

**Zaman Serilerine Dayalı Salgın Tespit Algoritmalarını
Karşılaştırma Amaçlı Bir Yazılım Aracının Tasarım ve
Gerçekleştirimi**

**Desing and Implementation of a Software Tool for
Comparing Methods in Outbreak Detection Based On
Time Series**

YASİN ŞAHİN

Hacettepe Üniversitesi
Lisansüstü Eğitim – Öğretim ve Sınav Yönetmeliğinin
BİLGİSAYAR Mühendisliği Anabilim Dalı İçin Öngördüğü
YÜKSEK LİSANS TEZİ
olarak hazırlanmıştır.

2013

Fen Bilimleri Enstitüsü Müdürlüğü'ne,

Bu çalışma jürimiz tarafından **BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI'nda YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Başkan (Danışman) :
(Prof.Dr. Ali SAATÇI)

Üye :
(Doç. Dr. Erdem KARABULUT)

Üye :
(Yrd. Doç. Dr. Kayhan İMRE)

Üye :
(Yrd.Doç. Dr. Harun ARTUNER)

Üye :
(Öğr. Gör. Dr. Ahmet Burak CAN)

ONAY

Bu tez/...../..... tarihinde Enstitü Yönetim Kurulunca kabul edilmiştir.

...../...../.....

Prof.Dr. Fatma Sevin DÜZ
FEN BİLİMLERİ ENSTİTÜSÜ MÜDÜRÜ

Geniřleyen aileme...

Zaman Serilerine Dayalı Salgın Tespit Algoritmalarını Karşılaştırma Amaçlı Bir Yazılım Aracının Tasarım ve Gerçekleştirimi

Yasin ŞAHİN

ÖZ

11 Eylül 2001 saldırılarının ardından, biyolojik teröre karşı erken uyarı sistemleri önem kazanmıştır. Bu bağlamda *EARS*(*Early Abberation Reporting System*) geliştirilmiş ve olası bir biyolojik terör saldırısının, aykırı bir durum olarak otomatik tespiti hedeflenmiştir. *EARS* yazılımının kullandığı klasik algılama yöntemlerinin, yavaş yayılan salgınların kaynağını bulmadaki başarımı düşük olduğundan, araştırmacılar, özellikle *Cusum*(*Cumulative Sum*), *EWMA*(*Exponentially Weighting Moving Average*), *NBC*(*Negative Binomial Cumulative Sum*) gibi istatistiksel süreç yöntemlerini (*SPC*) kullanarak kıyaslama yapmış; bunun sonucunda sistem başarımının veri kümesine bağımlı olduğu tespit edilmiştir. Bu tez kapsamında, bu amaca dönük olarak, WEB tabanlı bir yazılım aracı geliştirilmiş; bu araç ile, bir araştırmacının elindeki veri kümesine uygun veri ön-işleme ve uygun istatistiksel algılama (salgın tespit) yöntemini belirleyebilmesi amaçlanmıştır.

Anahtar Kelimeler: salgın tespiti, aykırı durum algılama, biyolojik terör, sürveyans sistemi, zaman serileri

Danışman: Prof. Dr. Ali SAATÇI, Hacettepe Üniversitesi, Bilgisayar Mühendisliği Bölümü

Desing and Implementation of a Software Tool for Comparing Methods in Outbreak Detection Based On Time Series

Yasin ŞAHİN

ABSTRACT

After the 9/11 terrorist attacks, early outbreak detection systems against bio-terrorism have gained a lot of importance. In this context, EARS (Early Abberation Reporting System) has been developed in order to detect automatically bio-terror attacks which can be observed as an anomaly in public health data. Because methods used in EARS have shown weaknesses in detecting slowly propagated epidemics, much more elaborated statistical methods such as Cusum, EWMA, NBC have been developed to replace old methods. Studies have shown that the performance of the method used is closely related to the nature of the data set scanned. In this thesis a WEB based software tool is designed and implemented to compare methods in order to find the adequate one for a given set of data.

Keywords: outbreak detection, anomaly detection, bio-terrorism, surveillance system, time series

Advisor: Professor ALİ SAATÇI, Hacettepe University, Department of Computer Engineering.

TEŐEKKÜR

Tez alıőmaları ve eserin yazılması sırasında bilgi ve tecrubesini esirgemeyen, tez danıőmanım Prof. Dr. Sn. Ali Saati'ye,

Tez alıőmaları sırasında maddi desteęi ile tez konusunun belirlenmesinde bulunduęu katkılardan dolayı Turkcell Akademi'ye,

Lisans ve Lisansüstü eęitimimi aldıęım ve byk bir baęlılık duyduęum blmm Hacettepe niversitesi Bilgisayar Mhendislięi'nin eski yeni tm alıőanlarına,

Lisansst eęitiminin ilk dneminde verdięi maddi ve manevi destekten tr Hacettepe niversite Bilgisayar Mhendislięi blm baőkanı Prof. Dr. Sn. Hayri Sever'e,

Tez alıőmalarımda fikirleri ile bana destek olan sevgili iő arkadaşlarıma,

Tez alıőmaları sırasında ve dięer zamanlarda manevi desteęini esirgemeyen ve eserin her virglne tesir eden *mstakbel* eőim Sn. Ceren Beőler'e ve aldıęım her kararda yanımda duran ok sevgili –geniőleyen- aileme teőekkr ederim.

İÇİNDEKİLER DİZİNİ

	<u>Sayfa</u>
ÖZ	i
ABSTRACT	ii
TEŞEKKÜR	iii
İÇİNDEKİLER DİZİNİ	iv
ÇİZELGELER DİZİNİ	ix
SİMGELER VE KISALTMALAR DİZİNİ	x
SÖZLÜK	xi
1. GİRİŞ	1
1.1. Salgın.....	1
1.2. Sendrom Sürveyansı	2
1.3. Veri Kaynakları	4
1.4. Mevcut Sistem Örnekleri.....	6
1.4.1. Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE).....	7
1.4.2. Real-Time Outbreak and Disease Surveillance (RODS).....	9
1.4.3. BioSense	9
2. SALGIN TESPİTİ İÇİN KULLANILAN İSTATİSTİKSEL YÖNTEMLER	11
2.1. Cusum	12
2.2. EWMA.....	13
2.3. Shewhart.....	14
2.4. EARS	15
2.5. Diğer Yöntemler	17
2.6. Başarım Değerlendirme Ölçütleri.....	18
3. VERİ ÖN İŞLEME YÖNTEMLERİ	22
3.1. Regresyon	23
3.1.1. Basit Doğrusal Regresyon	23
3.1.2. Çoklu Doğrusal Regresyon	23
3.1.3. Adaptive Regression.....	24

3.2. 7 Gün Farkı.....	25
3.3. Hareketli Ortalama.....	25
3.3.1. Basit Hareketli Ortalama.....	25
3.3.2. Ağırlıklı Hareketli Ortalama.....	26
3.3.3. Üstel Ağırlıklı Hareketli Ortalama.....	26
3.4. Holt Yöntemi.....	26
3.5. Holt ve Winters Yöntemi.....	27
3.6. Ayırışım (Decomposition).....	28
3.6.1. Haftanın Günü Etkisi Ayırışımı.....	30
3.6.2. Mevsimlik Etki Ayırışımı.....	31
3.7. Veri Ön İşleme Başarım Ölçütleri.....	32
4. GERÇEKLEŞTİRİM.....	33
4.1. Sistem Mimarisi ve Gereksinimler.....	34
4.2. Kullanılan Teknolojiler.....	37
4.2.1. ZKoss (ZK).....	38
4.2.2. PostgreSQL.....	38
4.2.3. Hibernate (Java Persistence API).....	38
4.2.4. Kullanılan Kütüphaneler.....	39
4.2.4.1. Apache Commons Math.....	39
4.2.4.2. OpenCsv.....	39
4.2.4.3. JFreeChart.....	39
4.3. Outbreak Detection Tool ve Örnek Süreç.....	40
4.3.1. Veri Kümesi İşlemleri.....	40
4.3.1.1. Veri Kaynağı Ekleme.....	40
4.3.1.2. Veri Kaynakları Listeleme / Silme.....	41
4.3.2. Veri Ön İşleme İşlemleri.....	42
4.3.2.1. Basit Doğrusal Regresyon Ekranı.....	42

