldığında sezonallık etkisi: Yapılan test sonuçlarına göre mevsimsellik etkisi şekil 4.6’da resmedilmiştir.



**Şekil 4.6:** Basınç değişkeni mevsimsellik etkisi diyagnostiği.

Mevsimsellik etkinin ve yıllık periodikliğin var olduğu şekil 4.5 ve şekil 4.6'daki grafikler aracılığı ile desteklenmektedir. Bu etkiden kaçınmak için blok maksimumları yıllık veya mevsimsel olarak alınabilir. Ancak deniz seviyesi değişkenimizi yıllık aldığımızdan dolayı yine blok maksimumları yıllık olarak alınacaktır.

Varsayımları sağlayacak şekilde blok maksimumları yıllık olarak alındığında: Maksimum olabilirlik estimatorü kullanıldığında negatif log olabilirlik değeri 81.63801 dir. Tahmin edilen konum, ölçek ve şekil parametre değerleri sırasıyla -993.438612, 2.9132751 ve 0.1289681 şeklinde hesaplanmıştır. Aynı şekilde standart hata tahminleri konum, ölçek ve şekil parametreleri için sırasıyla 0.5936094, 0.4487726 ve 0.1305949 değerlerine sahiptir. Bu tahmin edilen parametreler için Kovaryans Matrisi aşağıdaki değeri alır.

$$A = \begin{pmatrix} 0.35237214 & 0.130241114 & -0.022465456 \\ 0.13024114 & 0.201396854 & -0.004113823 \\ -0.022465456 & -0.004113823 & 0.017055031 \end{pmatrix}$$

AIC = 169.276 ve BIC = 173.4796 olarak hesaplanmıştır. Çizelge 4.4'de basınç şiddeti için GEV dağılım parametrelerinin %95 lik asimptotik güven aralıkları belirtilmiştir.

**Çizelge 4.4:** Basınç şiddeti için seçilen GEV dağılım parametrelerinin %95 lik asimptotik güven aralıkları.

%95 güven aralığında	Alt Sınır	Parametre Tahmini	Üst Sınır
Konum Parametresi	-994.6020651	-993.4386120	-992.2751589
Ölçek Parametresi	2.0336970	2.9132751	3.7928533
Şekil Parametresi	-0.1269933	0.1289681	0.3849294

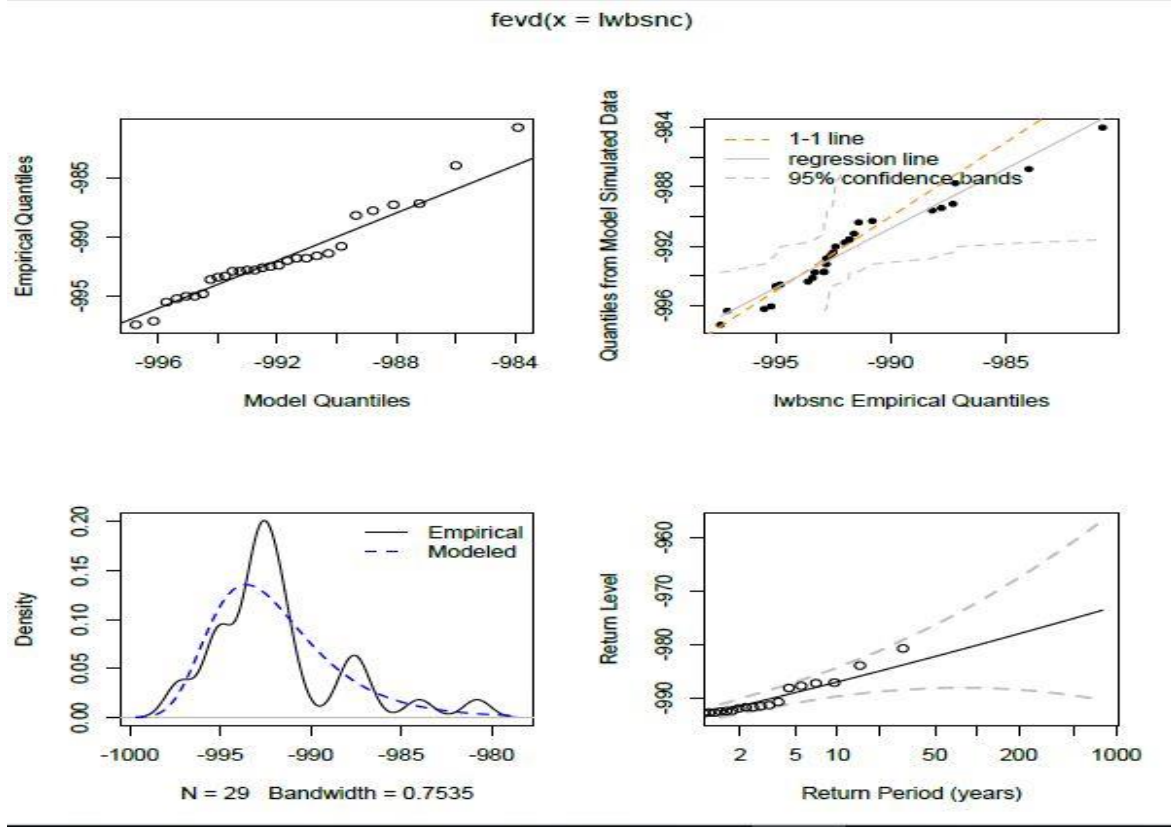
Verilerin durağanlığının sağlanmasından ve GEV dağılımına oturtulmasından sonra 2, 20 ve 100 yıllık tekrar seviyeleri için Çizelge 4.5’de bazı sonuçlar elde edilmiştir.

**Çizelge 4.5:** Basınç şiddeti için GEV dağılımına göre tekrar seviyesi.

Tekrar Periyodu	2 yıllık	20 yıllık	100 yıllık
Tekrar Seviyesi	992.3452	982.8948	975.1437

Yorum: (i) basınç seviyesinin 992.3452’ye düşme olasılığı %50’dir. Bir başka ifadeyle 2 sene içerisinde basınç seviyesi 992.3452 düzeyini görecektir. (ii) basınç seviyesinin 982.8948’e düşme olasılığı %5’dir. Bir başka şekilde 20 sene içerisinde basınç seviyesi 982.8948 düzeyini görecektir. (iii) basınç seviyesinin 975.1437’yi düşme olasılığı %1’dir. Bir başka deyişle 100 sene içerisinde basınç seviyesi 975.1437 düzeyini görecektir.

Burada uygun hale getirilen modelin uyumluluğu şekil 4.7’deki grafik yöntemleriyle de tasdik edilebilir. Şekil 4.7’de verilen Q-Q grafiğinde çizilen noktalar kümesinin yaklaşık olarak lineer olduğunun gözlemlenmesi, oturtulmuş genelleştirilmiş uç değer modelinin desteklendiğini söyler.



**Şekil 4.7:** Basınç şiddeti değişkeni için seçilen GEV dağılımının model uyumluluk grafikleri.

Bununla birlikte iki değişkenli GEV modeli çalışırken marjinal dağılımlara ihtiyaç duyulacaktır. Dolayısıyla, bunun içinde uç değer teoreminin gerektirdiği şekilde olasılık dağılımı Frechet, Weibull ya da Gumbel ailelerinden birine oturtulmaya çalışılmıştır.

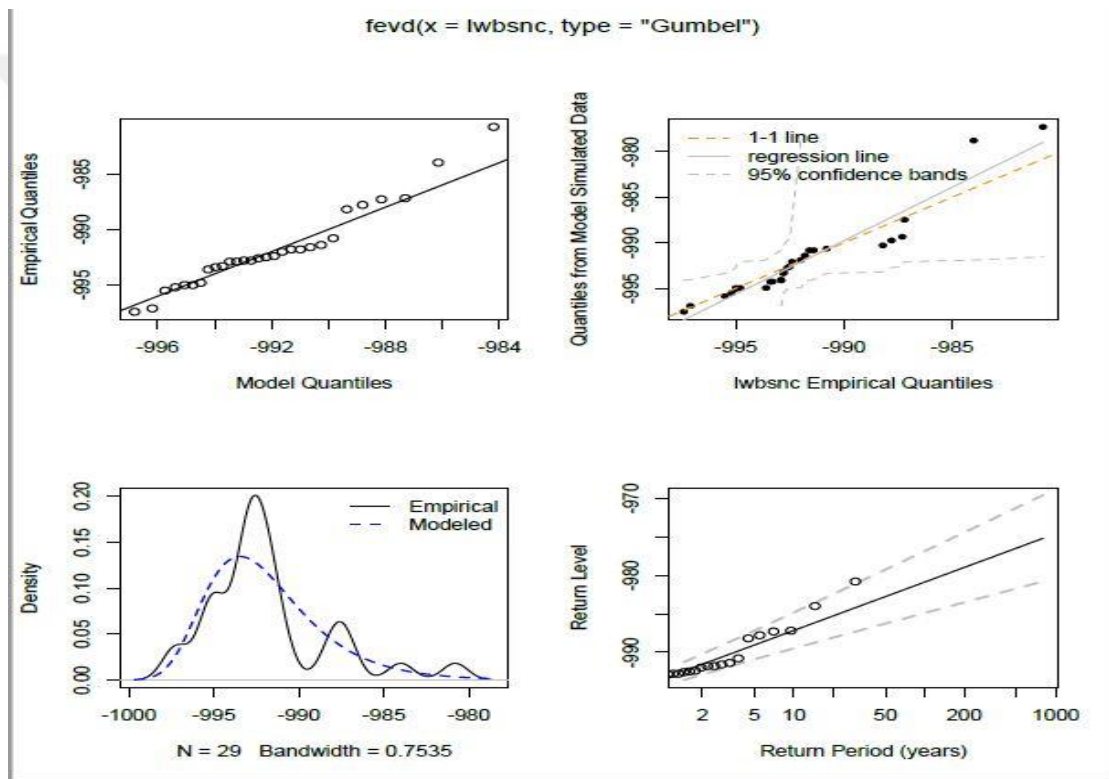
Uç değer dağılımı için burada 0.1305949 olarak tahmin edilen şekil parametresi  $-0.5$  'ten büyük olduğundan maksimum olabilirlik tahmin edicilerinin genel asimptotik özelliklere uygun olduğunu ve dağılımın sınırlı olduğunu gösterir. Diğer taraftan şekil parametresi için yukarıda verilen güven aralığı  $0$ 'ı içerdiğinden verilerin Gumbel dağılımına da oturtulabileceği söylenebilir. Böylelikle oturttuğumuz dağılımı 2 parametreye ifade edebiliriz ki bu da ilerde iki değişkenli GEV dağılımları oturturken kolaylık sağlayacaktır. Yine bir istatistiki kıyaslamayla 3 parametrelili GEV dağılımı ile 2 parametrelili Gumbel dağılımını kullanarak model oturtma arasında anlamlı bir farklılık var mı sorusuna oran testi yaparak cevap bulalım. Öncelikle 2 parametrelili Gumbel dağılımını oturtalım: Maksimum olabilirlik estimatorü kullanıldığında negatif log olabilirlik değeri = 82.23512 değerini alır. Tahmin edilen konum ve ölçek parametre değerleri sırasıyla  $-993.22950$  ve  $3.06742$

şeklinde hesaplanmıştır. Aynı şekilde standart hata tahminleri konum ve ölçek parametreleri için sırasıyla 0.5856156 ve 0.4509292 değerlerine sahiptir. Bu tahmin edilen parametreler için Kovaryans Matrisi aşağıdaki değeri alır.

$$A = \begin{pmatrix} 0.34294569 & 0.07719886 \\ 0.07719886 & 0.20333715 \end{pmatrix}$$

AIC = 168.4702 ve BIC = 171.2726 olarak hesaplanmıştır.

Burada bahsedilen modelin uyumluluğu şekil 4.8'deki grafik yöntemleriyle de tasdik edilebilir.



**Şekil 4.8:** Basınç şiddeti için seçilen Gumbel dağılımının model uyumluluk grafikleri.

Şekilde verilen Q-Q grafiğinde çizilen noktalar kümesinin yaklaşık olarak lineer olduğunun gözlemlenmesi oturtulmuş genelleştirilmiş uç değer modelinin desteklendiğini söyler.

Yapılan analizler sonucunda 3 parametrelî GEV ve 2 parametrelî Gumbel gibi iki farklı dağılıma verilerin uygun bir şekilde oturtulabileceği görülmüştür. Daha önce yapıldığı gibi bu iki dağılımdan hangisinin daha iyi sonuç verdiğini test etmek için olabilirlik oran testi kullanılmıştır.  $D=2(82.23512-81.63801) = 1.1942$  olabilirlik

oran test istatistiği ve bu değere karşı gelen 0.2745 p-değeri ile bulunan bu iki dağılımın birbirlerine üstünlükleri olmadığı ve dolayısıyla verilerin oturtulmuş 2 parametrelili Gumbel dağılımına indirgenebileceği sonucu çıkmaktadır. Buradaki sonuçları AIC yardımıyla da kontrol edebileceğimizden Gumbel dağılımı için ölçülen 168.4702 değeri GEV (Frechet) dağılımı için ölçülen 169.276 değerinden küçük olduğundan Gumbel dağılımının tercih edilebileceği bir kez daha belirtilmiştir.

Yukarıda bahsedilen oran testinin sonuçları: Olabilirlik oran testi = 1.1942, Ki-kare kritik değeri = 3.8415, alfa = 0.0500, serbestlik derecesi= 1.0000 ve p-değeri= 0.2745 olarak hesaplanmıştır.

Sonuç: İki model arasında anlamlı bir fark yoktur.

Sonuç olarak basınç düzeyi değişkeni için 2 parametrelili Gumbel dağılımının uygun olduğuna karar verilmiştir. Bu dağılımın tahmin edilen konum parametresi - 993.22950 ve ölçek parametresi ise 3.06742 değerini almıştır.

Saatlik maksimum basınç seviyesi verileri için durağanlık blok büyüklüğü yıllık alındığında sağlanmaktadır. Bu veriye Gumbel modeli oturtulduktan sonra 2, 20 ve 100 yıllık tekrar seviyeleri için çizelge 4.6’da sonuçlar elde edilmiştir.

**Çizelge 4.6:** Basınç şiddeti için Gumbel dağılımına göre tekrar seviyesi.

Tekrarlama Periyodu	2 yıllık	20 yıllık	100 yıllık
Tekrarlama Seviyesi	992.1053	984.1187	979.1189

Yorum: (i) basınç seviyesinin 992.1053’e düşme olasılığı %50’dir. Bir başka deyişle 2 sene içerisinde basınç seviyesi 992.1053 düzeyini görecektir. (ii) basınç seviyesinin 984.1187’ye düşme olasılığı %5’dir. Bir başka şekilde ifade edilirse 20 sene içerisinde basınç seviyesi 984.1187 düzeyini görecektir. (iii) basınç seviyesinin 979.1189’a düşme olasılığı %1’dir. Diğer bir ifadeyle 100 sene içerisinde basınç seviyesi 979.1189 düzeyini görecektir.

Geleceğe yönelik tahmin edilen basınç düzeyi için Gumbel ve GEV dağılımlarıyla oturtulan modeller arasında önemli bir fark gözükmemektedir.

### 4.3 Tek Değişkenli Eşik Seviyesini Aşan DeğerlerYöntemine Göre Uç Değer Yaklaşımları Üzerine Sonuçlar

#### 4.3.1 Deniz seviyesi değişkeninin pot yöntemiyle modellenmesi

Çalışmamızın bu bölümünde POT yöntemiyle, optimal eşik seviyesi tespit edilecek ve bu eşik seviyesini aşan tüm değerler dikkate alınıp GPD dağılımına oturtulacak ve buna göre modellemeler ileri sürülerek tahminler gerçekleştirilecektir. Bu kısımda yapılacak olan uç değer analizinin diğer blok maksimum yönteminden temel farkı, verinin alt gruplara ayrılmadan bir bütün halinde incelenecek olması ve bu sayede daha önceden belirtmiş olduğumuz sezonsallık etkisinden arındırılacak olmasıdır.

Durağanlık (Stationary):

Eğer GPD modelimiz belirlenen bir  $u$  eşik seviyesini geçen değerler için uygunsa, bu modelimizin  $m > u$  eşitsizliğini sağlayan  $m$  eşik seviyesini geçen değerler kümesi içinde uygun olması durumunu, yani durağan olması hali bazı testlerle kontrol edilecektir. Öncelikle eşik seviyesini aşan deniz seviyesi verilerinde durağanlığın kontrolü için Augmented Dickey-Fuller Test, KPSS Test ve Box-Ljung Test yöntemi kullanılmıştır.