4.3.2.2.Çoklu Doğrusal Regresyon Ekranı	43
4.3.2.3.Adaptive Regression Ekranı	44
4.3.2.4.7 Gün Farkı Ekranı	45
4.3.2.5.Hareketli Ortalama Ekranı	45
4.3.2.6.Holt Ekranı.....	46
4.3.2.7.Holt&Winters Ekranı	46
4.3.3. Salgın Tespit İşlemleri.....	47
4.3.3.1.CUSUM Yöntemi Ekranı.....	49
4.3.3.2.EWMA Yöntemi Ekranı.....	52
4.3.3.3.EARS Yöntemleri Ekranı	53
4.3.3.4.Shewhart Yöntemi Ekranı	54
4.3.3.5.NBC Yöntemi Ekranı	54
4.3.3.6.HLM Yöntemi Ekranı	55
4.3.4. Örnek Süreç ve Sonuçları Görüntüleme	56
5. SONUÇ VE ÖNERİLER	73
6. Kaynakça	78

ŞEKİLLER DİZİNİ

	<u>Sayfa</u>
Şekil 1-1 <i>ESSENCE</i> 'den veri filtreleme ekranı örneği	8
Şekil 1-2 <i>ESSENCE</i> raporlama ve paylaşma ekranı örneği	8
Şekil 1-3 RODS 6'dan bir kullanıcı arayüzü.....	9
Şekil 1-4 <i>EARS v5</i> 'e ait etkileşim ekranı örneği.....	10
Şekil 2-1 <i>AUC</i> 'nin Trapezoidal ile hesaplanması.....	21
Şekil 4-1 ODT* Uygulama Mimarisi(* <i>Outbreak Detection Tool</i>).....	35
Şekil 4-2 MVC mimarisi	36
Şekil 4-3 ODT Teknoloji-Kütüphane Tercihleri	37
Şekil 4-4 .csv uzantılı girdi biçimi.....	41
Şekil 4-5 Veri Ekleme Ekranı.....	41
Şekil 4-6 Veri Listeleme / Silme Ekranı	42
Şekil 4-7 Basit Doğrusal Regresyon Uygulama Ekranı	43
Şekil 4-8 Çoklu Doğrusal Regresyon Uygulama Ekranı	44
Şekil 4-9 " <i>Adaptive Regression</i> " Uygulama Ekranı	44
Şekil 4-10 7 Gün Farkı Uygulama Ekranı	45
Şekil 4-11 Hareketli Ortalama Uygulama Ekranı	45
Şekil 4-12 Holt Uygulama Ekranı.....	46
Şekil 4-13 Holt&Winter Düzleştirme Yöntemi Uygulama Ekranı	47
Şekil 4-14 Salgın Tespit yönteminden elde edilen sonuçların kaydedildiği csv uzantılı kütük biçimi	48
Şekil 4-15 Otomatik Eşik Değeri Oluşturma Ekranı	48
Şekil 4-16 Cusum Test Ekranı.....	49
Şekil 4-17 EWMA Test Ekranı.....	52
Şekil 4-18 <i>EARS</i> Yöntemlerinin Test Ekranı.....	53
Şekil 4-19 Shewhart Test Ekranı	54
Şekil 4-20 NBC Test Ekranı.....	55
Şekil 4-21 HLM Test Ekranı.....	56
Şekil 4-22 Araç ile bir Sürecin Modellenmesi	58
Şekil 4-23 <i>CDC</i> tarafından sağlanan veriye pencere boyu 7 olan " <i>adaptive regression</i> " uygulandığında elde edilen sonuç grafiği. Hata! Yer işareti tanımlanmamış.	

Şekil 4-24 CDC tarafından sağlanan veriye pencere boyu 14 olan " <i>adaptive regression</i> " uygulandığında elde edilen sonuç grafiği	60
Şekil 4-25 CDC tarafından sağlanan veriye pencere boyu 28 olan " <i>adaptive regression</i> " uygulandığında elde edilen sonuç grafiği	60
Şekil 4-26 CDC tarafından sağlanan veriye pencere boyu 56 olan " <i>adaptive regression</i> " uygulandığında elde edilen sonuç grafiği	61
Şekil 4-27 CDC tarafından sağlanan veriye pencere boyu 84 olan " <i>adaptive regression</i> " uygulandığında elde edilen sonuç grafiği	61
Şekil 4-28 CDC tarafından sağlanan veriye HW $\alpha = 0,4$; $\beta=0$; $\gamma=0,15$ düzleştirme katsayıları ile uygulandığında elde edilen sonuç grafiği	62
Şekil 4-29 CDC tarafından sağlanan veriye HW $\alpha = 0,1$; $\beta=0$; $\gamma=0,15$ düzleştirme katsayıları ile uygulandığında elde edilen sonuç grafiği	63
Şekil 4-30 CDC tarafından sağlanan veriye Haftanın Günü Etkisi Ayırıştırma ve Mevsimsel Ayırıştırma uygulandığında elde edilen sonuç grafiği	64
Şekil 4-31 C1 yöntemine ait ROC eğrisi. $AUC=0,025$ olarak çıkmıştır	67
Şekil 4-32 C2 yöntemine ait ROC eğrisi. $AUC=0,024$ olarak çıkmıştır	67
Şekil 4-33 C3 yöntemine ait ROC eğrisi. $AUC=0,012$ olarak çıkmıştır	67
Şekil 4-34 CDC tarafından sağlanan veriye DOW ve Mevsimsel Örüntü Ayırıştırması Uygulandığında elde edilen sonuç grafiği.....	70
Şekil 4-35 EWMA yöntemine ait ROC eğrisi. $AUC=0,021$ olarak çıkmıştır.....	71

ÇİZELGELER DİZİNİ

	<u>Sayfa</u>
Çizelge 2-1 Gerçek sonuç ile üretilen sonuç ilişki çizelgesi.....	19
Çizelge 4-1 k ve h değerleri ile <i>ATFS</i> arasındaki ilişkiyi gösterir çizelge.	51
Çizelge 4-2 " <i>Adaptive Regression</i> " Ön işleme sonuçları	62
Çizelge 4-3 Holt&Winters Düzleştirme Yöntemleri Başarım İstatistikleri	63
Çizelge 4-4 <i>EARS</i> Yöntemlerinin Farklı Pencere Boyu ve <i>ATFS</i> değerleriyle elde ettikleri başarımları	66
Çizelge 4-5 <i>Cusum</i> Yönteminin Farklı Pencere Boyu ve <i>ATFS</i> değerleriyle elde ettikleri başarımları	69
Çizelge 4-6 <i>EWMA</i> Yönteminin Farklı <i>ATFS</i> değerleriyle elde ettiği başarımları .	71

SİMGELER VE KISALTMALAR DİZİNİ

ATFOS	Average Time to First Outbreak Signals
ATFS	Average Time to First Signals
CDC	Centers for Disease Control and Prevention
CUSUM	Cumulative Sum
DOW	Day of Week
EARS	Early Aberration Reporting System
ESSENCE	Electronic Surveillance System for the Early Notification of Community-Based Epidemics
EWMA	Exponentially Weighted Moving Average
HLM	Historical Limit Method
HW	Holt and Winters
MVC	Model View Controller
NBC	Negative Binomial Cusum
OECD	Organisation for Economic Co-operation and Development
RODS	Real-Time Outbreak and Disease Surveillance
SPC	Statistical Process Control
VTYS	Veri Tabanı Yönetim Sistemi
WHO	World Health Organization

SÖZLÜK

7 Day Differencing	7 Gün Farkı
Additive	Toplanır
Cumulative Sum	Birikimli Toplam
Curve Fitting	Eğri Uydurma
Curvilinear	Eğrisel
Cyclical	Döngüsel
Day of Week Effect	Haftanın Günü Etkisi
Decomposition	Ayrışım
Forecast	Tahmin
Intercept	Kesişim
Lag	Gecikme(li)
Moving Average	Hareketli Ortalama
Multiple Linear Regression	Çoklu Doğrusal Regresyon
Multiplicative	Çarpımsal
Multivariate	Çok Değişkenli
Outbreak	Salgın
Quadratic	Karesel
Residual	Artık
Seasonal Effect	Mevsimsel Etki
Simple Linear Regression	Basit Doğrusal Regresyon
Simulation	Benzetim
Slope	Eğim
Smoothing	Düzleştirme
Surveillance	Sürveys / Gözetim-İzleme
Syndromic Surveillance	Sendrom Sürveysı
Trend	Eğilim
Univariate	Tek Değişkenli

1. GİRİŞ

1.1. Salgın

Salgın, hastalık belirtilerinin belirli bir coğrafya ve zaman için beklenenden daha sık gözlemlenmesi durumudur¹. Beklenen değer coğrafi konum ve mevsimsel şartlara göre değişeceği için, doğrudan bir formülle ifade edilmesi oldukça güçtür. Bazen bir hastalığın tek bir kez görünmesi bile salgın olarak yorumlanabilirken, bazen ise yüzlercesi ile karşılaşılması salgın olarak değerlendirilmeyebilir. Salgının istatistiksel olarak anlamlı bir artış olarak tanımlandığı durumlarda bile, bu anlamlı artışın tanımının yapılamadığı gözlemlenmektedir².

Bir salgın, belli bir alanla kısıtlı kalabilir ya da çok geniş alanlara, ülkelere ve hatta kıtalara yayılabilir. Yine aynı şekilde, bir salgın bir hafta ya da bir ay sürebileceği gibi yıllarca da sürebilir. Eğer salgın hastalık belirli bir bölgede sıklıkla görülüyorsa endemik, belirli bir yerde ve belirli bir zaman diliminde görülüyorsa epidemik, uluslararası boyutlara ulaşmışsa da pandemik olarak adlandırılır.

Dünya Sağlık Örgütü başta olmak üzere, pek çok resmi kurum ve çalışma ekibi, salgının engellenmesi, salgın oluşmuşsa yayılmasının engellenerek en az kayıpla sürecin atlatılması konularında araştırmalar yapmış, önerilerde bulunmuşlardır. Salgının engellenmesi elbette ki en doğru, ideal yaklaşımdır ve salgının engellenmesi sürekli bir mücadele ve eylem gerektirir. Ancak yaşanan tecrübeler göstermiştir ki, salgınlar hayatın bir parçasıdır ve ne kadar önlemler alınırsa alınsın, hastalıkların canlılar arasında yayılması tam manasıyla engellenememektedir. Bu durumda ikinci sürecin devreye alınması beklenir. Ancak salgının denetlenmesini sağlamayı hedefleyen bu süreç düzenli ve güvenilir bir sürveyans sisteminin bulunmasını gerektirir. Amerikan Hastalık Kontrol Merkezi (CDC) tarafından tanımlanan salgın incelemesi 10 aşamada şu şekilde tanımlanmıştır:

1. Hazırlıkların yapılması
2. Salgın olup olmadığının belirlenmesi
3. Tanının kesinleştirilmesi

¹ World Health Organization: http://www.who.int/topics/disease_outbreaks/en/

² Selçuk Tıp Fakültesi Salgın İnceleme Talimatı: <http://193.255.249.85/enfeksiyon/talimat7.html>

4. Vakaların belirlenmesi ve tanımlanması
5. Verilerin zaman, mekân ve kişilere göre tanımlanması ve yönlendirilmesi
6. Hipotezin geliştirilmesi
7. Hipotezin test edilmesi
8. Hipotezin iyileştirilmesi ve ek çalışmaların yapılması
9. Denetleme önlemlerinin geliştirilmesi
10. İlgililerin bilgilendirilmesi

Bu aşamalar ayrıntılı olarak *CDC* tarafından açıklanmıştır³.

Salgınlar, tarihte pek çok kez hem çok büyük can kayıplarına hem de büyük harcamalara sebep olmuştur. 14.yy'da başlayan ve 17.yy'a kadar süren kara veba 100 milyonların ölümüne sebep olmuştur. 19.yy başlarındaki ilk kolera salgını o dönemde yüz binlerce insanı öldürürken, halen her yıl binlerce insan bu hastalık yüzünden hayatını kaybetmektedir. 20.yy başlarında ortaya çıkan İspanyol gribi de 20 ila 50 milyon insanın ölümüne sebep olmuştur. Aynı hastalığa yakalananların sayısı 500 milyonlarla ifade edilmektedir⁴. O dönemde dünya nüfusunun 2 milyarın altında olduğu düşünülürse, bir salgının zamanında önlemi alınmadığında nelere yol açacağını gösteren en çarpıcı örneklerindendir. Son dönemde çıkan salgınlardan en tehlikelisi olan AIDS ise son 30 yılda 25 milyon insanın hayatına mal olmuştur. Maddi kayıplara verilebilecek en son örneklerden olan SARS, sadece Hong Kong'da 10 milyarlarca dolar maddi kayba sebep olmuştur [1]. Salgınlar yalnızca sağlık giderine yol açmamakta, turizm gibi yan gelirlerin de kaybedilmesine sebep olmaktadır.

1.2. Sendrom Sürveyansı

Sendrom sürveyansı sistemleri günümüzde, tanı ya da öntanı verilerinden sağlanan zaman serilerini salgın tespiti için kullanan erken uyarı sistemleri olarak özelleşmektedir [2]. Tanı verileri hastanelerden elde edilirken, öntanı verileri ilaç satışları, internet aramaları, iş veya okul devamsızlıkları gibi daha farklı kaynaklardan sağlanan verilerdir.

11 Eylül 2001'de New York'taki ikiz kulelere yapılan saldırılarının ardından

³ <http://www.cdc.gov/excite/classroom/outbreak/steps.htm>

⁴ http://en.wikipedia.org/wiki/1918_flu_pandemic

gerçekleşmesinden en çok korkulan şeylerin başında biyolojik terörist saldırılar gelmektedir. 2001'den önce sayıları daha seyrek olan ve birkaç araştırmaya konu olmaktan başka üzerinde durulmayan erken uyarı sistemleri, bu tarihten itibaren biyolojik teröre odaklanarak yaygınlaşmaktadır. Her ne kadar odaklandığı nokta biyolojik terör olsa da, doğal salgınlar da bu sistemler tarafından aynı yöntemlerle tespit edilmeye çalışılmaktadır. Küresel dünyanın hareketliliği, domuz gribi, SARS gibi yakın zamanın salgınlarındaki tecrübeler de gösteriyor ki, doğal yollarla oluşan ve yayılan salgınlar da hem büyük can kayıplarına hem de ciddi maddi kayıplara yol açmaktadır. Bu açıdan bakıldığında, sendrom sürveyansı sistemleri hem insan sağlığını hem de ülkelerin ekonomilerini doğrudan etkileyen bir sürece müdahil olmaktadır.

Sendrom sürveyansı için henüz kesin bir görev tanımı ve kesin bir terminoloji oluşmadığı için farklı kaynaklarda farklı isimlerle anıldığı gözlemlenmiştir. Halk sağlığını ilgilendiren bu çalışma alanının genelde erken bir salgın uyarı algoritmaları ve bu algoritmalara kaynak olan veri toplama/gözlem (sürveyans) üzerinde yoğunlaştığını ifade eden isimlendirmelerden bazıları "*Early Warning System*" [3], "*Outbreak Detection System*" [4], "*Biosurveillance System*" [2]'dir. Bunlardan başka isimlendirmeler de mevcuttur [5].