##### 1. Augmented Dickey-Fuller Test

Null hipotezi: Veri durağan değildir, Alternatif hipotezi: Veri durağan.

Şeklinde kurulmuştur. Dickey-Fuler test istatistiği = -4.9315, Lag order = 5 ve bu test istatistiğine karşılık gelen p-değeri = 0.01 bulunmuştur.

Sonuç: Durağanlıktan sapma yok.

##### 2. KPSS Test

Null hipotezi: Veri durağan, Alternatif hipotezi: Veri durağan değildir.

Şeklinde oluşturulmuştur. KPSS seviye = 0.40472, Truncation lag parameter = 3, ve p-değeri= 0.07512 bulunmuştur.

Sonuç: Durağanlıktan sapma yok.

##### 3. Box-Ljung Test

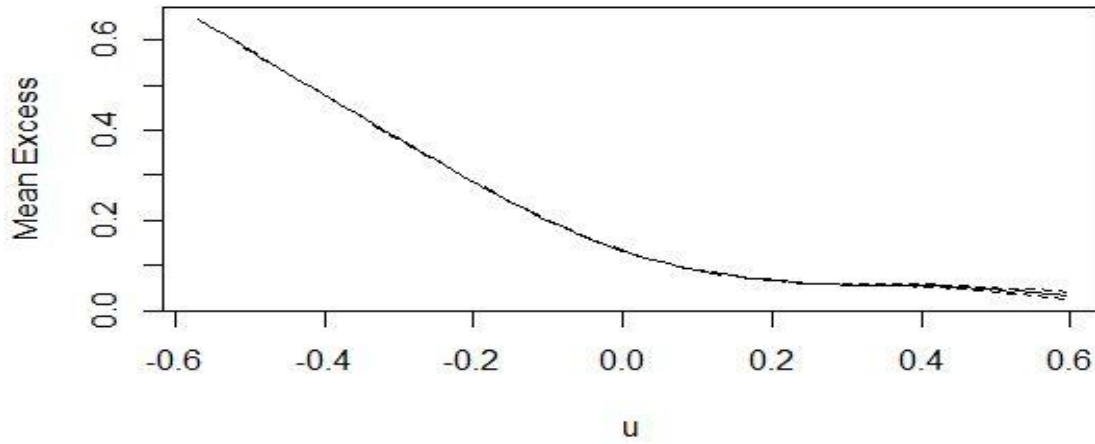
Null hipotezi: Veri durağan değildir, Alternatif hipotezi: Veri durağan.

Şeklinde kurulmuştur. Box-Ljung Ki-kare test istatistiği = 113.39, serbestlik derecesi

= 20 ve bu test istatistiğine karşılık gelen p-değeri =  $4.77 * 10^{-15}$  bulunmuştur.

Sonuç: Durağanlıktan sapma yok.

Burada durağanlık sağlandıktan sonra uygun eşik seviyesi belirlenmeye çalışılacaktır. Çok yüksek eşik değeri yüksek varyansa, çok düşük eşik değeri ise yanlış tahmin edicilere sebep olacağından uygun optimal eşik değeri şekil 4.9'daki ortalama aşım değeri grafiği (mean residual life plot) yardımıyla belirlenecektir.



**Şekil 4.9:** Deniz seviyesi için ortalama aşım değer grafiği.

Burada eşik seviyesi belirlemede kesin yöntemler olmamakla beraber lineerliğe yaklaştığı yerler önerilebilir eşik seviyelerini oluşturmaktadır. Bu çalışmada deniz seviyesi için uygun eşik seviyesi 0.371 olarak belirlenmiştir

Varsayımları sağlayacak şekilde eşik seviyesi 0.371 olarak alındığında eşik seviyesini aşan değer sayısı 2431'dir. MLE kullanıldığında negatif log olabilirlik değeri (Negative Log-Likelihood Value) -4512.472 değerini alır. Tahmin edilen ölçek (scale) ve şekil (shape) parametre değerleri sırasıyla 0.0638 ve -0.1049 şeklinde hesaplanmıştır. Aynı şekilde standart hata tahminleri ölçek ve şekil parametreleri için sırasıyla 0.00179 ve 0.0196 değerlerine sahiptir. Bu tahmin edilen parametreler için Kovaryans Matrisi (Covariance Matrix) aşağıdaki değeri alır.

$$A = \begin{pmatrix} 3.2234 * 10^{-6} & -2.7041 * 10^{-5} \\ -2.7041 * 10^{-5} & 3.8516 * 10^{-4} \end{pmatrix}$$

Verilerin durağanlığının sağlanmasının ve GPD dağılımına oturtulmasından sonra çizelge 4.7'de bazı sonuçlar 2, 20 ve 100 yıllık tekrar seviyeleri için verilmiştir.

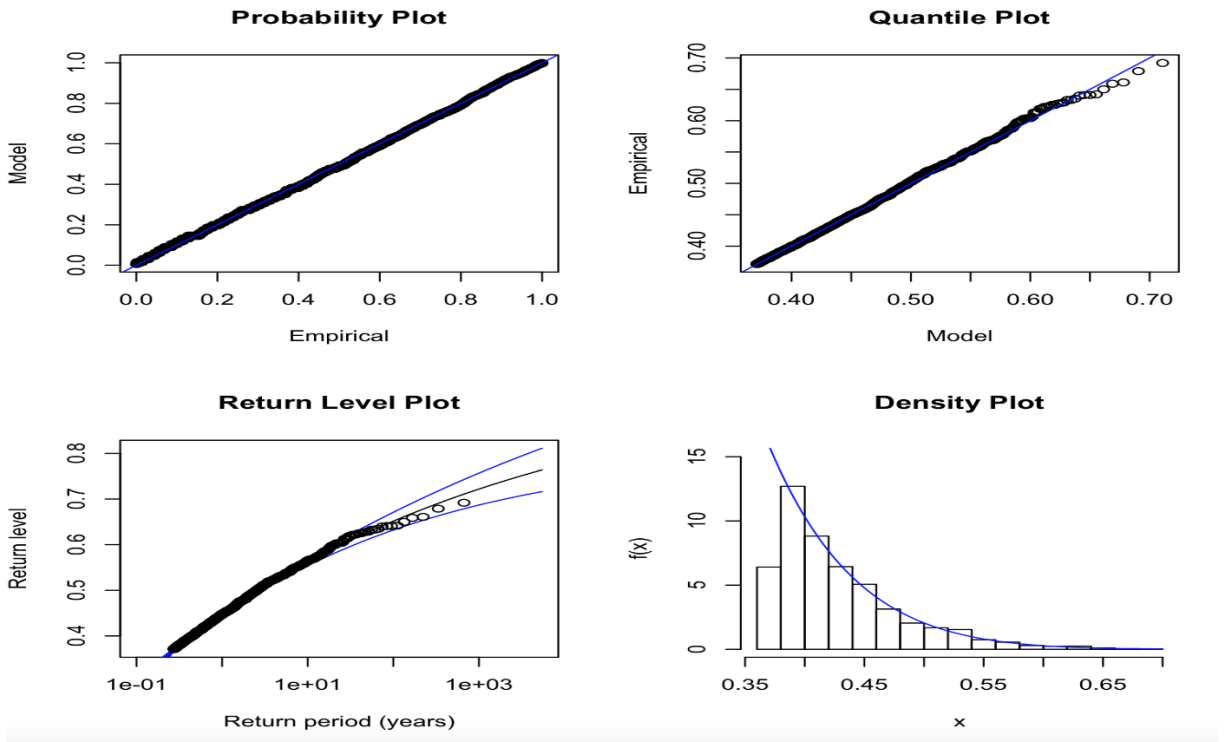


**Çizelge 4.7:** Deniz seviyesi GP dağılımına göre tekrar seviyeleri.

Tekrar Periyodu	2 yıllık	20 yıllık	100 yıllık
Tekrar Seviyesi	0.4979 metre	0.6063 metre	0.6679 metre

Yorum: Deniz seviyesi değişkeni için 2 parametrelili GPD dağılımı oturtulmuştur. Bu dağılım kullanılarak çıkarsamalar yapılabilir ve geleceğe yönelik tahminlerde bulunulabilir. Örneğin, (i) deniz seviyesinin 0.4979'u geçme olasılığı %50'dir. Bir başka ifadeyle 2 sene içerisinde deniz seviyesi 0.4979 m düzeyini görecektir. (ii) deniz seviyesinin 0.6063'ü geçme olasılığı %5'dir. Bir başka şekilde 20 sene içerisinde deniz seviyesi 0.6063m düzeyini görecektir. (iii) deniz seviyesinin 0.6679'u geçme olasılığı %1'dir. Bir başka deyişle 100 sene içerisinde deniz seviyesi 0.6679 m düzeyini görecektir.

Burada uygun hale getirilen modelin uyumluluğu şekil 4.10'daki grafik yöntemleriyle de tasdik edilebilir.



**Şekil 4.10:** Deniz seviyesi için seçilen GP dağılımının model uyumluluk grafikleri.

Şekil 4.10’da verilen Q-Q grafiğinde çizilen noktalar kümesinin yaklaşık olarak lineer olduğunun gözlemlenmesi, oturtulan genelleştirilmiş Pareto modelinin desteklendiğini söyler.

Uç değer dağılımı için burada -0.1049 olarak tahmin edilen şekil parametresi 0’dan küçük olduğundan maksimum olabilirlik tahmin edicilerinin genel asimptotik özelliklere uygun olduğunu ve dağılımın sınırlı olduğunu gösterir.

Yorum: Burada yapılan analizler ve uygulanan modeller, Menteş’te bugüne kadar toplanan veriler üzerinden tekrarlar periyotlarını vermektedir. Buna göre; deniz seviyesi 0.4979m, 0.6063m, 0.6679m yüksekliklerine sırasıyla 2, 20, 100 yıllık periyotlarla ulaşabilmektedir. Ancak Harita genel komutanlığı tarafından yapılan ölçümler neticesinde Menteş/İzmir mareograf istasyonunda deniz seviyesi yükselme trendi  $3.8 \pm 0.9$  mm/yıl olarak tespit edilmiştir. İyimser bir tahminle yıllık yükselme trendini 3 mm/yıl olarak alırsak 20 yıl sonra MSL bugünküne nazaran 6 cm daha yüksek olacaktır. Bu durumda yukarıdaki tekrarlar periyotlarında deniz seviyesinin hangi değerlere ulaşabileceğini öngörebilmek için mevcut deniz seviyelerine 6 cm (0.06m) ekleyerek aynı modelleri uyguladık. Bu model sonuçlarına göre 20 yıl sonraki MSL baz alındığında; deniz seviyesi 0.5547m, 0.6748m, 0.7528m yüksekliklerine sırasıyla 2, 20, 100 yıllık periyotlarla ulaşabilecektir. Bu da bugün 100 yılda bir görülen 0.6679m lik seviyenin 20 yılda bir görülmesi anlamına gelmektedir ki su baskınlarının daha sık gözlenebileceğini söyleyebiliriz. Çizelge 4.8’de MSL baz alınarak GP dağılımı için tekrar seviyesi sonuçları deniz seviyesi için verilmiştir.

**Çizelge 4.8:** MSL baz alındığında deniz seviyesi GP dağılımına göre tekrar seviyeleri.

Tekrar seviyesi	2 sene	10 sene	20 sene	100 sene
Deniz seviyesi	0.5547484	0.6397747	0.6748500	0.7528788

#### 4.3.2 Basınç şiddeti değişkeninin pot yöntemiyle modellenmesi

Şimdi bu yaptığımız aşamaları basınç şiddeti değişkeni için uygulayacağız.

Durağanlık (Stationary):

Eşik seviyesinin altında kalan basınç şiddeti (Minimum basınç şiddeti) verilerinde durağanlığın kontrolü için Augmented Dickey-Fuller Test, KPSS Test ve Box-Ljung Test yöntemi kullanılmıştır.

### 1. Augmented Dickey-Fuller Test

Null hipotezi: Veri durağan değildir, Alternatif hipotezi: Veri durağan.

Şeklinde kurulmuştur. Dickey-Fuler test istatistiği = -2.1393, Lag order = 1 ve bu test istatistiğine karşılık gelen p-değeri = 0.03441 bulunmuştur.

Sonuç: Durağanlıktan sapma yok.

### 2. KPSS Test

Null hipotezi: Veri durağan, Alternatif hipotezi: Veri durağan değildir.

Şeklinde oluşturulmuştur. KPSS seviye = 0.3616, Truncation lag parameter = 1, ve p-değeri= 0.09371 bulunmuştur.

Sonuç: Durağanlıktan sapma yok.

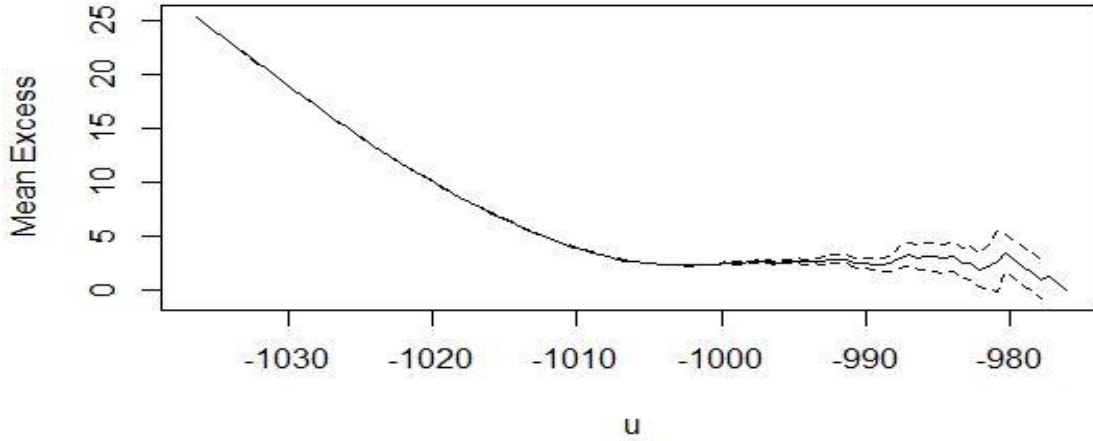
### 3. Box-Ljung Test

Null hipotezi: Veri durağan değildir, Alternatif hipotezi: Veri durağan.

Şeklinde kurulmuştur. Box-Ljung ki kare test istatistiği = 69.42, serbestlik derecesi = 20 ve bu test istatistiğine karşılık gelen p-değeri =  $2.265 \times 10^{-7}$  bulunmuştur.

Sonuç: Durağanlıktan sapma yok.

Burada durağanlık sağlandıktan sonra uygun eşik seviyesi belirlenmeye çalışılacaktır. Minimum basınç şiddeti belirlenmeye çalışıldığından negatif basınç şiddeti değerlerinin maksimumu alınarak uygun optimal eşik değeri şekil 4.11'de ortalama aşım değeri grafiği (mean residual life plot) yardımıyla belirlenmektedir ve burada verilen negatif eşik seviyesi değerlerinin pozitif olarak belirlenmesiyle basınç şiddeti için lineerliğe yaklaştığı yerler önerilebilir eşik seviyelerini oluşturmaktadır. Bu çalışmada basınç şiddeti için uygun eşik seviyesi 998.6 olarak belirlendi.



**Şekil 4.11:** Basınç şiddeti için ortalama aşım değer grafiği.

Varsayımları sağlayacak şekilde eşik seviyesi 998.6 olarak alındığında eşik seviyesinin altındaki değer sayısı 2419'dur. MLE kullanıldığında negatif log olabilirlik değeri = 4696.495 değerini alır. Tahmin edilen ölçek (scale) ve şekil (shape) parametre değerleri sırasıyla 2.47771965 ve 0.03416423 şeklinde hesaplanmıştır. Aynı şekilde standart hata tahminleri ölçek ve şekil parametreleri için sırasıyla 0.07284554 ve 0.02124732 değerlerine sahiptir. Bu tahmin edilen parametreler için kovaryans matrisi aşağıdaki değeri alır:

$$A = \begin{pmatrix} 0.005306473 & -0.0010823919 \\ -0.0010823919 & 0.0004514487 \end{pmatrix}$$

Verilerin durağanlığının sağlanmasının ve GPD dağılımına oturtulmasından sonra çizelge 4.9'da bazı sonuçlar 2, 20 ve 100 yıllık tekrar seviyeleri için verilmiştir.