Bir önceki kesimde(1.1) belirtilen 10 maddelik salgın inceleme adımlarını da göz önünde bulundurursak, ideal bir sendrom sürveyansı sisteminin üç aşamadan oluştuğunu söyleyebiliriz. Bunlardan birincisi gözlemlerin toplandığı aşamadır. Tüm veri kaynaklarının doğru olarak ve zamanında ilgili birimlere bildirilmesi sağlanmalıdır. Yanlış tarihli ya da iyi organize edilmemiş veriler bundan sonraki aşamaların doğru işlememesine sebep olacaktır. İkinci aşamada toplanan verilerin çözümlenerek, eğer varsa salgın uyarısı oluşturulması safhasıdır. Bu aşamada elde edilen başarı iki şeye bağlıdır; ilk aşamanın başarımı ve kullanılan salgın tespit yönteminin başarımı. Sendrom sürveyansının son aşaması ise, ilgili birimlerin salgın konusunda uyarılarak daha önce tanımlanmış olması gereken eylem planının devreye alınmasını sağlamaktır.

Başta Amerika Birleşik Devletleri (ABD) olmak üzere, İngiltere, Hollanda, Kanada, Avustralya gibi pek çok ülke sendrom sürveyansına titizlikle yaklaşmaktadır. Bu amaçla yıllardır toplanan veriler üzerinden çalışmalar yapılmakta, gerçek zamanlı

ya da gerçek zamanlıya yakın sistemler geliştirilmektedir. Ülkemizde ise henüz salgın araştırmaları hastaneler düzeyinde kalmakta, aylık raporlarla il sağlık müdürlükleri bilgilendirilmekte ve sadece geçmiş aya ait salgın durum raporu sunulmaktadır. Gerçek zamanlı herhangi bir sistemin varlığından ise bahsetmek mümkün değildir. 2007 yılında raporlanan bir çalışma [6] ülkemizin o tarihte etkin kullanılacak bir sürveyans sisteminden ve eylem planlarından yoksun olduğunu ifade etmektedir. Bir yıl sonra hazırlanan ve taslak metnine ulaşılabilen *Erken Uyarı, Salgın Araştırma ve Kontrol Sistemi*⁵ yapılması önerilmiştir.

1.3. Veri Kaynakları

Geleneksel sürveyans sistemlerinde, sebebi belli ölümler ve belli başlı laboratuvar sonuçlarının günlük ya da haftalık olarak sayılarının tutulması öngörülmektedir. Ne var ki, bu verilerin toplanması, çözümlenmesi ve bir sonuca varılması, salgının tespit edilmesi gereken dönemden günler, haftalar hatta aylar sonrasına denk gelmektedir. Modern sistemlerde ise sendrom sürveyansının en önemli hedefi, doğal sebeplerden ya da biyolojik terör saldırılarından kaynaklanan salgınları en hızlı şekilde tespit etmektir. Bu amaca yönelik olarak, ikame veri olarak kullanılacak olan ilaç satışlarındaki artışlar, okul devamsızlıklarının artması, internet aramalarındaki bölgelere göre oluşan eğilimler salgın tespiti için kullanılabilir [7]. Her ne kadar doğrudan hastane kayıtlarında sağlanabilecek vaka sayısına ulaşmamızı sağlayamasa da, hasta tecrübeleri gösteriyor ki, bu ikame veri kaynakları hastalık yoğunluğu ile ilgili ön bilgi olabilmektedir. Bu veri kaynakların göreceli olarak daha erken oluşması erken teşhis için de ayrı bir önem kazanmaktadır. Normal koşullarda, bir hastanın önce kendini tedavi etme yoluna gideceği ve bunun için ilaç tüketeceği ya da internette diğer çözümleri araştıracağı varsayılabilir.

ABD'de yer alan veri toplama merkezleri, hastane tabanlı, şehir tabanlı ya da bölgesel tabanlı, coğrafi konum bilgilerini de ele alabilecek bir altyapı ile doğru kanuni düzenlemelerle çalışmaktadır. Aynen ülkemizde olduğu gibi [8], veriler faturalandırma kaygılarından ötürü Hastalıkların Uluslararası Sınıflandırılması (International Classification of Disease) ICD-10 kodları girilerek merkeze iletilmektedir. Ancak bu verilerin ne kadar doğru girildiği ve güvenilir olduğu

⁵ http://www.istanbul.edu.tr/duyurular/yok_duyurulari/bulasici_hastaliklar.pdf

konusunda herhangi bir çalışma bulunmamaktadır.

Salgın tespitinde kullanılan veriler zaman içinde farklı istatistiklere sebep olacak şekilde oluşmaktadır. Ortalama ve varyans gibi değerler de buna bağlı olarak zaman içinde değiştiği gibi, zaman serisinin kaynağına bağlı olarak bu değişimin derecesi değişiklik gösterir. Ancak salgın için kullanılan zaman serilerinde varlığı öngörülen bazı örüntülere rastlamak mümkündür. Hafta sonu ile hafta içi günleri arasında zaman serisinde dalgalanma meydana gelmesi bunlara verilebilecek en bariz örneklerdendir [9]. Poliklinik hizmetlerinin hafta sonlarında işlemediği düşünüldüğünde, birikimli olarak hafta başında hastane ziyaretçileri ve doğal olarak vaka sayısı artış gösterebilir. Goldenberg yaptığı çalışmada hafta sonları için ilaç satışlarının arttığını gözlemlemiştir [10]. Dolayısıyla, zaman serilerinin, aynı amaç için kaydedilse bile hangi koşulları kapsadığı önemli farklara sebep olabilmektedir. Haftanın gününe benzer bir etki de, tatiller ve tatil dönüşleri için söz konusu olabilmektedir [11]. Mevsimsel etkiler, hava sıcaklığına bağlı değişimler gibi öngörülerin de gözlemlendiği [7] göz önünde bulundurulduğunda, salgın başlangıcını tespit etmenin oldukça karmaşık olduğu düşünülebilir. Tüm bu karmaşıklığa, eksik veya tekrarlı veriler, gecikmeli kayıt ihtimali ve kodlama hataları gibi ihtimaller dâhil edilirse istatistiksel yöntemlerle doğru sonucun nasıl elde edileceği sorusu ile karşı karşıya kalmaktayız.

Salgın tespiti hızla önemi artan bir araştırma alanı olmasına rağmen, önündeki en büyük engel, kullanılan zaman serilerinin, her araştırmacının ulaşımına açık olmayıp, yalnızca sürveyans ekibinde yer alan araştırmacılar tarafından kullanılabilen veriler olmasıdır [12]. Erişim imkânı bulunan verilerin ise tespit yöntemlerinin idealleştirilmesi için kullanılacak öğrenmeyi ve doğrulamayı karşılaması oldukça güçtür. Bu gibi durumlarda kullanılmak üzere, veri benzetimlerine başvurulması gerekmektedir. Örneğin CDC tarafından geliştirilen *EARS* sisteminde test edilmek için üretilmiş 56 adet veri benzetimi, kullanıma açık bir biçimde sistemin web sayfasından yayınlanmaktadır⁶. Bu veriler, farklı şiddette, uzunlukta ve büyüklükte salgınları içerecek şekilde yaratılmışlardır. Buckeridge araştırmasında salgın tespit algoritmaları değerlendirirken üç çeşit veri kümesi olduğunu ifade etmektedir [13]. Bunlar tam-gerçek, tam-benzetim ve gerçek veriye

⁶ <http://www.bt.cdc.gov/surveillance/ears/datasets.asp>

salgın eklenmiş veri kümeleridir. Tamamen gerçek veri elde etmek oldukça güçtür. Çünkü bir salgın dönemini belirlemek için hem epidemiyolojiden hem de istatistikten uzmanların birlikte çalışarak vardıkları ortak sonuçlar gerekmektedir. *The Defense Advanced Research Projects Agency (DARPA)* tarafından desteklenen *BioALIRT* programı ABD'nin çeşitli şehirlerinden sağlanan tam-gerçek veri sunulmaktadır. Yukarıda bahsettiğimiz *EARS*, tam-benzetim veri kümelerini erişime açmıştır. Gerçek veriye salgın ekleme yöntemi ise diğer iki yönteme göre daha yaygın olarak kullanılmaktadır [9] [14]. Bu yöntemle üretilen veriler, salgın barındırmadığı düşünülen veri kümelerine değişik ölçeklerde yapay salgınlar eklenerek geliştirilir. Belirtilen yöntemler genel itibari ile tek değişkenlidir. Lotze ise geliştirdiği bir uygulama ile çok değişkenli, birbirlerini etkileyen veri kümeleri yaratan bir çalışma yayınlamıştır [12]. Bu çalışmalarda yer alan ortak örüntüler ise haftanın günü (Day of Week – *DOW*) etkisi, mevsimsel etki, tatil ve tatil sonrası etkilerdir. Tüm bu benzer çalışmaların yanı sıra, 1994 ile 2003 yılları arasında toplanan veriler üzerine yapay salgınlar yerleştiren Burr, bu işi yaparken *sıradüzenli ve sıradüzensiz* olarak adlandırdığı yöntemleri kullanmıştır [15]. Sıradüzenli yöntem, her yılın mevsimsel örüntünün sabit etki yarattığını varsayarken, sıradüzensiz yöntem bunu reddetmektedir. Burr'un çalışmasından elde edilen sonuçlar sıradüzensiz yöntemin gerçek verilerle daha uyumlu ve başarılı olduğunu göstermektedir.

Benzetilmiş veri kullanmak her ne kadar bir çıkar yol olarak gözükse de, araştırma üzerinde bazı sınırlandırmalara sebep olacağı da muhakkaktır. Mandl çalışmasında bu sınırlandırmaları dört maddede açıklamalı olarak ifade etmiştir [16].

1.4. Mevcut Sistem Örnekleri

Sendrom sürveyansı sistemleri, ABD, Kanada, Hollanda gibi gelişmiş ülkelerde hızla yaygınlaştığı gibi, AB, *WHO*, *OECD* gibi uluslararası etkinliği bulunan kuruluşların bildirimleri ile hem yaygınlaşmakta hem de merkezileştirilmeye çalışılmaktadır [6]. Uluslararası ticari ve turizm ilişkileri göz önünde bulundurulduğunda, bir bölgede bulunan salgın riskinin, halkının o bölgeye seyahat yoğunluğu bulunan yönetimler tarafından bilinmesinin arzulanması gayet

doğaldır. Bazı sistemler yoğun olarak kullanılıyor olmasına karşın henüz standartlaşmış bir kullanım veya sistem söz konusu değildir.

Yoğun olarak kullanılan sistemlerden bazılarında aşağıda değinilmektedir. Değinilmeyen bazı sistemler ve bu sistemlerin daha ayrıntılı araştırması, Mukhi'nin çalışmasında mevcuttur [17].

1.4.1. Electronic Surveillance System for the Early Notification of Community-Based Epidemics (ESSENCE)

ESSENCE, Department of Defence Global Emerging Infections System ile John Hopkins University Applied Physics Laboratory ortak çalışması ile ortaya çıkmış bir sistemdir [18]. Sistemin öncelikli hedefi, Washington D.C. bölgesindeki biyolojik terör saldırılarından kaynaklanabilecek bir salgın durumunun erken tespit edilebilmesidir. Bu amaçla, özel klinikler ile pek çok acil servisten 1999 Aralık ayından bu yana elektronik olarak günlük veri toplamaktadır. Toplanan veriler filtrelenerek (Şekil 1-2) işlemlerin sonuçları raporlanabilmektedir (Şekil 1-3). *ESSENCE* uygulaması geliştirilerek *ESSENCE II* sürümü ile varlığını sürdürmektedir. Birinci sürümden farklı olarak, *ESSENCE II* ICD kodlarının yanında, erken teşhisi kolaylaştıracak klinik dışı veri kaynaklarını kullanmaktadır. *ESSENCE II* doğal dil işleme ile metin olarak ele alınan hastalık belirtilerinin sendromlar halinde gruplandırılması yapılarak işletilmektedir. *ESSENCE* yalnızca zamansal tespit yöntemlerini kullanılırken, *ESSENCE II* ile hem zamansal hem de mekansal tespit yöntemleri kullanılmaktadır [17].

Data Configuration			
Data Source:	ER by Patient	Region:	Osage, MO OTHER_REGION Ozark, MO Perniscot, MO Perry, MO
Syndrome:	Hemr_ill Neuro Other Rash Resp	MedicalSubGrouping:	All MedicalSubGroupings AcuteBronchitis Bronchitis CongestionChest CongestionNasal
Detector:	Regression/EWMA 1.2		
Age Range:	Unknown 00-04 05-17 18-44 45-64	Sex:	All Sexes Unknown Male Female
Start Date:	10 Jul 09	End Date:	08 Oct 09
Change Configuration		Adv Qry	

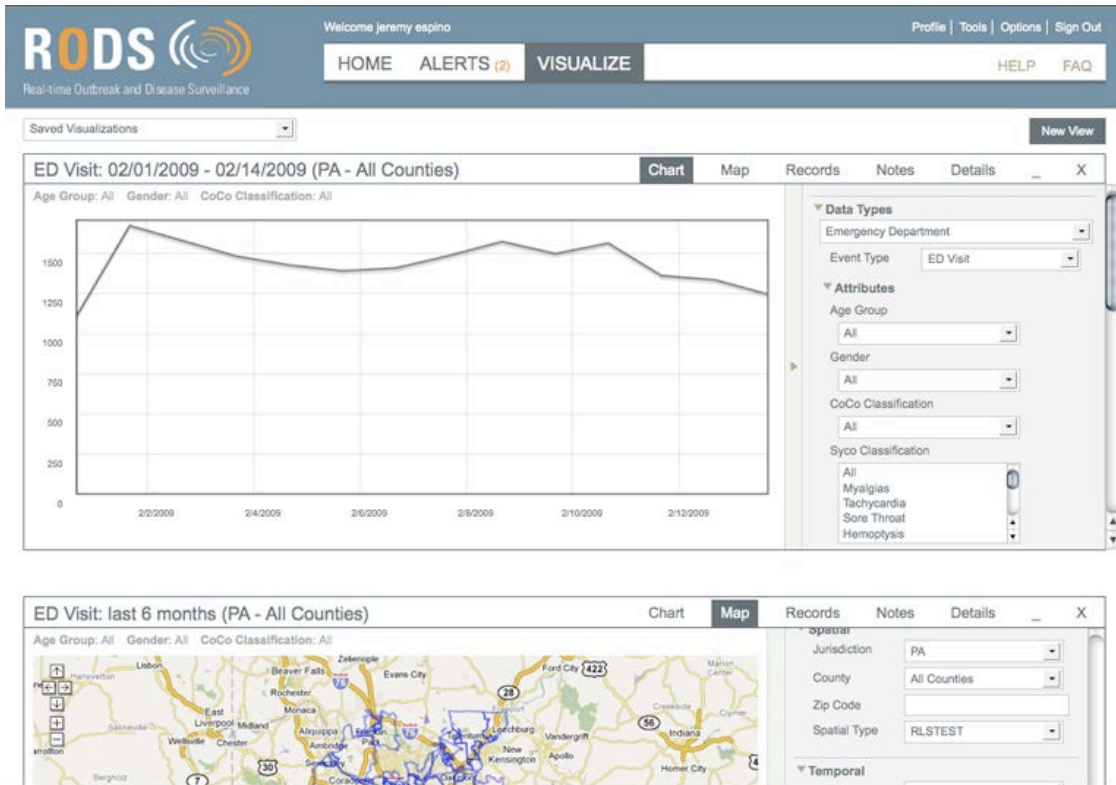
Şekil 1-1 ESSENCE'den veri filtreleme ekranı örneği

Create Event			
Datasource:	Real Time ER Data By Hospital Emergency Room Data by Patient Location Emergency Room Data by Hospital Location Percentage ER Data by Hospital Location NRDM	Status:	Open Closed
Category:	General Multi-Syndrome Bot_Like Fever GI	Age:	00-04 05-17 18-44 45-64 65+
Rank:	Info Investigating Monitoring Not Concerned Responding	Medical Grouping:	Bot_like Fever GI Hemr_ill Neuro
Geography:	Adair, MO Andrew, MO Atchison, MO Audrain, MO BOND, IL		
Start Date:	09Oct09		
End Date:	09Oct09		
Event Title:			
Message:			
<input type="checkbox"/> This message does not contain sensitive data.			
Preview			
Cancel			

Şekil 1-2 ESSENCE raporlama ve paylaşma ekranı örneği

1.4.2. Real-Time Outbreak and Disease Surveillance (RODS)

RODS, Pittsburgh Üniversitesi tarafından 1999'da şarbon salgınına gözlemlemek amacıyla geliştirilmiştir [19]. Java ile geliştirilmiş olan RODS'a açık kaynak olarak erişim mümkündür⁷. Hali hazırda 6.1. sürümü kullanılmakta ve 6.2. sürümü geliştirilmektedir⁸. RODS günümüzde yüzlerce halk sağlığı birimi tarafından hem ABD'de hem uluslararası başka yerlerde kullanılmaktadır [2]. Veriler ESSENCE'te olduğu gibi, geniş bir coğrafyadan toplanarak gerçeğe yakın zamanlı bir şekilde çözümlenmektedir (Şekil 1-3).