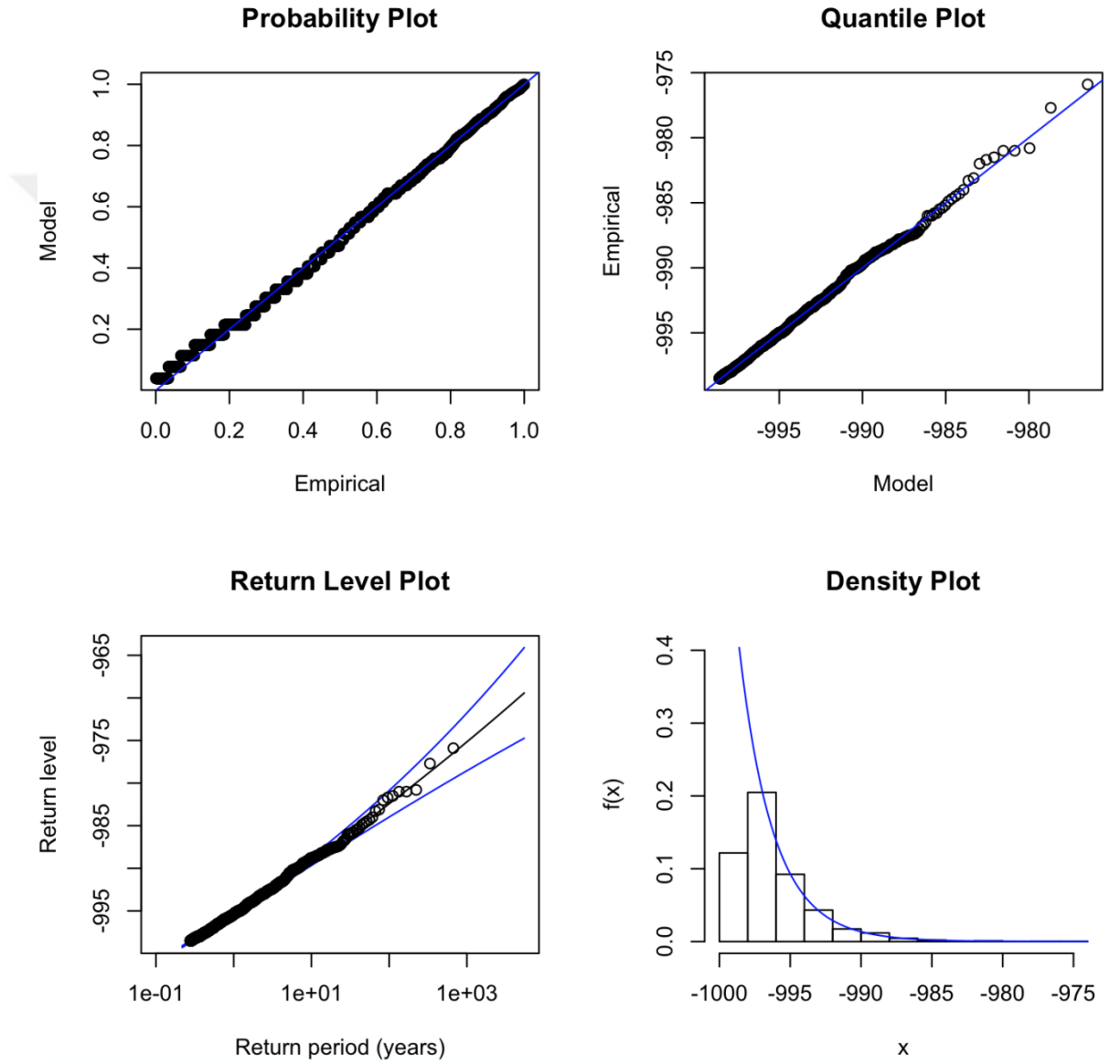
**Çizelge 4.9:** Basınç şiddeti GP dağılımına göre tekrar seviyeleri.

Tekrar Periyodu	2 yıllık	20 yıllık	100 yıllık
Tekrar Seviyesi	993.5649	987.3306	982.7581

Yorum: Basınç şiddeti değişkeni için 2 parametrelili GPD dağılımı oturtulmuştur. Bu dağılım kullanılarak çıkarsamalar yapılabilir ve geleceğe yönelik tahminlerde bulunulabilir. Örneğin, (i) basınç şiddetinin 993.5649'a düşme olasılığı %50'dir. Bir başka ifadeyle 2 sene içerisinde basınç şiddeti 993.5649 düzeyini görecektir. (ii) basınç şiddetinin 987.3306'ya düşme olasılığı %5'dir. Bir başka şekilde 20 sene

içerisinde basınç şiddeti 987.3306 düzeyini görecektir. (iii) basınç şiddetinin 982.7581'e düşme olasılığı %1'dir. Bir başka deyişle 100 sene içerisinde basınç şiddeti 982.7581 düzeyini görecektir.

Burada uygun hale getirilen modelin uyumluluğu şekil 4.12'deki grafik yöntemleriyle de tasdik edilebilir.



**Şekil 4.12:** Basınç şiddeti için seçilen GP dağılımının model uyumluluk grafikleri.

Şekil 4.12'de verilen Q-Q grafiğinde çizilen noktalar kümesinin yaklaşık olarak lineer olduğunun gözlemlenmesi, oturtulmuş genelleştirilmiş Pareto modelinin desteklendiğini söyler. GPD modeli için burada 0.03416423 olarak tahmin edilen

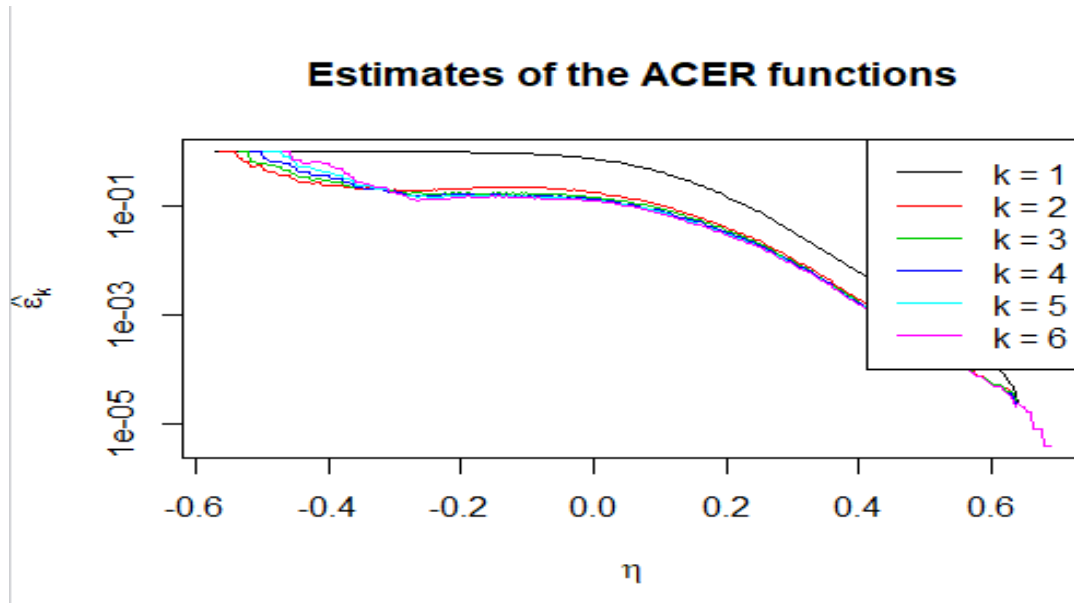
şekil parametresi 0'dan büyük olduğundan dağılımın üstten sınırlı olmadığını gösterir.

#### 4.4 Tek Değişkenli ACER Yöntemine Göre Uç Değer Yaklaşımları Üzerine Sonuçlar

Bu bölümde uç değer analizindeki son yaklaşımımız olan ACER yöntemi ile ilgili uygulamalar R programı üzerinde acer paketi kullanılarak verilecektir. Belli bir seviyeyi aşan değerlerin koşullu saptanmış (k-1) önceki değerlerinin o seviyeyi aşmaması şartıyla ACER yönteminin uygulanacağından bahsedilmiştir. Veriler için durağanlık şartı aranmadığından ötürü herhangi bir durağanlık testi uygulanmadan deniz seviyesi ve basınç şiddeti değerleri için yöntem kolayca uygulanabilecektir. ACER yöntemi verilerin bağımsız olması varsayımına gerek duymaz fakat koşullu saptanmış (k-1) önceki değer ile datanın bağımlılığı temsil edilebilir. Yani k=1 durumu datanın bağımsızlığını temsil edecektir. Tabi k=2 durumu bir adım şartlı yakınsama, k=3 durumu da iki adım şartlı yakınsama olarak ifade edilir.

##### 4.4.1 Deniz seviyesi değişkeninin acer yöntemiyle modellenmesi

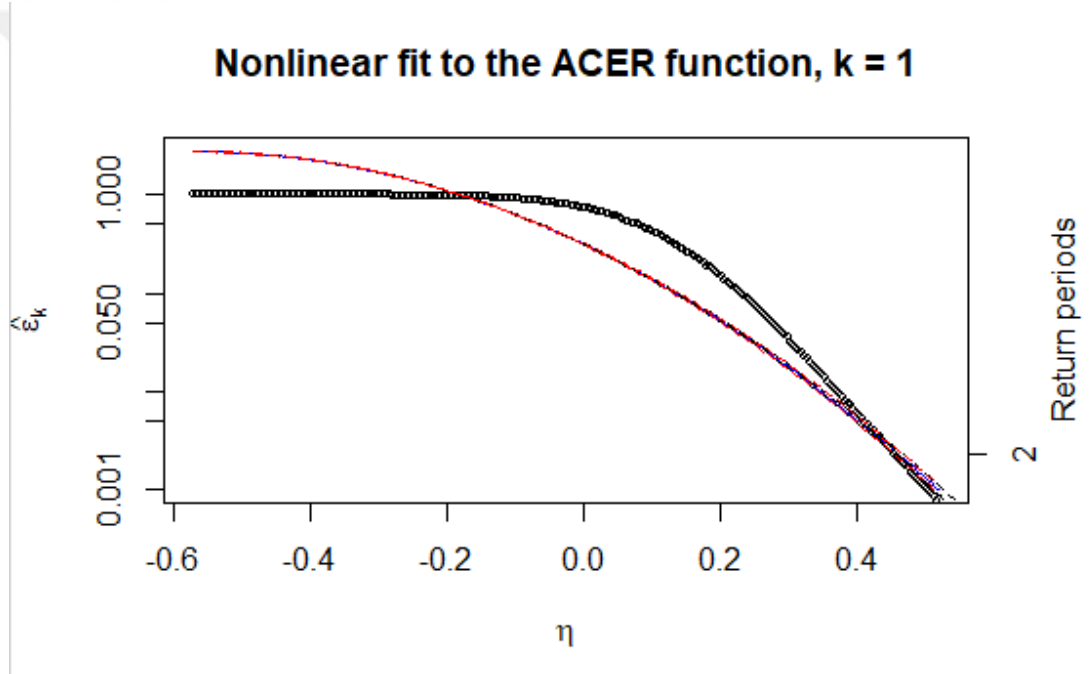
Saatlik deniz seviyesi değişkeni için farklı k tam sayılarına karşılık gelen ACER fonksiyonunun gözlemsel tahmini olan ( $\hat{\epsilon}_k$ ) değerlerinin  $\eta$  eşik seviyeleriyle birlikte gösterimi şekil 4.13'deki gibidir.



Şekil 4.13: Deniz seviyesi için farklı k değerlerine karşılık ACER fonksiyonu Tahminleri.

Şekil 4.13’de de görüldüğü gibi  $k=1$  ve  $k \geq 2$  için tahmin edilen ACER fonksiyonları ciddi farklılık göstermektedir.  $k \geq 2$  durumu için tahmin edilen ACER fonksiyonları ise birbirine yakın olduğundan  $k=2$  alınabilir. Denklem (2.21)’de de belirtildiği üzere ACER fonksiyonunun kuyruk yapısı formu  $\varepsilon(\eta) = -q(\eta)(1 + a(\eta - b)^c)^{-1/\xi}$  şeklinde verilmişti. Öncelikle  $k=1$  için R programı kullanılarak elde edilen sonuçlara göre oturtulan ACER fonksiyonunun parametreleri olan  $q, a, b$  ve  $c$  değerleri sırasıyla 2.7649, 6.6681, -0.5690 ve 2 şeklinde hesaplanmıştır.  $k=1$  için 2, 20 ve 100 yıllık periyotlarda deniz seviyesine oturtulan ACER fonksiyonları ve sonuçları şekil 4.14, şekil 4.15 ve şekil 4.16’da belirtilmiştir.

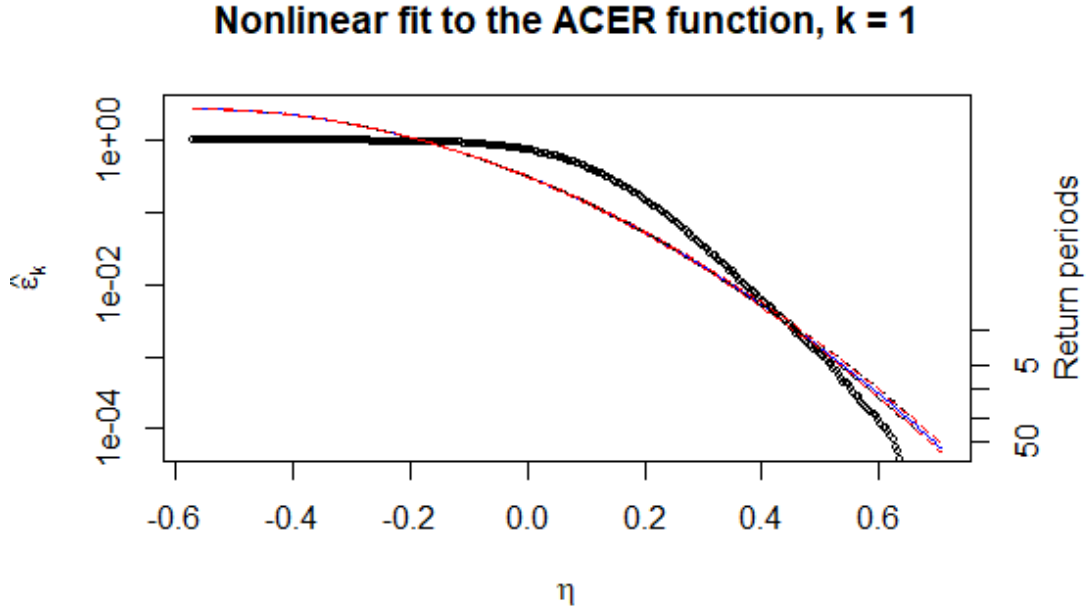
a)  $k=1$  için oturtulan ACER fonksiyonu ve 2 yıllık deniz seviyesi sonuçları



Şekil 4.14: Deniz seviyesi için  $k=1$  ve 2 yıllık periyotta oturtulan ACER fonksiyonu.

%95 güven aralığı için 2 yıllık deniz seviyesi tekrarı 0.46 m olarak öngörülmüştür.