Şekil 1-3 RODS 6'dan bir kullanıcı arayüzü

1.4.3. BioSense

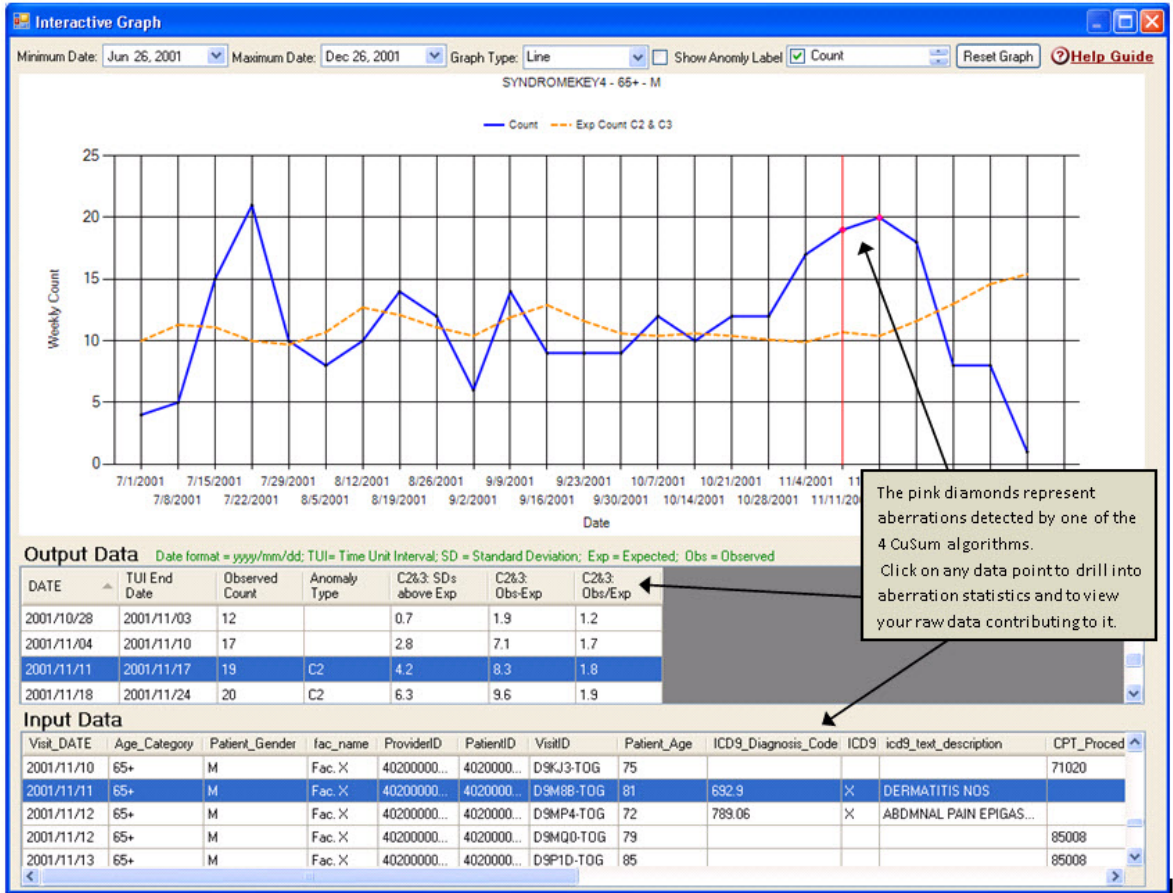
EARS, 11 Eylül saldırılarından sonra New York ve ABD'nin ulusal belli başlı bölgelerinde biyolojik terörü gözlemlemek, geniş bir alanı gözetim altında tutmak için CDC tarafından geliştirilmiş bir takip sistemidir [20]. EARS, kaynak verinin uzunluğuna ve seçilen yöntemin gereksinimlerine göre salgın tespit yöntemlerini üç kategoride toplamıştır [17]. EARS'ın bir başka özelliği de, daha önce de

⁷ <https://www.rods.pitt.edu/site/content/view/15/36/>

⁸ <http://betaweb.rods.pitt.edu/wiki/index.php/Changelog>

değinildiği gibi, önermiş olduğu C1, C2 ve C3 yöntemlerinin diğer yayınlarda referans seçilerek karşılaştırmalı araştırmalarda sıkça yer almasıdır.

BioSense, 2003 yılında CDC tarafından geliştirilmiş olan ve *EARS*'ı da kapsayan bir sistemdir. *BioSense* ile CDC, halk sağlığından sorumlu görevlilerin kendi halkının sağlığına katkıda bulunmayı hedeflemektedir⁹. *BioSense* geliştirdiği ikinci sürümü olan *BioSense 2*'yi 2011 Kasım itibari ile kullanmaya başlamıştır.



Şekil 1-4 EARS v5'e ait etkileşim ekranı örneği

⁹ <http://www.cdc.gov/biosense/>

2. SALGIN TESPİTİ İÇİN KULLANILAN İSTATİSTİKSEL YÖNTEMLER

Denetim grafikleri (*Shewhart* grafikleri) ilk olarak 1920'lerin başlarında Walter Shewhart tarafından kalite denetimi amacıyla ortaya atılmış [21] ve İstatistiksel Süreç Denetimi (Statistical Process Control - SPC)'nin temelini oluşturmuştur. Bu tarihten sonra, istatistiksel süreç denetimleri, başta üretim olmak üzere pek çok alanda yaygın olarak kullanılmaya başlanmıştır. Öyle ki, 2. Dünya Savaşı'nın ardından, Japonya Dr. Deming'in önerdiği kalite denetim süreçlerini uygulayarak hızla ayağa kalkmıştır. Deming'in felsefesi ise, üretkenlik ve kaliteyi artırırken, değişkenliği azaltmak ve bu değişkenliği anlayarak denetim altında tutmak üzerinedir [22].

Denetim grafikleri iki yönlü olarak uyarı üretir. Birincisi alt sınırın altına düşme durumu (Lower Control Limit – LCL), ikincisi ise üst sınırın üstüne çıkma durumudur (Upper Control Limit – UCL). Biyosürveyans için, yalnızca UCL eşiği değerli olmaktadır. Kesim 1.1'de verilen tanım gereği, grafikteki düşüşler değil yalnızca yükselişlerin eşik değeri aşp aşmadığı değer kazanmaktadır.

Temelde üç farklı denetim grafiği vardır. Bunlar *Shewhart*, birikimli toplam (*Cusum*) ve üstel ağırlıklı hareketli ortalamadır (*EWMA*). *SPC* literatüründeki bu yöntemlerin yanı sıra, salgın tespitinde kullanılan birçok yöntem ortaya atılmıştır. Bunların en çok dikkat çeken ve üzerine konuşulanı *EARS* (*Early Aberration Reporting System*) içerisinde yer alan C1, C2 ve C3 modelleridir. Bu üç model de yine esas olarak, *SPC*'yi temel almaktadır. Bu üç modelin yanı sıra W2 diye adlandırılan ve duruma göre C2'yi ya da C3'ü seçen bir model de tanımlanmıştır. C1 ile C2 neredeyse tamamen aynıdır. Tek farkları C2 modelinin, kullandığı kayan pencereyi iki gün geriden başlatarak salgının yavaş yayılarak gözden kaçmasını engellemeyi hedeflemesidir. C3 modeli ise diğer iki modelden farklı bir yöntem kullanmaktadır. Kesim 2.4.'te bu modeller daha ayrıntılı olarak ele alınacaktır.