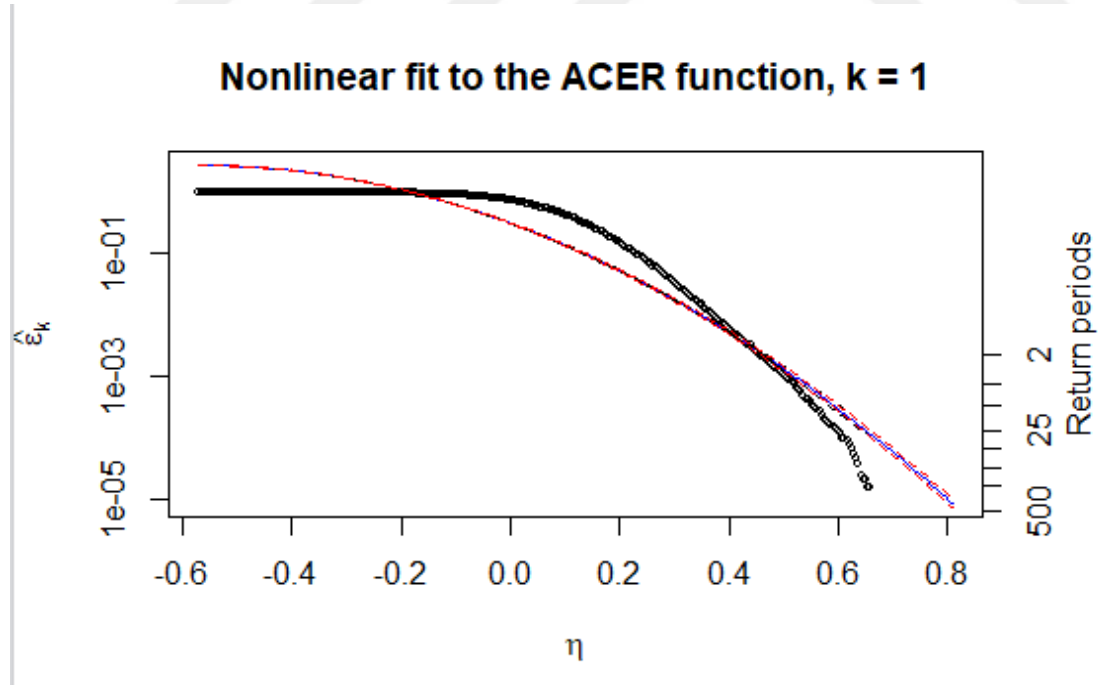
b)  $k=1$  için oturtulan ACER fonksiyonu ve 20 yıllık deniz seviyesi sonuçları



**Şekil 4.15:** Deniz seviyesi için k=1 ve 20 yıllık periyotta oturtulan ACER fonksiyonu.

%95 güven aralığı için 20 yıllık deniz seviyesi tekrarı 0.64 m olarak öngörülmüştür.

c) k=1 için oturtulan ACER fonksiyonu ve 100 yıllık deniz seviyesi sonuçları



**Şekil 4.16:** Deniz seviyesi için k=1 ve 100 yıllık periyotta oturtulan ACER fonksiyonu.



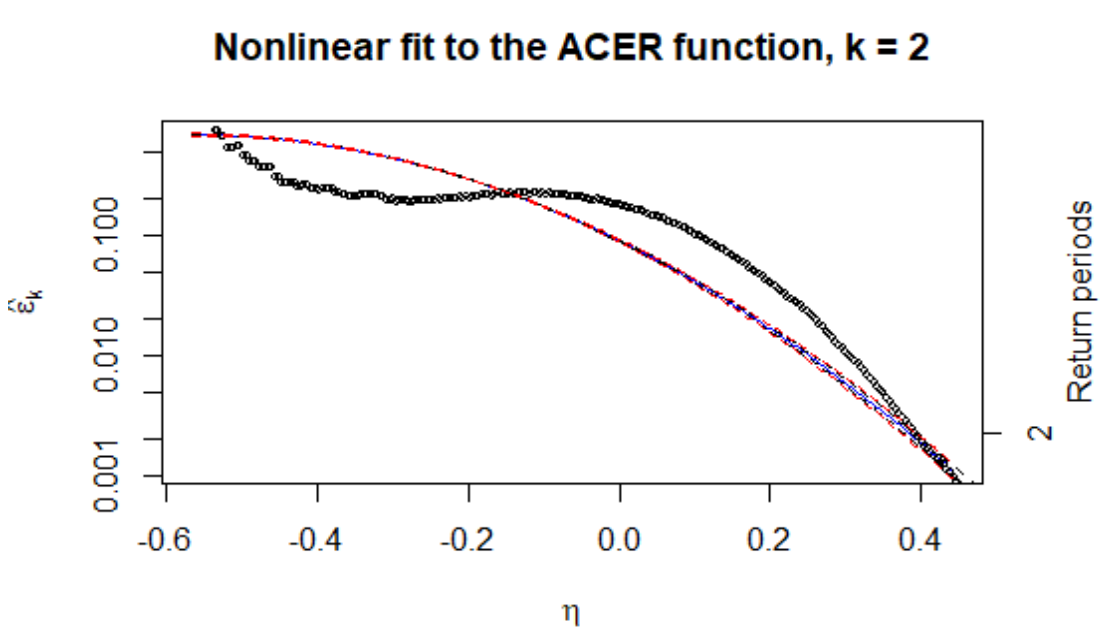
%95 güven aralığı için 100 yıllık deniz seviyesi tekrarı 0.73 m olarak öngörülmüştür. Sonuç olarak saatlik deniz seviyesi için k=1 alındığındaki ACER yöntemi sonuçlarına göre risk öngörülerini çizelge 4.10'da belirtilmiştir.

**Çizelge 4.10:** Deniz seviyesi için k=1 durumundaki ACER yöntemine göre tekrar seviyeleri.

Tekrar seviyesi	2 sene	20 sene	100 sene
Deniz	0.46	0.64	0.73

Şimdi de k=2 için R programı kullanılarak elde edilen sonuçlara göre oturtulan ACER fonksiyonunun parametreleri olan q, a, b ve c değerleri sırasıyla 0.6907, 6.3442, -0.5648 ve 2 şeklinde hesaplanmıştır. k=2 için 2, 20 ve 100 yıllık periyotlarda deniz seviyesine oturtulan ACER fonksiyonları ve sonuçları şekil 4.17, şekil 4.18 ve şekil 4.19'da belirtilmiştir.

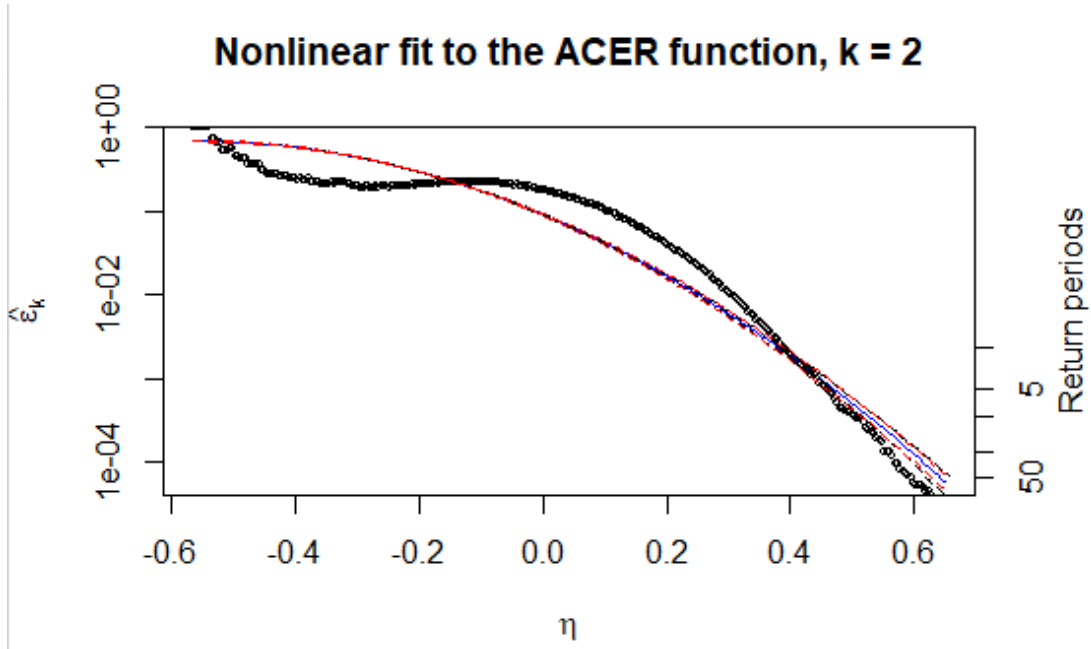
a) k=2 için oturtulan ACER fonksiyonu ve 2 yıllık deniz seviyesi sonuçları



**Şekil 4.17:** Deniz seviyesi için k=2 ve 2 yıllık periyotta oturtulan ACER fonksiyonu.

%95 güven aralığı için 2 yıllık deniz seviyesi tekrarı 0.38 m olarak öngörülmüştür.

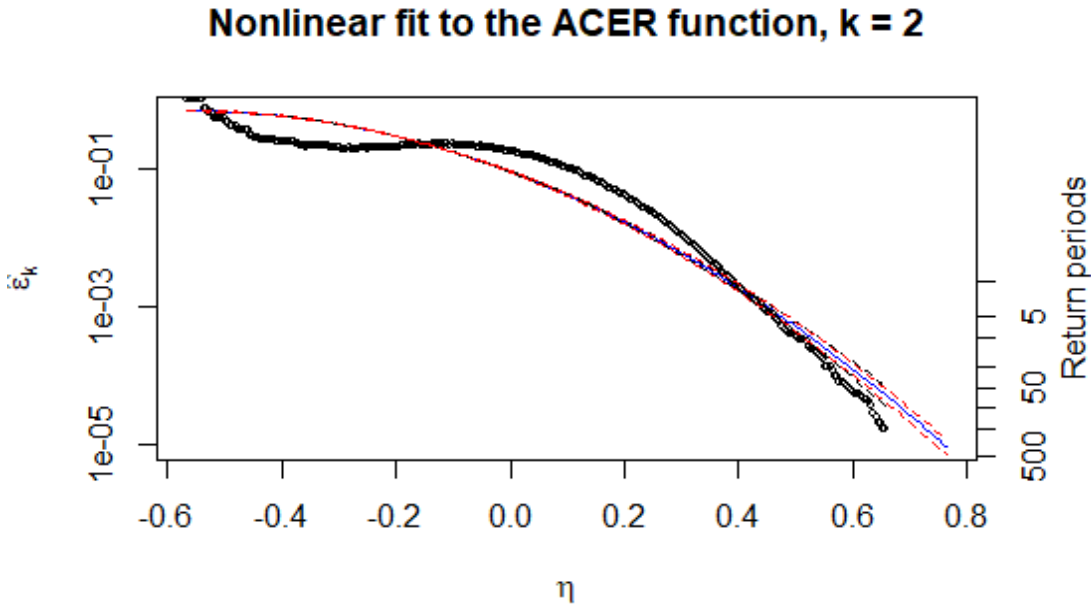
b) k=2 için oturtulan ACER fonksiyonu ve 20 yıllık deniz seviyesi sonuçları



**Şekil 4.18:** Deniz seviyesi için  $k=2$  ve 20 yıllık periyotta oturtulan ACER fonksiyonu.

%95 güven aralığı için 20 yıllık deniz seviyesi tekrarı 0.58 m olarak öngörülmüştür.

c)  $k=2$  için oturtulan ACER fonksiyonu ve 100 yıllık deniz seviyesi sonuçları



**Şekil 4.19:** Deniz seviyesi için  $k=2$  ve 100 yıllık periyotta oturtulan ACER fonksiyonu.

%95 güven aralığı için 100 yıllık deniz seviyesi tekrarı 0.69 m olarak öngörülmüştür.

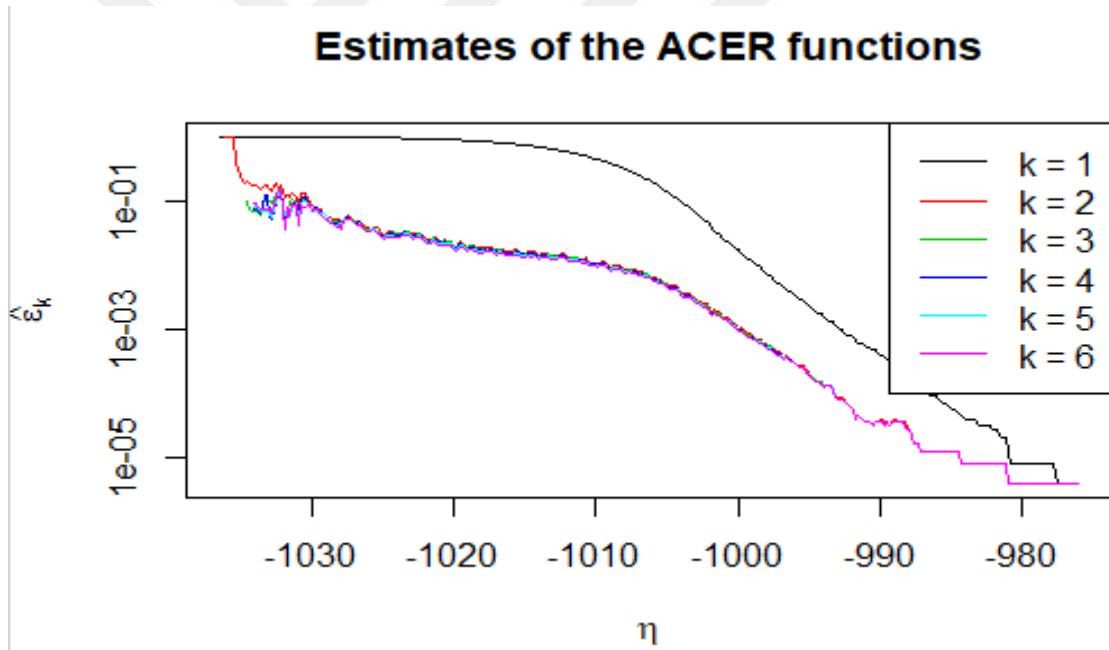
Sonuç olarak saatlik deniz seviyesi için  $k=2$  alındığındaki ACER yöntemi sonuçlarına göre risk öngörülleri aşağıdaki çizelge 4.11’de belirtilmiştir.

**Çizelge 4.11:** Deniz seviyesi için  $k=2$  durumundaki ACER yöntemine göre tekrar seviyeleri.

Tekrar seviyesi	2 sene	20 sene	100 sene
Deniz	0.38	0.58	0.69

#### 4.4.2 Basınç şiddeti değişkeninin acer yöntemiyle modellenmesi

Saatlik basınç şiddeti değişkeni için farklı  $k$  tam sayılarına karşılık gelen ACER fonksiyonunun gözlemsel tahmini olan  $(\hat{\epsilon}_k)$  değerlerinin  $\eta$  eşik seviyeleriyle birlikte gösterimi şekil 4.20’deki gibidir.

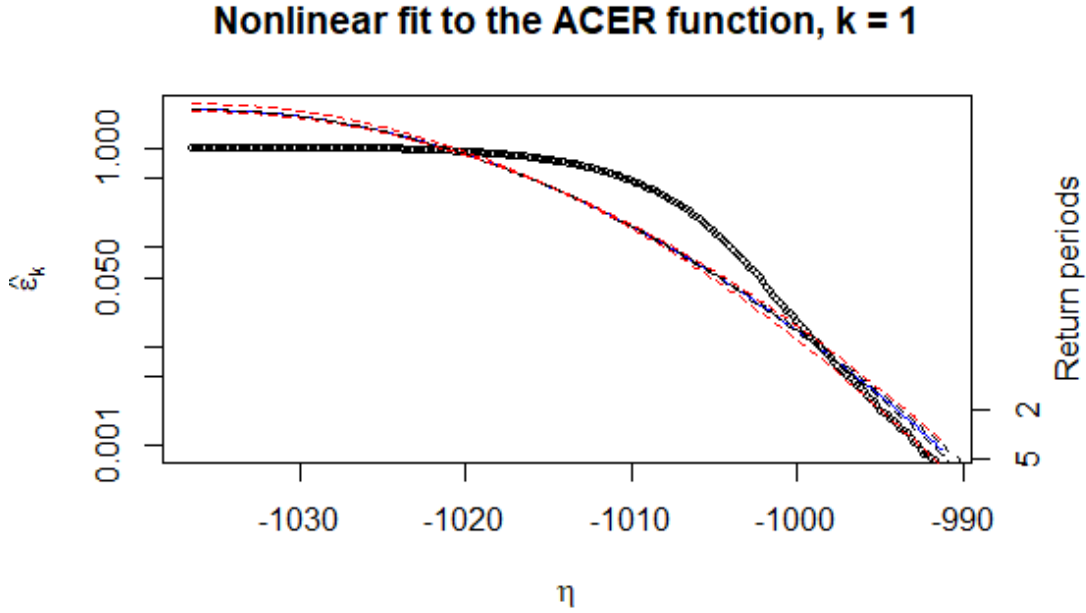


**Şekil 4.20:** Basınç şiddeti için farklı  $k$  değerlerine karşılık ACER fonksiyonu Tahminleri.

Şekil 4.20’de de görüldüğü gibi  $k=1$  ve  $k \geq 2$  için tahmin edilen ACER fonksiyonları ciddi farklılık göstermektedir.  $k \geq 2$  durumu için tahmin edilen ACER fonksiyonları ise birbirine yakın olduğundan  $k=2$  alınabilir. Öncelikle  $k=1$  için R programı kullanılarak elde edilen sonuçlara göre oturtulan ACER fonksiyonunun parametreleri olan  $q$ ,  $a$ ,  $b$  ve  $c$  değerleri sırasıyla 2.4615, 0.0039, -1036.4000 ve 2 şeklinde

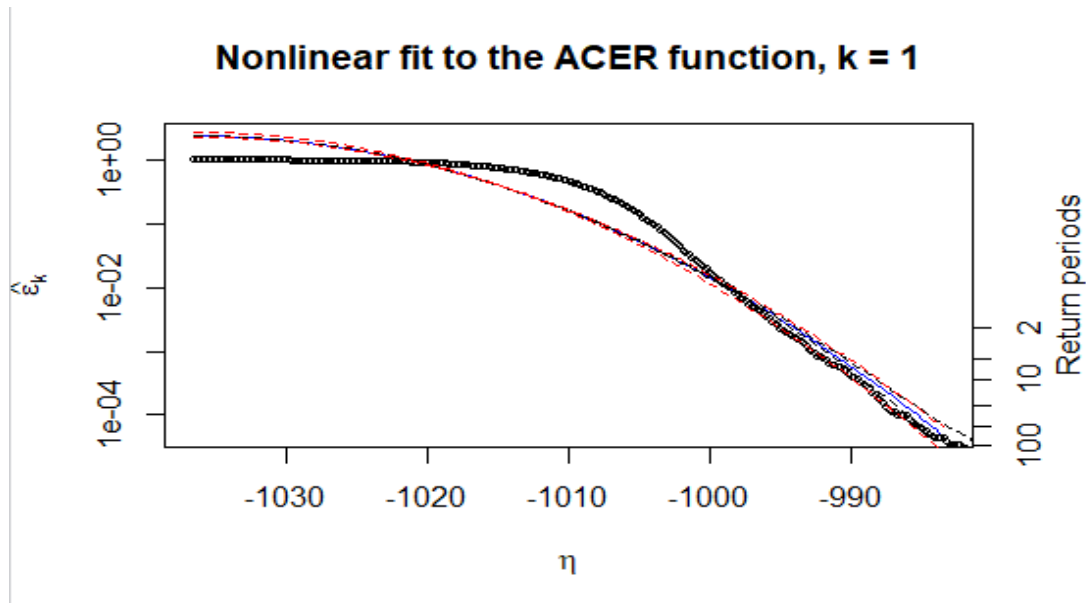
hesaplanmıştır.  $k=1$  için 2, 20 ve 100 yıllık periyotlarda basınç şiddetine oturtulan ACER fonksiyonları ve sonuçları şekil 4.21, şekil 4.22 ve şekil 4.23’de belirtilmiştir.

a)  $k=1$  için oturtulan ACER fonksiyonu ve 2 yıllık basınç şiddeti sonuçları



**Şekil 4.21:** Basınç şiddeti için  $k=1$  ve 2 yıllık periyotta oturtulan ACER fonksiyonu. %95 güven aralığı için 2 yıllık basınç şiddeti tekrarı 993.98 olarak öngörülmüştür.

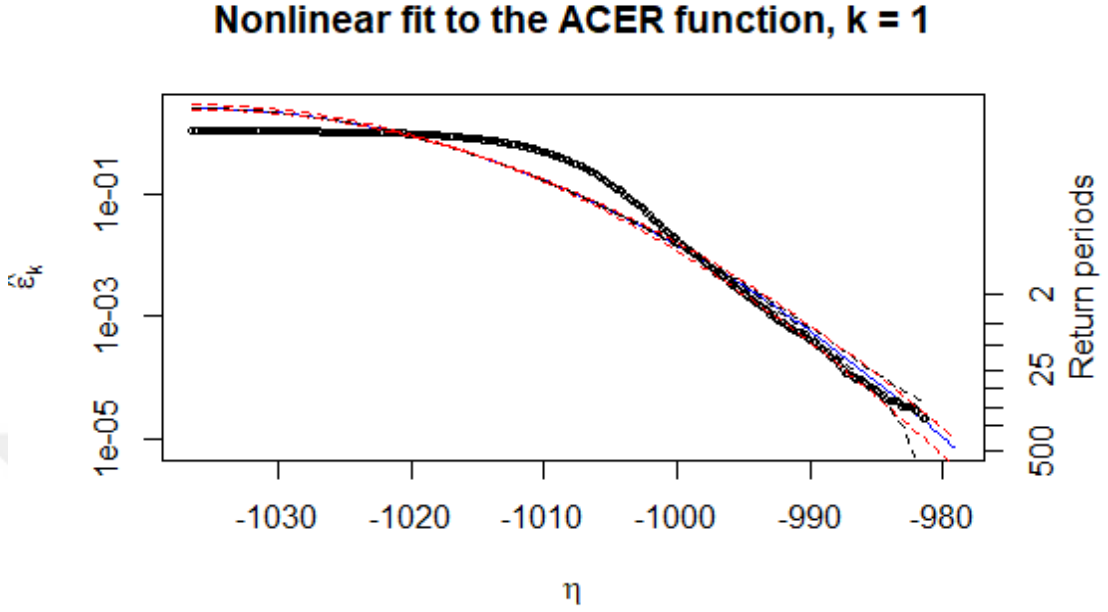
b)  $k=1$  için oturtulan ACER fonksiyonu ve 20 yıllık basınç şiddeti sonuçları



**Şekil 4.22:** Basınç şiddeti için  $k=1$  ve 20 yıllık periyotta oturtulan ACER fonksiyonu.

%95 güven aralığı için 20 yıllık basınç şiddeti tekrarı 986.69 olarak öngörülmüştür.

c) k=1 için oturtulan ACER fonksiyonu ve 100 yıllık basınç şiddeti sonuçları



**Şekil 4.23:** Basınç şiddeti için k=1 ve 100 yıllık periyotta oturtulan ACER fonksiyonu.

%95 güven aralığı için 100 yıllık basınç şiddeti tekrarı 982.62 olarak öngörülmüştür. Sonuç olarak saatlik basınç şiddeti için k=1 alındığındaki ACER yöntemi sonuçlarına göre risk öngörülürleri çizelge 4.12’de belirtilmiştir.

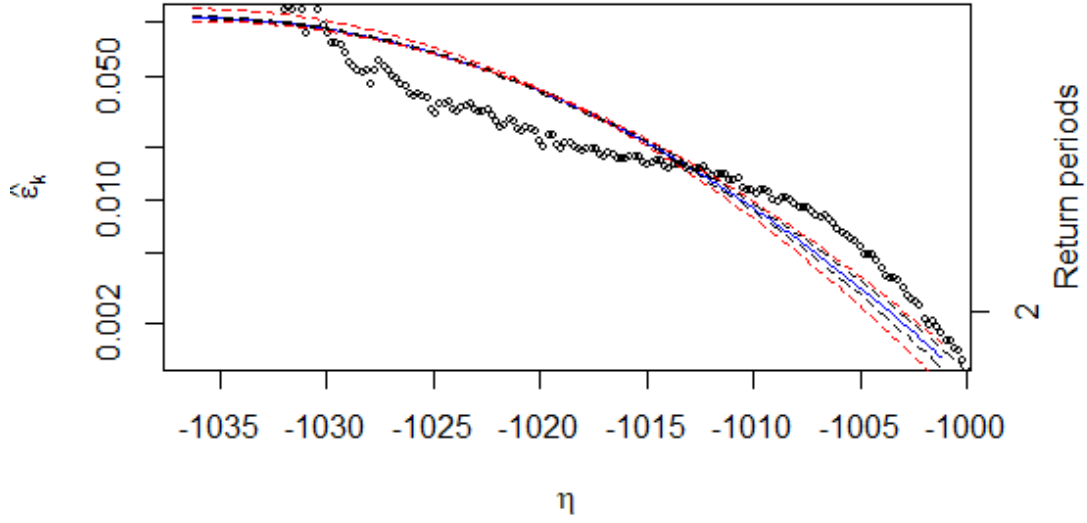
**Çizelge 4.12:** Basınç şiddeti için k=1 durumundaki ACER yöntemine göre tekrar seviyeleri.

Tekrar seviyesi	2 sene	20 sene	100 sene
Basınç	993.86	986.69	982.62

Şimdi de k=2 için R programı kullanılarak elde edilen sonuçlara göre oturtulan ACER fonksiyonunun parametreleri olan q, a, b ve c değerleri sırasıyla 0.1058, 0.0036, -1036.1977 ve 2 şeklinde hesaplanmıştır. k=2 için 2, 20 ve 100 yıllık periyotlarda basınç şiddetine oturtulan ACER fonksiyonları ve sonuçları şekil 4.24, şekil 4.25 ve şekil 4.26’da belirtilmiştir.

a) k=2 için oturtulan ACER fonksiyonu ve 2 yıllık basınç şiddeti sonuçları

### Nonlinear fit to the ACER function, k = 2

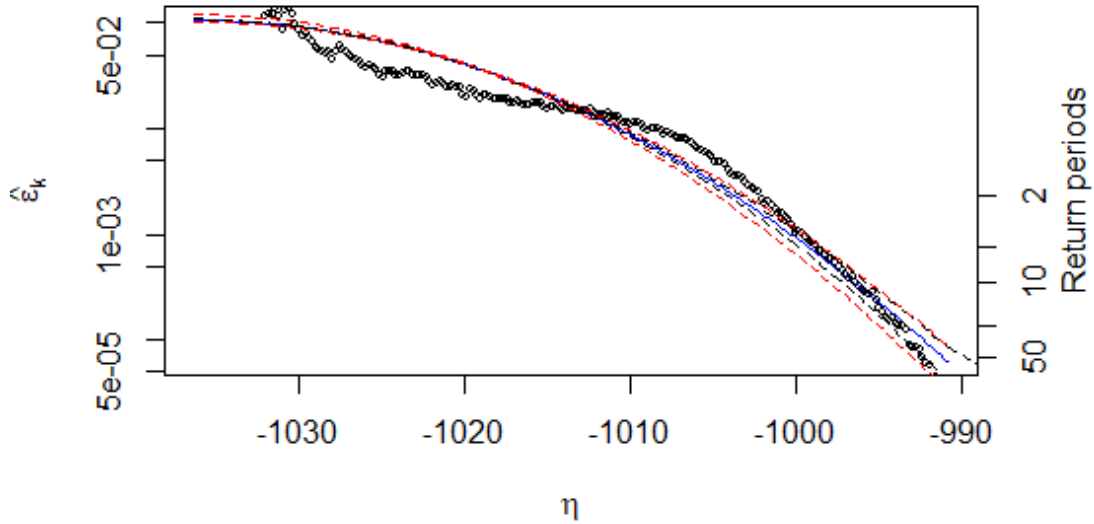


**Şekil 4.24:** Basınç şiddeti için k=2 ve 2 yıllık periyotta oturtulan ACER fonksiyonu.

%95 güven aralığı için 2 yıllık basınç şiddeti tekrarı 1003.62 olarak öngörülmüştür.

b) k=2 için oturtulan ACER fonksiyonu ve 20 yıllık basınç şiddeti sonuçları

### Nonlinear fit to the ACER function, k = 2

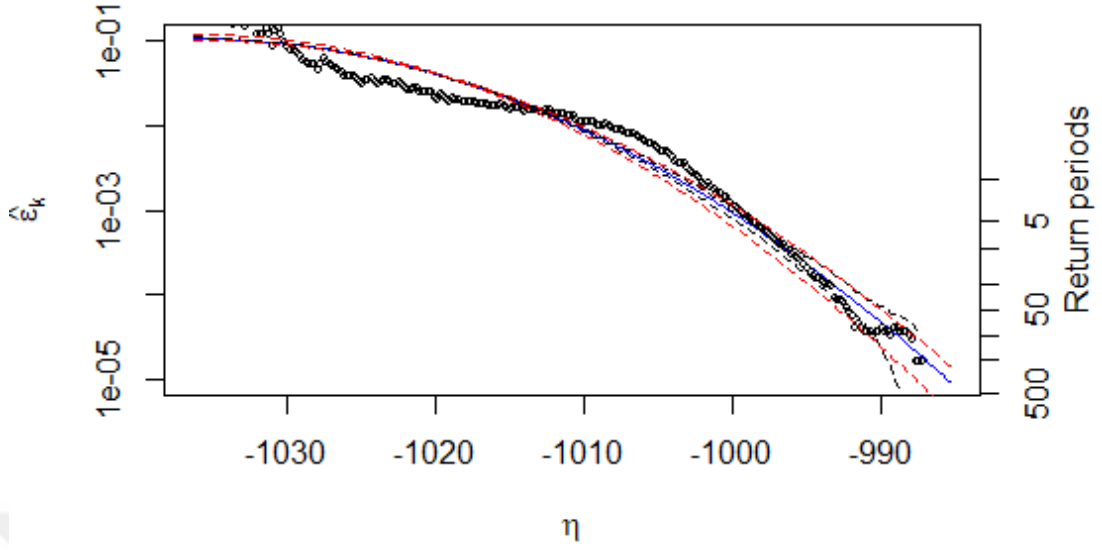


**Şekil 4.25:** Basınç şiddeti için k=2 ve 20 yıllık periyotta oturtulan ACER fonksiyonu.

%95 güven aralığı için 20 yıllık basınç şiddeti tekrarı 993.96 olarak öngörülmüştür.

c) k=2 için oturtulan ACER fonksiyonu ve 100 yıllık basınç şiddeti sonuçları

## Nonlinear fit to the ACER function, k = 2



**Şekil 4.26:** Basınç şiddeti için k=2 ve 100 yıllık periyotta oturtulan ACER fonksiyonu.

%95 güven aralığı için 100 yıllık basınç şiddeti tekrarı 982.62 olarak öngörülmüştür. Sonuç olarak saatlik basınç şiddeti için k=2 alındığındaki ACER yöntemi sonuçlarına göre risk öngörülleri çizelge 4.13’de belirtilmiştir.

**Çizelge 4.13:** Basınç şiddeti için k=2 durumundaki ACER yöntemine göre tekrar seviyeleri.

Tekrar seviyesi	2 sene	20 sene	100 sene
Basınç	1003.62	993.96	982.62

## 4.5 İki Değişkenli Uç Değer Yaklaşımları Üzerine Sonuçlar

İki değişkenli modellemede tek değişkenliden farklı olarak değişkenler arasındaki bağımlılık dikkate alınmaktadır

### 4.5.1 Blok maksimum üzerine sonuçlar

Öncelikle günlük, aylık ve yıllık olmak üzere blok seçimi çalışmaları yapılmıştır. Günlük blok seçimi değişkenler arasında bağımlılık ve durağanlık problemi doğurduğundan bu blok büyüklüğü tercih edilememiştir. Diğer taraftan blok büyüklüğü yıllık alındığında çeşitli modeller için uyum iyiliği testleri

gerçekleştirilmiş ve bu blok büyüklüğünün iyi sonuç vermediği gözlenmiştir. Blok büyüklüğü aylık olarak alındığında çeşitli modeller için uyum iyiliği testleri uygun sonuç verdiği için çalışmalar bu blok seçimi üzerinden yapılmıştır. Uyum iyiliği testi olarak Cramer–Von Mises istatistiği kullanılmıştır. Kopula ve iki değişkenli dağılım seçimleri aşağıda belirtilmiştir.

#### 4.5.1.1 Kopula yaklaşımına göre

Bazı kopula aileleri için uyum iyiliği test sonuçları Çizelge 4.16’da belirtilmiş olup aylık blok büyüklüğünün uygun olduğu hipotez testleriyle (yüksek p-değerlerinden dolayı) teyit edilmiştir. Ayrıca aylık veriler için durağanlık testi yapılmış olup, durağanlıktan sapma olmadığı KPSS istatistik testi ile onaylanmıştır. Kopula seçimi için marjinal dağılımlara ihtiyaç duyulmaktadır. Tek değişkenli modellemelerde olduğu gibi yapılan testler sonucu deniz seviyesi değişkeni için konum, ölçek ve şekil parametreleri sırasıyla 0.3000, 0.0952 ve -0.1385 olan üç parametrelili GEV dağılımı oturtulmuştur. Diğer yandan, basınç şiddeti değişkeni için konum ve ölçek parametreleri sırasıyla 1002.042 ve 4.251 olan iki parametrelili Gumbel dağılımı oturtulmuştur. Blok büyüklüğünü aylık aldığımız zaman parametrik ve parametrik olmayan korelasyon ilişkileri çizelge 4.14’de sunulmuştur.