SPC yöntemleri ile *EARS* modelleri dışında, Negatif İkiterimli Birikimli Toplam (*Negative Binomial CuSum - NBC*) [23], *HLM* (*Historical Limits Method*) [24] ve başka değişik algoritmalar uygulanmıştır. Her ne kadar *NBC* ve *HLM* gibi başka yöntemler ortaya sürülmüş olsa da, *SPC* yakın tarihten itibaren salgın tespiti için bir istisnadan ziyade bir standart yöntem olarak görülmeye başlanmıştır [25].

2.1. Cusum

Cusum, Shewhart'ın ardından süreç yönetimi için Page'in önerdiği bir yöntemdir [26]. Page, kalite denetim sürecinde, kusurlu ürünleri denetim altında tutmayı amaçlamıştır. Cusum Shewhart'ın aksine [27], süreçte karşılaşılan küçük kaymaları tespit etmek amacıyla kullanılır. Cusum yalnızca o anlık gözlemi değil, aynı zamanda yakın geçmişin gözlemlerini de hesaba kattığı için küçük değişikliklerin tespiti mümkün olmaktadır. Ancak Cusum denetim grafiklerinin de önemli bir eksiği vardır. Gözlemlerin dönemsel dalgalanmalar göstermesi durumunda süreç hakkında isabetli karar vermekte yetersiz kalabilmektedir [28].

$$CS_t = \sum_{j=1}^n (\bar{x}_j - \mu_0) \quad (2.1)$$

$CS_t = t$ anındaki *CuSum* sonucu

$\bar{x}_j = j$ 'ninci gözlemin ortalaması

$\mu_0 =$ hedef ortalama

$n =$ gözlem sayısı

Eşitlik (2.1) incelendiğinde, birikimli toplam işleminin nasıl gerçekleştiği anlaşılmaktadır. Eşitliği açarsak:

$$CS_1 = x_1 - \mu_0$$

$$CS_2 = (x_1 - \mu_0) + (x_2 - \mu_0) = (x_2 - \mu_0) + CS_1$$

⋮

$$CS_t = \sum_{j=1}^n \bar{x}_j - \mu_0$$

sonucuna varırız. Bu durumda $CS_t = (x_t - \mu_0) + CS_{t-1}$ eşitliğini elde ederiz.

$$CS_t = \max[0, (x_t - \mu_0) + CS_{t-1}] \quad (2.2)$$

Biyosürveyans kapsamında yalnızca artışlar ele alındığı için *Cusum* sonucunda elde edilen eksi değerler anlamsız olduğu için 0'ın altına düşmesinin önlenmesi gerekir [29]. Eşitlik (2.2) bunu sağlamak için önerilen *Cusum* hesaplama yöntemidir. Eşitlik (2.1) ve (2.2) ile yapılan hesaplamalar bir denetim grafiği elde etmek için kullanılan eşitliklerdir. Yapılan hesaplamalardan bir aykırı durum (sürveyans sistemlerinde bu aykırı durum salgındır) oluştuğunu anlamak için belirlenen eşik değeri aşıp aşmadığı denetlenir.

$$CS_t^+ = \text{maks}[0, x_t - (\mu_0 + k) + CS_{t-1}] \quad (2.3)$$

Eşitlik (2.2)'den farklı olarak eşitlik (2.3)'te kaynak değer kullanılmıştır. Kaynak değer, *Cusum*'ün o anki değerine dâhil edilmeden, sadece eşik değere ulaştığının test edilmesinde kullanıldığı için grafik eşitliklerinde yer almamaktadır. k değeri genel olarak, ortalamadaki kaymanın ortası olarak belirlenir [30]. Başka bir deyişle $k = \sigma/2$ olarak belirlenir. Ancak bu en iyi k değerinin belirlenmesi biyosürveyans araştırmalarında herhangi bir sabit hesaplama yapılmamıştır.

2.2. EWMA

Üstel Ağırlıklı Hareketli Ortalama (*Exponentially Weighted Moving Average-EWMA*) denetim grafikleri, ilk olarak Robert'in çalışmasında *Geometrik Hareketli Ortalama* adıyla ortaya atılmıştır [31]. Ortaya atıldıktan sonra birkaç araştırma dışında [32] [33], kalite-denetim çözümleyicileri tarafından pek tercih edilmemiştir. Ancak 1980'lerin ikinci yarısından itibaren faydalanıp analitik olarak değerlendirilmeye başlanmıştır [34] [35] [36]. Süreçte gerçekleşen küçük değişimlerin tespitinde başarılıdır [36]. Shewhart ve *Cusum* denetim grafikleri gibi gerçekleştirilmesi oldukça kolay bir denetim grafiğidir. Aynı zamanda bir düzleştirme yöntemi de olan *EWMA*'nın her iki işlemin benzerlikleri kesim 3.3.3'te görülebilir.

$$E_t = \alpha x_t + (1 - \alpha)E_{t-1} \quad (2.4)$$

$$\alpha = \text{ağırlık çarpanı} \quad 0 < \alpha \leq 1$$

Eşitlik(2.4)'te yer alan gözlem sayısı, t anına ait gözlem sayısı olabileceği gibi, belirlenen boyuttaki bir döneme ait gözlemlerin ortalaması da olabilir. Eşitlik (2.4)'te bir başlangıç sorunu olduğu görülmektedir. Çünkü özyineli ilerleyen bir eşitliktir ve başlangıç değerlerinin doğru seçilmesi gerekmektedir. *EWMA* için ilkendirme için *hedef* değer seçilmektedir [37]. Hedef değer, gözlemlerin ortalaması olarak hesaplanır.

$$E_0 = \mu$$

$$E_1 = \alpha x_1 + (1 - \alpha)E_0$$

$$E_2 = \alpha x_2 + (1 - \alpha)(\alpha x_1 + (1 - \alpha)E_0)$$

⋮

$$E_i = \alpha \sum_{j=0}^{i-1} (1 - \alpha)^j x_{i-j} + (1 - \alpha)^i E_0 \quad (2.5)$$

Eşitlik (2.5) ile *EWMA*'nın geometrik açılıma dönüştüğünü görüyoruz. E_0 'nın etkisinin ise üstel olarak gerilediğini ve belirlenen α büyüklüğüne göre zamanla etkisinin 0'a yaklaştığını görebiliriz.

$$\sigma_{ewma}^2 = \frac{\alpha}{2-\alpha} \sigma^2 \quad (2.6)$$

$$E_t^+ = E_0 + k\sigma_{ewma} \quad (2.7)$$

Eğer hesaplanan *EWMA* değeri hedef değerden k kez uzaklaşırsa, eşik değerini aşmış ve aykırı durum tespit edilmiş olur. Biyosürveyans kapsamında yalnızca artışları incelediğimiz için yukarı yönlü hareketleri denetliyoruz. Eğer gözlenen serideki düşüşleri de ele alacak olsaydık $E_t^- = E_0 - k\sigma_{ewma}$ eşitliğini de denetlememiz gerekecekti. k seçimiyle ilgili ayrıntılı bir çizelgeyi Lucas'ın çalışmasında bulabilirsiniz [37].

$$E_t = \text{maks}[\mu, \alpha x_t + (1 - \alpha)E_{t-1}] \quad (2.8)$$

Eşitlik (2.8) biyosürveyans kapsamında ele alınması için Fricker'ın ders notlarında belirtilen bir yöntemdir. *Cusum*'da olduğu gibi, gözlem miktarının 0'ın altına düşmesini engellemesi sağlanabilir. Ancak eğer bir tahmin yöntemi ile düzleştirilmiş bir veri kaynağı varsa, eşitliğin eşitlik (2.9)'da gösterildiği gibi ele alınması gerekir.

$$E_t = \text{maks}[0, \alpha x_t + (1 - \alpha)E_{t-1}] \quad (2.9)$$

2.3. Shewhart

Shewhart, *SPC* yöntemlerinin başlangıç noktası olarak karşımıza çıkmaktadır. Endüstriyel amaçlarla ortaya koyduğu, üretim sürecini ölçmeyi hedeflediği yöntem, zaman içinde belki de en çok ele alınan yöntem olmuş ve hiçbir zaman değerini yitirmemiştir. Biyosürveyans çalışmalarında da Shewhart denetim grafikleri sık ele alınmış, hatta kesim 2.4.'te ele alındığı üzere, *EARS* sistemindeki modellere temel oluşturmuştur.

$$S_t = \frac{x_t - \mu}{\sigma} \quad (2.10)$$

μ ve σ hesabı yapılırken aykırı durum içermeyen dağılıma sahip geçmiş verilerden yararlanılır. Bu sayede gözlemin sorunsuz bir dönemde ortalaması ve standart

sapması hesaplanmış olur. Shewhart yöntemi ile bir gözlemin ortalamadan ne kadar uzaklaştığına bakılır. Eğer bir gözlem, ortalamadan standart sapmanın belirlenen katından daha büyük bir değerle uzaklaşmışsa, Shewhart bu durumu tespit eder.