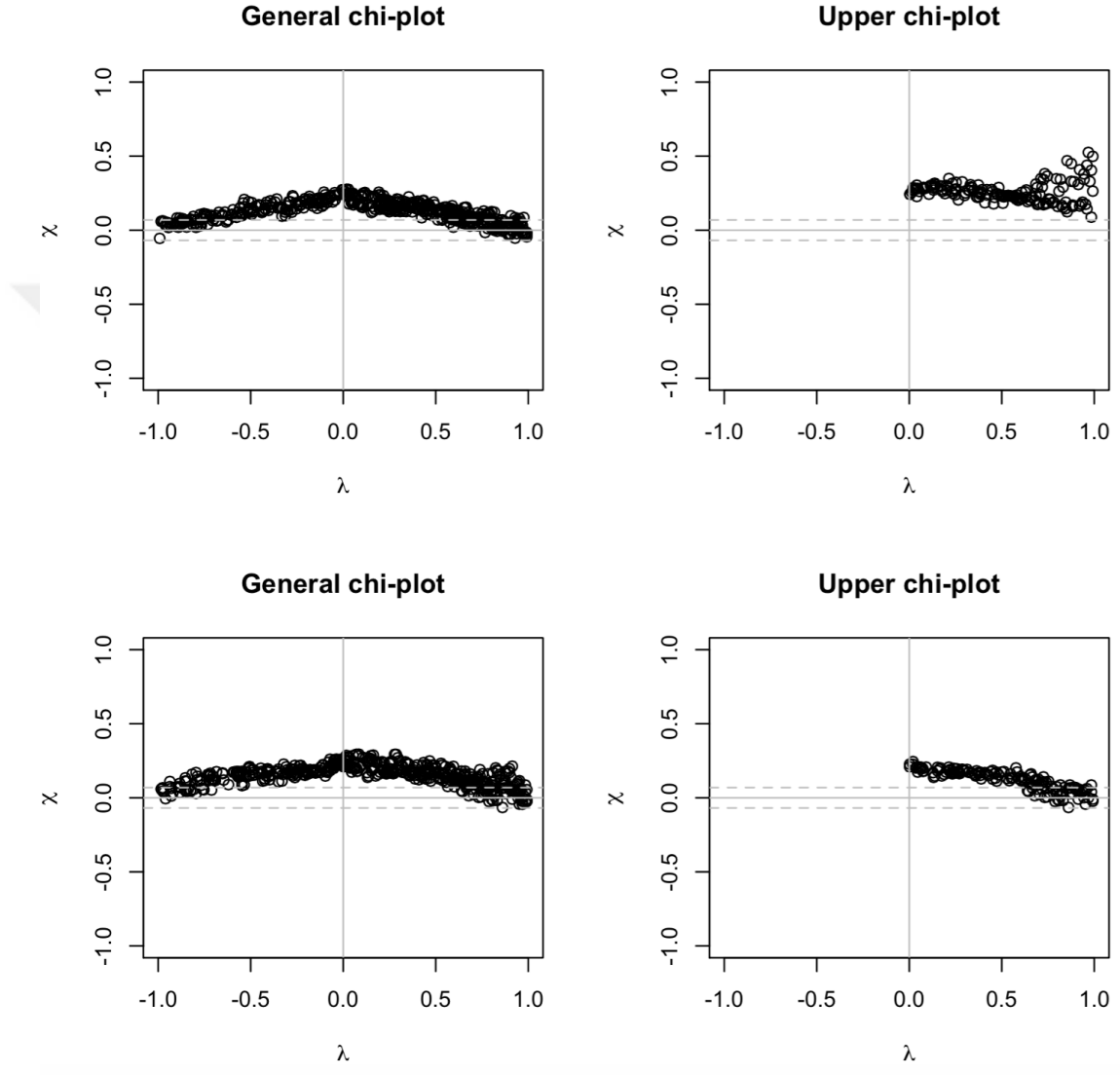
**Çizelge 4.14:** Blok büyüklüğü aylık alındığında parametrik ve parametrik olmayan korelasyon ilişkileri.

Korelasyon	Kendalls Tau	Spearman Rho	Pearson
Aylık sonuçlara göre	0.1973636	0.2901372	0.2751298

Çizelge 4.14’de aylık maksimum deniz seviyesi ve aylık minimum basınç şiddeti gözlemleri arasında düşük seviyede bir korelasyon olduğu söylenebilir (Not: aylık maksimum deniz seviyesi ve aylık maksimum basınç şiddeti arasındaki korelasyon negatiftir). Ancak bu bölümdeki modellemelerde iki değişken arasındaki bağımlılık ilişkisini Kendal Tau, Spearman Rho veya Pearson korelasyonlarından direk olarak alamayız. Deniz seviyesi ve basınç şiddeti arasındaki bağımlılık ilişkisini kopula yaklaşımıyla hesaba katarak modelleme yapılacaktır. Farklı kopula ailelerinden söz etmek mümkündür, ancak bu iki değişken arasında üst kuyruk bağımlılığı beklemek doğal olduğundan buna Chi-Plot grafikleriyle bakılmıştır.



Şekil 4.27'nin birinci sütunu pozitif bir bağımlılık olduğunu söylerken, ikinci sütunu uç değer ailesinden olan Gumbel kopula için (üstteki) pozitif bir üst kuyruk bağımlılığı ifade ederken, Arşimet ailesinden olan Frank kopula (alttaki) için bir üst kuyruk bağımlılığından söz edilememektedir.



**Şekil 4.27:** Gumbel ve Frank kopulalar için Chi-Plot ve üst kuyruk bağımlılığı.

Bazı kopulalar için üst kuyruk bağımlılık katsayısı  $\lambda_U$  kopula bağımlılık katsayısı ile ifade edilebilmektedir. Bu bilgiden hareketle bu kopulalar için üst kuyruk bağımlılık parametresi  $\theta_n$ , üst sınırları ile birlikte çizelge 4.15'deki gibi verilmiştir.

**Çizelge 4.15:** Kopulalar için üst kuyruk bağımlılık parametresi ve üst sınır.

Kopula	$\theta_n$	Üst sınır
Gumbel	1.2458940	0.2557143
Galambos	0.5016260	0.2511260
Husler-Reiss	0.8674480	0.2490000
TEV	0.5032890	0.2549893
Tawn	0.5289444	0.2644722

Çizelge 4.15’de uç değer kopula ailesi üyelerinden yukarıda belirtilen kopulalar deniz ve basınç değişkenleri arasında kuyruk dağılımının üst bölgesinde bir bağımlılık ifade etmektedir. Ancak Clayton, Frank ve Plackett kopulalar için üst kuyruk bağımlılık katsayısı sıfır olduğu bilinmektedir ki bu bilgi şekil 4.27’de Frank kopula için gösterilmiştir. Dolayısıyla uç değer kopula ailesi, deniz ve basınç değişkenlerinin arasındaki yapıyı açıklamada kullanılabilir.

Uç değer kopulaları kullanılacağı kabulünden hareketle hangi kopulanın daha iyi sonuç vereceğini uyum iyiliği testi ve AIC bilgi kriterleri yardımıyla belirlemesi amacıyla Çizelge 4.16 oluşturulmuştur. Ayrıca bu çizelgede MLE parametre tahmin metodu kullanılarak uyum iyiliği testi için elde edilen kopula parametre tahmini  $\theta_n$ , test istatistiği  $S_n$  ve hesaplanan p-değeri çizelge 4.16’da verilmiştir.

**Çizelge 4.16:** Kopulalar için parametre tahmini  $\theta_n$ , uyum iyiliği test istatistiği  $S_n$  ve AIC bilgi kriteri.

Kopula	$\theta_n$	$S_n$	p-değeri	AIC-bilgi kriteri
Gumbel	1.2052	0.0074053	0.8716	-22.88
Galambos	0.45302	0.0100320	0.6778	-23.38
Husler-Reiss	0.80081	0.0129720	0.5919	-23.42
TEV	0.88051, 24.537	0.0113390	0.6149	-21.42
Tawn	0.43291	0.0170170	0.6518	-19.16

Uyum iyiliği testleri ile elde edilen yüksek p-değerleri uç değerler kopula ailesinde yer alan dağılımların modellemede uygulanabileceğini göstermektedir. En yüksek p-değeri Gumbel kopula için elde edilmesine rağmen diğer kopulaların da istatistiksel olarak anlamlı olduğu gözükmemektedir. Gumbel, Galambos ve Husler-Reiss kopula modellerinin AIC bilgi kriteri birbirine çok yakın olduğundan ve uyum iyiliği testlerinde en yüksek p-değeri Gumbel için elde edildiğinden bu problem için en uygun kopulanın Gumbel olduğu sonucuna varılmıştır.

#### 4.5.1.2 İki değişkenli uç değer dağılımları kullanarak

İki değişkenli (çok değişkenli) dağılımlar uygun kopula seçimleriyle türetilmektedir. Bunlardan uç değer problemleri için olanları Çizelge 4.17’de verilmiş olup bizim problemimiz için uygun model seçimi belli kriterler gözetilerek yapılmıştır. Buradaki amaçlardan biri de seçilen bu model yardımıyla iki değişkenli POT yöntemini çalışabilmek ve 2, 20 ve 100 sene için tekrar seviyesi hesapları yapabilmektir. Dokuz tane iki değişkenli uç değer dağılımları için aylık değerler göz önüne alınarak yapılan hesaplar çizelge 4.17’de verilmiştir. Fakat öncesinde iç içe yuvalanmış modellerin birbirlerinin yerine kullanılabilirliğinden söz edeceğiz. Yani, LOG dağılım ile ALOG dağılımın ve aynı şekilde NEGLOG dağılım ile ANEGLOG dağılımın iç içe yuvalanmış dağılımlar olduğunu düşünürsek bunların birbirlerine indirgenebiliyor olup olmadığını testlerini yapacağız.

Null hipotezi: LOG ile ALOG arasında fark yoktur,

Alternatif hipotezi: LOG ile ALOG arasında fark vardır şeklinde oluşturulmuştur.

Ki-kare test istatistiği = 1.1533 ve Ki-kare dağılımından elde edilen p-değeri = 0.5618 şeklinde hesaplanmıştır.

Sonuç: Null hipotezi reddedilemez.

Yukarıdaki sonucu destekleyecek şekilde Yue ve Wang çalışmalarında korelasyon katsayısı  $2/3$ ’ten küçük olduğunda LOG ve ALOG modelleri arasında fark olmadığını belirtmişlerdir [60]. Benzer şekilde,

Null hipotezi: NEGLOG ile ANEGLOG aynıdır,

Alternatif hipotezi: NEGLOG ile ANEGLOG aynı değildir şeklinde oluşturulmuştur.

Olabilirlik oran testinden elde edilen p-değeri = 0.2061 şeklinde hesaplanmıştır.

Sonuç: Null hipotezi reddedilemez.

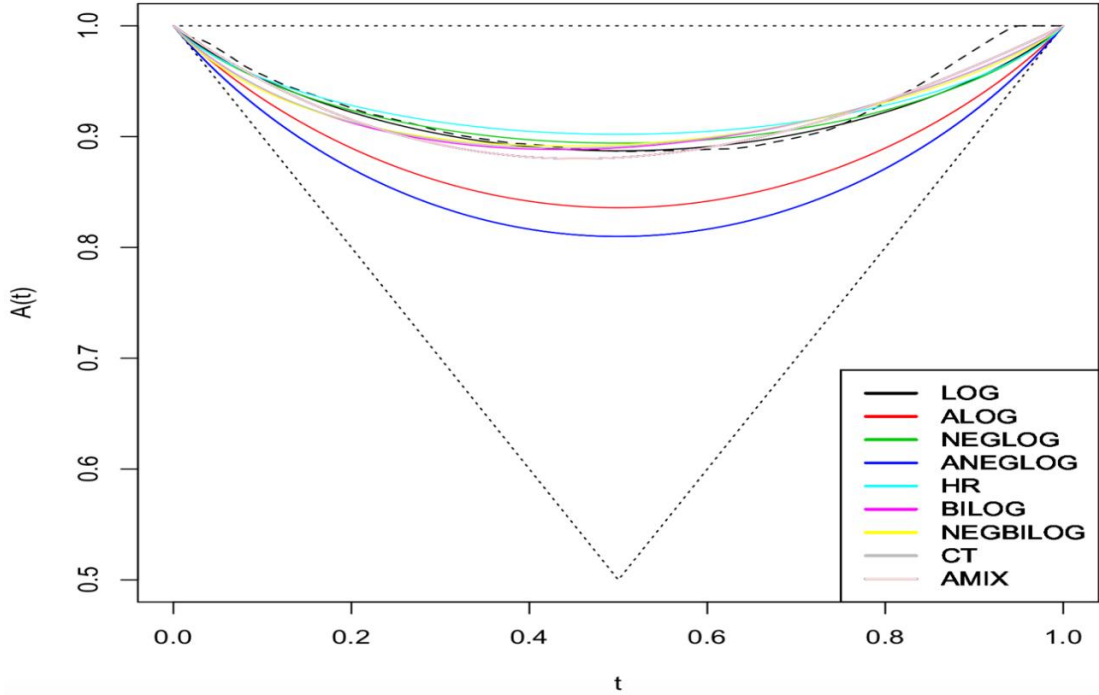
Bu testlerden LOG ve NEGLOG dağılımlarının asimetrik karşılıkları arasında istatistiksel bir fark olmadığı ortaya çıkmıştır ve çizelge 4.17’ de modellerin AIC ve bağımlılık ölçüleri verilmiştir.

**Çizelge 4.17:** İki değişkenli uç değer dağılımları ve AIC bilgi kriteri.

İki değişkenli uç değer dağılımları	-2* negatif log olabilirlik değeri	AIC-bilgi kriteri	Dependence
<b>Log</b>	1407.481	<b>1421.481</b>	0.2256801
<b>Alog</b>	1406.328	1424.328	0.2528097
Hr	1409.738	1423.738	0.1959184
Neglog	1408.739	1422.739	0.2117667
<b>Aneglog</b>	1405.58	1423.58	0.2254345
Bilog	1405.869	<b>1421.869</b>	0.219253
<b>Negbilog</b>	1406.724	1422.724	0.2094607
Ct	1405.928	<b>1421.928</b>	0.2178006
Amix	1407.088	1423.088	0.2377948

Şimdi, Çizelge 4.17’de verilen modellerden hangisinin daha iyi oturduğuna AIC ve Pickands bağımlılık ölçülerinin karşılaştırılmasıyla karar vereceğiz.

Daha önce bahsedildiği üzere bağımlılık yapılarını Pickands bağımlılık fonksiyonu ile de inceleyebiliriz. Pickands fonksiyonunun konveks bir yapıda olduğunu ve  $t \in [0,1]$  için  $\max(1-t, t) \leq A(t) \leq 1$  sınırları arasında değer aldığını biliyoruz.  $A(t) = \max(1-t, t)$  olan alt sınırimız tam bağımlılığı temsil ettiğinden ve  $A(t) = 1$  olan üst sınırimız bağımsızlığın sembolü olacağından Şekil 4.28’e göre kuyruk bağımlılığı en yüksek değerle ifade olunan modellerimiz gri renkle temsil edilen CT ve mavi renkle temsil edilen ANEGLOG modellerimizdir. Ancak bu modellerden hiç biri tam bağımsızlığı veya tam bağımlılığı işaret etmemektedir.



**Şekil 4.28:** İki değişkenli dağılımlar için Pickand bağımlılık fonksiyonları.

Kesikli çizgi ile belirtilen eğri parametrik olmayan yöntemler kullanılarak elde edilen bağımlılık yapısıdır. Grafikte ALOG (kırmızı) ve ANEGLOG (mavi) bağımlılık fonksiyonları diğerlerinden daha yüksek bağımlılık göstermelerine karşılık bunlarla kıyaslandığında kesikli çizgi ile belirtilen veriyi kullanarak parametrik olmayan yöntemler ile elde edilen bağımlılık yapısından uzak olduğu gözükmemektedir. Diğer bağımlılık fonksiyonları birbirlerine yakın olmakla beraber LOG (siyah) bağımlılık fonksiyonu kesikli çizgi ye diğerlerine nazaran daha yakın olduğundan LOG yapısını

bir adım öne çıkarmaktadır. Ayrıca, bu dokuz modelin AIC bilgi kriteri kıyaslandığında en küçük değere iki değişkenli LOG dağılım fonksiyonu sahiptir. Bu iki gösterge çalışılan problem için en uygun dağılım fonksiyonunun LOG olduğu sonucunu doğurmuştur. Buradan elde edilen sonuç da kopula seçimiyle uyuşmaktadır.

Tek değişkenli modellemelerde olduğu gibi iki değişkenli modellemelerde de risk seviyesi öngörülere olan tekrar seviyeleri hesaplamaları yapılmıştır. Çizelge 4.18’de  $T_x$  deniz tekrar seviyesi,  $T_y$  basınç tekrar seviyesi (tek değişkenli blok maksimum yönteminden elde edilen) ve  $T_{x,y}$  deniz veya basınç değişkenlerinin birleşik tekrar seviyesi olmak üzere iki değişkenli blok maksimum yöntemine göre seçtiğimiz LOG/Kopula modeliyle elde edilen birleşik tekrar seviyeleri kıyaslamaları verilmiştir.

**Çizelge 4.18:** Tek değişkenli blok maksimum yöntemiyle elde edilen 2, 20, 100 senelik tekrar seviyelerinin iki değişkenli blok maksimum modeline göre eriştiği tekrar seviyeleri.

Marjinal Dağılım (Deniz,Basınç)	Tekrar Seviyesi, $T_x=T_y$	Deniz Seviyesi	Basınç Şiddeti	Tekrar Seviyesi, $T_{x,y}$ (Gumbel Kopula veya LOG Modeline)
(Gev,Gumbel)	2	0.4694863	992.1053	5.57
(Gev,Gumbel)	20	0.6675585	984.1187	59.16
(Gev,Gumbel)	100	0.791555	979.1189	218.27

Marjinal dağılımları, deniz seviyesi için 3 parametrelili GEV ve basınç şiddeti için 2 parametrelili Gumbel alındığında uygun kopula modelinin Gumbel olduğundan bahsedilmiştir. Bu model kullanılarak tek değişkenli blok maksimum için 2, 20 ve 100 yıla karşılık gelen deniz ve basınç tekrar seviyeleri kullanılarak iki değişkenli tekrar seviyeleri elde edilmiştir. Buna göre, tekrar seviyelerinin 2 ve 20 yıl için yaklaşık 3 katı ve 100 yıl için ise yaklaşık 2 katı bir sürede aşılacağı tespit edilmiştir. Özetle iki değişkenli blok maksimum yöntemi daha iyimser bir risk öngörüsünde bulunmaktadır.

İki deęişkenli blok maksimum için seçilen LOG modeline göre dağılım dilimi kullanarak elde edilen eşik seviyeleri için hesaplanan birleşik tekrar seviyeleri çizelge 4.19’da verilmiştir.

**Çizelge 4.19:** İki deęişkenli blok maksimum için seçilen LOG modeline göre dağılım dilimi kullanarak elde edilen eşik seviyeleri için hesaplanan birleşik tekrar seviyeleri.

Dağılım dilimi	Deniz Eşik Seviyesi	Basınç Eşik Seviyesi	Bileşik Tekrar Seviyesi $T_{x,y}$ (LOG modeli)
0.900	0.4756	994.500	4.210
0.950	0.5418	992.520	7.190
0.990	0.6281	987.244	20.29
0.995	0.6646	983.104	38.54
-	0.77	980	100.22

Bu yaklaşımı kullanarak (0.5418, 992.52) seviyesinin, yani deniz seviyesinin 0.5418’i aşması ve/veya basınç deęerinin 992.52 seviyesinin altında kalması ortalamada 7 senede bir gerçekleşecektir. Benzer şekilde, (0.62816, 987.244) seviyesinin, yani deniz seviyesinin 0.62816’yı aşması ve/veya basınç deęerinin 987.244 seviyesinin altında kalması ortalamada 20 senede bir gerçekleşecektir. Veride olmadığı için dağılım dilimi verilmeyen, ancak seçilen (0.77, 980) seviyesinin, yani deniz seviyesinin 0.77’yi aşması ve/veya basınç deęerinin 980 seviyesinin altında kalması ortalamada 100 senede bir gerçekleşecektir.

Çizelge 4.18’den, tek deęişkenli blok maksimum yöntemine göre elde edilen tekrar seviyeleri iki deęişkenli blok maksimum yöntemine göre daha yüksek deniz seviyesi tahmininde bulunmaktadır. Literatürde de farklı problemler için benzer çalışmalarda da tek deęişkenli blok yönteminin iki deęişkenli blok yönteminden daha yüksek deęerler verdiği gözlenmektedir.

#### 4.5.2 Eşik seviyesini aşan değerler yöntemi üzerine sonuçlar

İki Değişkenli POT yönteminden Bölüm 3.2’de bahsedilmiştir ve yukarıda elde edilen LOG modeline bağlı eşik değeri  $u = (0.371, 998.6)$  alındığında elde edilen sonuçlar Şekil 4.29’da belirtilmiştir. Burada 2, 20 ve 100 yıllık tekrar seviyeleri kontor eğrileri ile ifade edilmiştir.

Çizelge 4.20’de dağılım dilimleri ve buna karşılık gelen deniz seviyesi ve basınç şiddeti değerleri ve bunlardan en az birinin tüm verideki aşım oranı ve sayıları listelenmiştir.

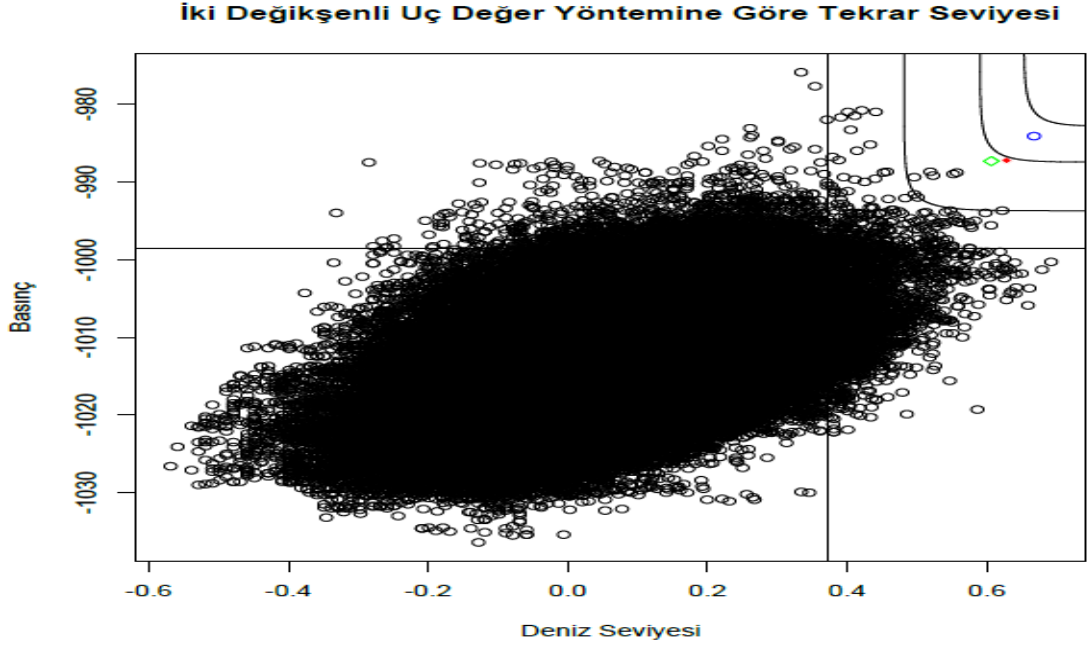
**Çizelge 4.20:** Dağılım dilimleri ve buna karşılık gelen deniz seviyesi, basınç şiddeti değerleri ve bunlardan en az birinin tüm verideki aşım oranı.

Dağılım dilimi (quantile)	Deniz Seviyesi	Basınç şiddeti	Tüm veri setindeki aşım sayıları	Tüm verideki aşım oranı
0.90	0.231	1004.0	41607	0.1703
0.95	0.277	1002.3	21597	0.0884
0.99	0.371	998.60	4571	0.0187

Çizelge 4.20’deki eşik değer seçimi, saatlik gözlemlerdeki 0.99 dağılım dilimi (quantile) olarak alınmıştır.

Ayrıca kıyaslama amacıyla tek değişkenli ve iki değişkenli blok maksimum yöntemlerine göre elde edilen 20 yıllık tekrar seviyeleri Şekil 4.29’da mavi ve kırmızı noktalarla, tek değişkenli POT yöntemine göre elde edilen 20 yıllık tekrar seviyesi ise yeşil nokta ile işaretlenmiştir.





**Şekil 4.29:** İki Değişkenli POT yöntemine göre 2, 20 ve 100 tekrar seviyeleri kontör eğrileri.

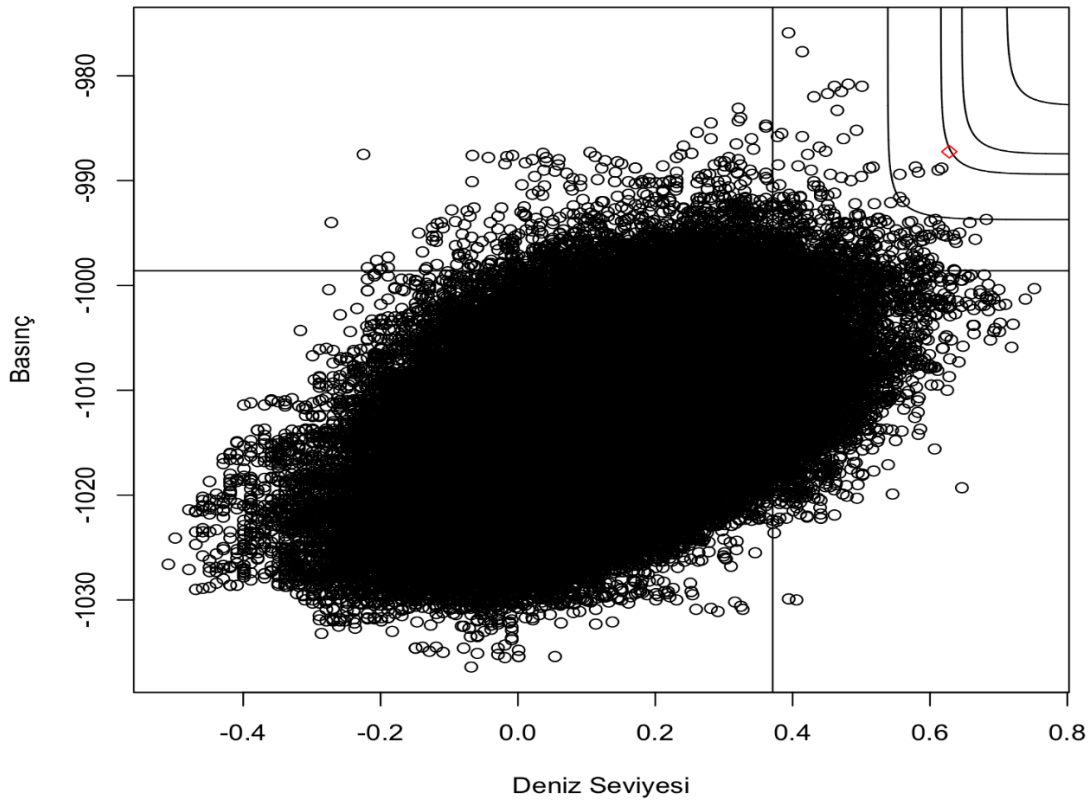
**Şekil 4.29'a bağlı yorumlar:**

1. (0.371, 998.6) değeri eşik seviyesi olarak seçilmiştir. Bu nokta grafikte belirtilen yatay ve dikey doğrularının kesişim noktasıdır.
2. Bu eşik seviyesinin üstünde kalan 3 tane kontör eğrisi sırasıyla 2, 20 ve 100 senelik tekrar seviyelerini temsil etmektedir. Örneğin 20 senelik kontör eğrisi üzerindeki her nokta aynı riski temsil etmektedir.
3. Bu şekil iki değişkenli blok maksimum metodu kullanılarak LOG (Gumbel kopula) modeline göre elde edilen ortalamada 20 senede bir karşılaşılabilecek seviyenin bu bölümde tanımladığımız iki değişkenli POT yöntemiyle 19 senede bir erişilebileceğini göstermektedir ve bu seviye kırmızı nokta ile işaretlenmiştir.
4. Basınç seviyesinin 985 mmbar a düşmesi kendi başına deniz seviyesinin 0.60' a yükselmesine neden olabileceği öngörülebilir.
5. Bu grafikte deniz seviyesi ve basınç şiddeti verileri için tek değişkenli POT modeli uygulayarak ortalamada 20 senede bir karşılaşılabilecek seviyelere iki değişkenli POT yöntemi kullanılarak 16 senede bir erişileceği gösterilmektedir ve bu seviye yeşil nokta ile işaretlenmiştir.

6. Bu grafikte deniz seviyesi ve basınç şiddeti verileri için tek değişkenli blok maksimum metoduna göre ortalamada 20 senede bir karşılaşılabilecek seviyelere bu bölümde tanımladığımız iki değişkenli POT yöntemi kullanılarak 55 senede erişilebileceği gösterilmektedir ve bu seviye mavi nokta ile işaretlenmiştir.

Tek değişkenli POT yönteminde olduğu gibi burada da iyimser bir tahminle yıllık yükselme trendi 3 mm/yıl olarak alınınca 20 yıl sonra MSL bugünküne nazaran 6 cm daha yüksek olacağından bahsedilmiştir. Bu durumda yukarıdaki tekrarlar periyotlarında deniz seviyesinin hangi değerlere ulaşabileceğini öngörebilmek için mevcut deniz seviyelerine 6 cm (0.06m) ekleyerek iki değişkenli POT yöntemi uygulandı. Şekil 4.30'daki kontör eğrileri sırasıyla 2, 10, 20 ve 100 senelik riskleri vermektedir. Kırmızı noktanın bulunduğu eğri 10 seneye karşılık gelmektedir. Bu noktanın şekil 4.29'da 20 senelik kontör eğrisi üzerinde gözüktüğü bilgisinden yararlanarak bu model sonuçlarına göre bugün 20 yılda bir gözükene seviyenin 10 senede bir gözükene anlamına gelmektedir.

#### İki Değişkenli Uç Değer Yöntemine Göre Tekrar Seviyesi



**Şekil 4.30:** MSL baz alındığında iki değişkenli POT yöntemine göre tekrar seviyeleri kontör eğrileri.

## 5. SONUÇ VE ÖNERİLER

Bu çalışmada İzmir Körfezindeki deniz seviyesi ölçümlerinden ve basınç şiddeti değerlerinden yararlanılarak ani su baskınlarının modellenmesi ve ileriye yönelik 2, 20 ve 100 yıllık risk öngörülerinde bulunulmuştur. Öncelikle deniz seviyesi ve basınç şiddeti değerleri için ayrı ayrı tek değişkenli uç değerler teorisinden yararlanılarak blok maksimum, POT ve ACER yöntemleriyle istatistiksel modelleme üzerine uygulamalar gerçekleştirilmiştir. Blok maksimum için blok büyüklüğü yıllık olarak seçilmiş ve üç parametrelili GEV ve iki parametrelili Gumbel modellerinin her iki değişken içinde uygun olduğu istatistiksel testlerle kanıtlanmıştır. İstatistiksel olarak daha basit model tercihimiz olacağından her iki değişken için Gumbel modeli tercih edilmiştir. Saatlik veri seti alınarak her iki değişken içinde POT yöntemi çalışılmış ve Genelleştirilmiş Pareto modelinin uygun olduğu sonucuna varılmıştır. POT yöntemi, blok maksimum yöntemine nazaran veriyi alt gruplara ayırmadan bir bütün halinde inceleyerek daha fazla veriden yararlandığından ötürü tercih sebebi olmuştur. Aynı şekilde saatlik veri seti alınarak her iki değişken için daha güncel bir metot olan ACER yöntemi uygulanmıştır. Bu yöntem asimptotik argümanlara ihtiyaç duyulmadan çalışılmıştır ve diğer iki yöntemin aksine durağanlık şartı aranmadan uygulanabilmiştir. Blok maksimum yöntemiyle kıyaslandığında hem daha çok veriden yararlanılarak veriyi bir bütün halinde incelemiş olması hem de durağanlık şartı aranmadan yani mevsimsellik etkisine maruz bırakılmadan uygulanabilmiş olmasından ötürü tercih sebebi olmuştur. POT yöntemi ile kıyaslandığında birbirlerine benzer mantıkta iki yöntem olsalar da ACER yönteminin daha güncel bir metot oluşu, olasılıksal bir dağılım fonksiyonunu temsil etmesi yerine numerik bir hesaplama yöntemi sunmuş olmasından ötürü tercih edilebilir. Ayrıca  $(k-1)$  önceki koşullu saptanmış adımın sifıra eşit olması halinde POT yöntemindeki gibi klasik yaklaşımı temsil etmesi ve metodun alt formunun özel halde uç değer dağılım fonksiyonlarına (GEV) tekabül etmesinden ötürü ACER yöntemi tercih edilecek modelimiz olacaktır.

ACER yönteminin tek değişkenli deniz ve basınç değerlerinin 2, 20 ve 100 yıllık risk öngörülere Çizelge 5.1’de verilmiştir.

**Çizelge 5.1:** Deniz ve Basınç değişkenleri için tek değişkenli ACER yöntemi tekrar seviyesi sonuçları.

Tekrar seviyesi	2 sene	20 sene	100 sene
Deniz	0.46	0.64	0.73
Basınç	993.98	986.69	982.62

Tez çalışmasının ikinci aşamasında ise deniz seviyesi ve basınç seviyesi ve aralarındaki bağımlılık ilişkisi dikkate alınarak iki değişkenli modellemeler yapılmıştır. Öncelikle iki değişkenli blok maksimum için blok büyüklüğünün aylık alınmasının uygun olduğu istatistiksel testlerle ortaya konmuştur. Daha sonra hem kopula hem de iki değişkenli uç değerler modellemeleri çalışılmıştır. Kopula seçimi için uç değer kopula ailesinin seçiminin uygunluğu tartışılmış ve daha sonra bu uç değer kopulalar arasında marjinal dağılımları deniz seviyesi için GEV ve basınç şiddeti için ise Gumbel alınarak Gumbel Kopulanın uygun olduğu sonucuna varılmıştır. Yine aynı yöntem için iki değişkenli 9 farklı uç değer dağılımları arasında en uygun olanı, Pickand fonksiyonları ve uyum iyiliği testleri aracılığı ile, LOG modeli olarak seçilmiştir. Bu da bulunan kopula yöntemiyle uyuşmaktadır. Çizelge 4.18’de tek değişkenli blok maksimum için 2, 20 ve 100 yıla karşılık gelen deniz ve basınç tekrar seviyeleri kullanılarak Gumbel kopula yardımı ile iki değişkenli tekrar seviyeleri elde edilmiştir. Buna göre, tekrar seviyeleri 2 ve 20 yıl için yaklaşık 3 katı ve 100 yıl için ise yaklaşık 2 katı bir sürede aşılacağı tespit edilmiştir. Özetle iki değişkenli blok maksimum yöntemi daha iyimser bir risk öngörüsünde bulunmaktadır.

Çizelge 4.19’da ise iki değişkenli blok maksimum için seçilen LOG modeline göre dağılım dilimi kullanarak elde edilen eşik seviyeleri için hesaplanan bileşik tekrar seviyeleri verilmiştir. Bu yaklaşımı kullanarak (0.5418, 992.52) seviyesinin, yani deniz seviyesinin 0.5418’i aşması ve/veya basınç değerinin 992.52 şiddetinin altında kalması ortalama 7 senede bir gerçekleşecektir. Benzer şekilde, (0.62816, 987.244) seviyesinin, yani deniz seviyesinin 0.62816’yı aşması ve/veya basınç şiddetinin 987.244 seviyesinin altında kalması ortalama 20 senede bir gerçekleşecektir.

Seçilen (0.77, 980) seviyesinin, yani deniz seviyesinin 0.77'yi aşması ve/veya basınç şiddetinin 980 seviyesinin altında kalması ortalama 100 senede bir gerçekleşecektir. Bu sonuçlardan, tek değişkenli metodun daha yüksek deniz seviyesi tahmininde bulunduğunu söyleyebiliriz. Literatürde de farklı problemler için benzer çalışmalarda da tek değişkenli blok maksimum yönteminin iki değişkenli blok maksimum yönteminden daha yüksek değerler verdiği gözlenmektedir.

Çalışmamızın son bölümünde, iki Değişkenli POT yöntemi eşik değeri (0.371, 998.6) seçildiğinde oturtulan LOG modelin seçimine göre 2, 20 ve 100 yıllık tekrar seviyeleri kontör eğrileri ile Şekil 4.29'da ifade edilmiştir. Burada eşik değer seçimi, saatlik gözlemlerdeki 0.99 dağılım dilimi (quantile) olarak alınmıştır.

Dört farklı yöntemlere göre risk öngörü kıyaslaması.

1. Tek değişkenli blok maksimum yöntemiyle iki değişkenli blok maksimum yöntemi kendi aralarında kıyaslandığında tek değişkenli yöntem çok daha büyük bir risk seviyesi öngörmektedir.
2. Diğer yandan tek değişkenli POT yöntemi iki değişkenli POT yöntemiyle kıyaslandığında risk seviyeleri yakın olmakla beraber iki değişkenli yöntem bir miktar daha yüksek risk seviyesini öngörmektedir.
3. Tek değişkenli blok maksimum yöntemi ile tek değişkenli POT yöntemi aralarında kıyaslandığında blok maksimum yöntemi çok daha yüksek risk seviyesi öngörmektedir.
4. Sonuç olarak tek ve iki değişkenli blok maksimum ve POT yöntemlerini incelediğimizde en yüksek risk seviyesi öngörüsü tek değişkenli blok maksimum metoduyla diğer yandan en düşük risk seviyesi öngörüsü ise iki değişkenli blok maksimum metoduyla gözlemlenmiştir.

Farklı yöntemler farklı risk seviyeleri öngörmektedir. Ancak iki değişkenli yaklaşımlar deniz ve basınç değişkenleri arasındaki bağımlılığı dikkate aldığından daha hassas bir modelleme ortaya çıkmaktadır. Diğer yandan POT metotları seçilen bir eşik seviyesine bağlı olarak blok maksimum yöntemleriyle kıyaslandığında daha çok veriden yararlanabildiğinden ve ayrıca mevsimsellik etkisinden arınabildiğinden tercih edilmesi görüşü ağır basmaktadır. Bu gerekçelerden ötürü bu çalışmada iki değişkenli POT yöntemi ile yapılan modelleme daha ön plana çıkmaktadır.



## KAYNAKLAR

- [1] **Adlouni, S. E., Bobe'e, B., Quarda, T.B.M.J.** (2008). On the tails of extreme event distributions in hydrology, *Journal of Hydrology*, 355, 16-33.
- [2] **Fisher, R., Tippett, L. H.** (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample, In: *Proceedings of the Cambridge Philosophical Society*, 24, 180-190.
- [3] **Gnedenko, B.V.** (1943). Sur la distribution limite du terme maximum d'une serie aleatoire, *Annals of Mathematics*, 44, 423-453.
- [4] **Gumbel, E.J.** (1958). *Statistics of Extremes*, Columbia University Press, New York, 371.
- [5] **Balkema, A. A., De Haan, L.** (1974). Residual life time at great age, *Annals of Probability*, 2, 792-804.
- [6] **Pickands, J.** (1975). Statistical inference using extreme order statistics, *Annals of Statistics* 3, 119-131.
- [7] **Embrechts, P., Küppelberg, C., Mikosch T.** (2003). *Modelling Extremal Events for Insurance and Finance Applications of Mathematics*, 33, Springer, 648.
- [8] **Reiss, R.D., Thomas, M.** (2001). *Statistical Analysis of Extreme Values*, Second ed. Birkhauser, Basel/Boston/Berlin.
- [9] **Coles, S., Heffernan, J., Tawn, J.** (1999). Dependence measures for extreme value Analyses, *Extremes*, 2(4), 339-365.
- [10] **Jenkinson, A.F.** (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements, *Q.J. Roy. Meteorol. Soc.*, 87, 158-171.
- [11] **Todorovic, P., Zelenhasic, E.** (1970). A stochastic model for flood analysis, *Water Resources Research*, 6(6), 1641-1648.
- [12] **Smith, R.L.** (1984). Threshold methods for sample extremes, *Statistical extremes and applications*, 621-638.
- [13] **Davison, A.C.** (1984). Modelling excesses over high thresholds, with an application, *Statistical Extremes and Applications*, 461-482.
- [14] **Davison, A.C., Smith, R.L.** (1990). Models for exceedances over high thresholds (with discussion), *J. R. Statist. Soc. B*, 52, 237-254.
- [15] **Kharin, V. V., Zwiers, F. W., Zhang, X., Hegerl, G. C.** (2007). Changes in temperature and precipitation extremes in the IPCC ensemble of global coupled model simulations, *Journal of Climate* 20, 1419-1444.
- [16] **Katz, R. W. Parlange, M. B., Naveau, P.** (2002). Statistics of extremes in hydrology, *Advances in Water Resources* 25, 1287-1304.

- [17] **Madsen, H., Rasmussen, P. F., Rosbjerg, D.** (1997). Comparison of annual maximum series and partial duration series methods for modeling extreme hydrologic events1, At-site modeling, *Water Resources Research* 33, 747–757.
- [18] **Naveau, P., Guillou, A., Cooley, D., Diebolt, J.** (2009). Modelling pairwise dependence of maxima in space, *Biometrika* 96, 1–17.
- [19] **Van Den Brink, H. W., Können, G. P., Opsteegh, J. D.** (2005). Uncertainties in extreme surge level estimates from observational records, *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* 363, 1377–1386.
- [20] **De Valk, C.** (1993). Estimation of marginals from measurements and hindcast data, WL—Delft Hydraulics Report H1700.
- [21] **Prescott, P., Walden, A.T.** (1980). Maximum likelihood estimation of the parameters of the generalized extreme-value distribution, *Biometrika* 67, 723-724.
- [22] **Hosking, J.R.M., Wallis, J.R., Wood, E.F.** (1985). Estimation of the Generalized Extreme-Value Distribution by the Method of Probability Weighted Moments, *Technometrics* 27, 251–261.
- [23] **Galiatsatou, P., Prinos, P.** (2012). Reducing uncertainty in extreme waves and storm surges using a combined extreme value model and wavelets, *Journal of Coastal Engineering*, 2.
- [24] **Sklar, A.** (1959). Fonctions de repartition a n dimensions et leurs marge, *Publ. Inst. Statist. Univ. Paris*, 8, 229-231.
- [25] **Nelsen, R. B.** (2006). An introduction to copulas (2<sup>nd</sup> Edition), New York: Springer.
- [26] **Ferguson, T. S.** (1983). Bayesian density estimation by a mixture of normal distributions, *Recent Advances in Statistics: Papers in Honor of Herman Chernoff on his 60<sup>th</sup> Birthday*, 202-287.
- [27] **Genest, C., Gendron, M., Bourdeau, B. M.** (2009). The advent of copulas in finance, *Eur J Financ* 15(7–8), 609–618.
- [28] **Fisher, N. I.** (1997). Copulas, In: *Encyclopedia of Statistical Sciences*, 1, 159-163.
- [29] **Galambos, J.** (1987). *The Asymptotic Theory of Extreme Order Statistics*, 2nd edn., R.E. Krieger Publishing Co. Inc, Melbourne, FL.
- [30] **Ghoudi, K., Khoudraji, A., Rivest, L. P.** (1998). Propriétés statistiques des copules de valeurs extrêmes bidimensionnelles, *The Canadian Journal of Statistics* 26, 1, 187-197.
- [31] **Gudendorf, G., Segers, J.** (2010). Extreme-value copulas, in: P. Jaworski (Ed.), *Proceedings of the Workshop on Copula Theory and Its Applications*, Springer Media.
- [32] **Gumbel, E.** (1960). Bivariate exponential distributions, *Journal American Statistical Association*, 55, 698-707.
- [33] **Hüsler, J., Reiss, R. D.** (1989). Maxima of Normal Random Vectors: Between Independence and Complete Dependence, *Statistics and Probability Letters*, 7, 283–286.



- [34] **Demarta, S., McNeil, A. J.** (2005). The t copula and related copulas, *International Statistical Review*, 73(1), 111-129.
- [35] **Tiago de Oliveira, J.** (1980). Bivariate extremes: foundations and statistics, In *Multivariate Analysis V* (ed Krishnaiah PR), 349-366.
- [36] **Pickands, J.** (1981). Multivariate extreme value distributions, In *Proceedings of the 43rd session of the International Statistical Institute*, 2(49), 859–878, 894–902.
- [37] **Deheuvels, P.** (1991). On the limiting behavior of the Pickands estimator for bivariate extreme value distributions, *Statistics & Probability Letters*, 12, 429-439.
- [38] **Cap'era`a, P., Foug`eres, A. L., Genest, C.** (1997). A nonparametric estimation procedure for bivariate extreme value copulas, *Biometrika*, 84, 567-577.
- [39] **Gumbel, E. J.** (1961). Bivariate logistic distributions, *Journal of the American Statistical Association*, 56(294), 335–349.
- [40] **Tawn, J. A.** (1988). Bivariate extreme value theory: Models and estimation, *Biometrika*, 75(3), 397–415.
- [41] **Galambos, J.** (1975). Order statistics of samples from multivariate distributions, *Journal of the American Statistical Association*, 70, 674–680.
- [42] **Joe, H.** (1994). Multivariate extreme-value distributions with applications to environmental data, *The Canadian Journal of Statistics*, 1, 47–64.
- [43] **Næss, A., Gaidai, O.** (2009). Estimation of extreme value from sampled time series, In *Structural Safety*, 31, 325–334.
- [44] **Naess, A., Gaidai, O., Karpa, O.** (2013). Estimation of extreme values by the average conditional exceedance rate method, *Journal of Probability and Statistics*, 15.
- [45] **Von Mises, R.** (1954). La distribution de la plus grande de n valeurs, *Selected Papers*, Providence, R.I.: Amer. Math. Soc., 2, 271-294.
- [46] **Kotz, S., Nadarajah, S.** (2000). *Extreme Value Distributions: Theory And Applications*. London: Imperial Collage Press.
- [47] **Smith, R. L.** (1989). Extreme Value analysis of environmental time series: Application to trend detection in Ground-Level Ozone, *Statist. Sc.*, 4, 367-393.
- [48] **Leadbetter, M. R.** (1991). On a basis for Peaks over Threshold modelling, *Statistics and Probability Letters*, 12(4), 357-362.
- [49] **Pickands, J.** (1971). The two dimensional Poisson process and extremal processes, *J. Appl. Probab.*, 8, 745-756.
- [50] **De Haan, L., Resnick, S.I.** (1977). Limiting theory for multivariate extremes, *Z. Wahrsch Theor*, 40, 317-337.
- [51] **De Haan, L.** (1985). Extremes in higher dimensions: the model and some statistics, In: *Proc 45<sup>th</sup> Session Inter. Statist. Inst. Amsterdam*, paper 26,3.
- [52] **Coles, S.G., Tawn, J. A.** (1994). Statistical methods for multivariate extremes: an application to structural design (with discussion), *Appl. Statist.*, 43, 1-48.
- [53] **Salvadori, G., De Michele C.** (2007). On the use of copulas in hydrology theory and practice, *J. Hydrol. Eng.*, 12(4), 369-380.
- [54] **Poulin, A., Huard, D., Favre, A.C., Pugin, S.** (2007). Importance of tail dependence in bivariate frequency analysis, *J. Hydrol. Eng.*, 12(4), 394-403.

- [55] **Coles, S. G., Tawn, J. A.** (1991). Modelling extreme multivariate events, *J. Roy. Statist. Soc., B*, 53, 377-392.
- [56] **Joe, H.** (1990). Families of min-stable multivariate exponential and multivariate extreme value distributions, *Statist. Probab. Letters*, 9, 75–81.
- [57] **Tawn, J. A.** (1990). Modelling multivariate extreme value distributions, *Biometrika*, 77, 245-253.
- [58] **Rootzén, H., Tajvidi, N.** (2006). Multivariate generalized Pareto distribution, *Bernoulli* 12, 917-930.
- [59] **Nelsen, R. B.** (1999). *An Introduction to Copulas*, Springer, New York.
- [60] **Yue, S., Wang C. Y.** (2004). A comparison of two bivariate extreme value distributions, *Stochastic Environmental Resea*, 18, 61-66.
- [61] **Schafer, J. L., Olsen, M. K.** (1998). Multiple imputation for multivariate missing-data problems: A data analyst's perspective, *Multivariate Behavioral Research*, 33(4), 545-571.
- [62] **Abberger, K.** (2005). A simple graphical method to explore tail -dependence in stock-return pairs, University of Konstanz.
- [63] **Fisher, N. I., Switzer, P.** (1985). Chi-plots for assessing dependence, *Biometrika*, 72, 253-265.
- [64] **Fisher, N. I., Switzer P.** (2001). Graphical assessment of dependence: is a picture worth 100 tests, *The American Statistician*, 55, 233-239.
- [65] **Coles, S.** (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer Series in Statistics, Springer Verlag London, 208.

## ÖZGEÇMİŞ

**Ad-Soyad** : Halil İbrahim ALTUN  
**Doğum Tarihi ve Yeri** : 19.01.1992, İSTANBUL  
**E-posta** : altunh@itu.edu.tr

### ÖĞRENİM DURUMU:

- **Lisans** : 2015, İTÜ, Fen Edebiyat Fakültesi, Matematik Mühendisliği
- **Yüksek Lisans** : 2018, İTÜ, Fen Bilimleri Enstitüsü, Matematik Mühendisliği