Endüstriyel uygulamalarda, iki Shewhart grafiği aynı anda işletilir. Bunlardan birincisi süreç ortalamasını gözlemlemek için, ikincisi ise süreçteki diğer değişimler içindir [38]. Shewhart'ın farklı uygulama alanları için modellenmesi ile ilgili daha ayrıntılı bilgi için bir çalışması mevcuttur [39].

Biyosürveyans sistemleri için, Shewhart'ın en önemli avantajı hızlı tepki verebilmesidir. Ne var ki, Shewhart ortalamadaki kaymanın küçük olmadığı durumlarda etkindir. Ani ve keskin yükselişleri tespit etmek için doğru bir yöntemdir.

2.4. EARS

Early Aberration Reporting System (EARS), Amerika'da bulunan *Centers for Disease Control and Prevention* tarafından, 11 Eylül 2001'de New York'ta gerçekleşen saldırıdan sonra biyolojik terör ihtimaline karşın oluşturulmuş, ulusal düzeyde gözlem yapan bir sistemdir [20]. *EARS* yöntemlerinin yöneldiği salgın tipi, biyolojik terörün ani yükselişlere sebep olacağı verilerdir. *EARS* yöntemleri veri kümesinin çok az geçmişe sahip olduğu veya hiçbir geçmiş veri içermeyen durumlar için önerilmiştir [5]. *EARS* tarafından kullanılan üç farklı yöntem vardır. Bunlar C1, C2 ve C3'tür. Bu üç yöntemin yanı sıra, hafta içi ile hafta sonunu ayrı ayrı ele alan W2 yöntemi vardır. Pek çok araştırmada *EARS* yöntemlerinin *Cusum* tabanlı olduğu belirtilmişse de [40] [41] Shewhart tabanlı yöntemlerdir [38].

EARS yöntemleri, biyosürveyans araştırmalarında en çok ele alınan yöntemlerdir. Bu yüzden de farklı yorumlar ile karşımıza çıkmaktadırlar. Yeni yöntemler öneren pek çok araştırmacı, kendi yöntemlerini *EARS* yöntemleri ile karşılaştırarak *EARS* yöntemlerinin aslında birer referans nokta olduğunu kabul etmiş olmaktadır.

$$C_1(t) = \frac{x_t - \bar{X}_t}{s_t} \quad (2.11)$$

$$\bar{X}_t = 7 \text{ günlük döneme ait gözlemlerin hareketli ortalaması}$$

$s_t = 7$ günlük döneme ait gözlemlerin standart sapması

Eşitlik (2.11) C1 istatistiğinin hesaplanması için kullanılan eşitliktir.

$$\bar{X}_t = \frac{1}{7} \sum_{i=t-1}^{t-7} x_i \quad (2.12)$$

$$s_t = \sqrt{\frac{1}{6} \sum_{i=t-1}^{t-7} [x_i - \bar{X}_t]^2} \quad (2.13)$$

Eşitlik (2.12) ve (2.13) ile C1 istatistiğinde kullanılan terimlerin hesaplamaları yapılmaktadır. C1 istatistiği t anında h eşiğini aşarsa, sistem uyarı verecektir. CDC, C1 eşik değerinin 3 olmasını önerirken Fricker'ın araştırmasında eşik değerini, karşılaştırılmanın mümkün olması için *ATFS(Average Time to False Signal)*'nin sabitlendiği farklı eşik değerleri önermiştir [42]. Eşitlik (2.11)'de standart sapma hesabı küçük gözlemler için 0'a çok yaklaşabilir ya da 0 olabilir. Bu durumlar için *SPC* yöntemlerindeki genel kullanım, standart sapmanın bir önceki değerini kullanmak ya da asgari bir standart sapma belirlemektir.

Bir başka *EARS* yöntemi olan C2'nin C1'den farkı, 2 günlük bir gecikme kullanmasıdır. C2 hem ortalama hesaplanırken hem de standart sapma hesaplanırken, son iki günü dâhil edilmeden, $\bar{X}_t = \frac{1}{7} \sum_{i=t-3}^{t-9} x_i$ ve $s_t = \sqrt{\frac{1}{6} \sum_{i=t-3}^{t-9} [x_i - \bar{X}_t]^2}$ şeklinde hesaplanır. Bu sayede, C2 yavaş artış gösteren serilerde meydana gelmesi mümkün olan manipülasyonun önüne geçilmesi hedeflenmiştir.

C3 yöntemi, C2 yönteminin t anına ait istatistik ile birlikte son 2 istatistiğini de kullanan bir denklem ile hesaplanır.

$$C_3(t) = \sum_{i=t}^{t-2} \max[0, C_2(i) - 1] \quad (2.14)$$

C3 istatistiğinin *EARS* içerisindeki kullanımında C1 ve C2'den ayrılan bir başka özelliği de eşik değerinin 2 olarak tanımlanmış olmasıdır.

C1, C2 ve C3 yöntemleri dışında, *BioSense* tarafından gerçekleştirilen W2 yöntemi vardır. W2 yöntemi, C2 yönteminin hafta içi ve hafta sonu için ayrı ayrı hesaplamalar yapan bir yöntemdir. W2 yöntemi ise kendi içinde iki farklı formda kullanılmaktadır. Bunlardan birincisi W2c, hafta içi ve hafta sonu için 2 günlük

gecikmeyi ve 7lik bir pencereyi kullanmaktadır. Diğer W2 yöntemi ise W2r'dir. W2r yönteminde, gözlem yapılan kaynakta karşılaşılan bir sendromun toplam ziyaretçi sayısına oranı kullanılarak hesaplamalar yapılmaktadır. W2 yöntemi ile ilgili daha ayrıntılı bilgi ve çözümler için Szarka'nın çalışması incelenebilir [43].

2.5. Diğer Yöntemler

Zaman serilerinde aykırı durum tespitinde kullanılan farklı yöntemlerin salgın tespitinde de uygulandığı görülmektedir. Bunlardan bazıları "*Hidden Markov Model*"(HMM), "*Negative Binomial Cusum*"(NBC), "*Historical Limits Method*"(HLM), "*Poisson Salgın Tespit Yöntemi* ve "*Temporal SaTScan*" çözümleridir.

HMM, çok etkin bir salgın tespit yöntemi olarak göze batan sonuçlar vermiş olmasına rağmen, bugüne değin çok tercih edilmemiştir [44]. HMM, posta kodları ile komşu ilişkisi kurarak, mekân etkisini de birleştirir. Bu yöntemi kullanabilmek ancak çok değişkenli zaman serilerinde mümkündür.

NBC, Cusum tabanlı bir salgın SPC yöntemidir. NBC'nin salgın tespitinde EARS yöntemlerinden daha etkin olduğu yönünde sonuçlara varan araştırmalar mevcuttur [40]. NBC, yanlış uyarıları asgari seviyelere indirgemeyi sağlayan bir yöntemdir. Bu yüzden de kurulan sistemin amacına göre kullanımı oldukça kullanışlı olabilir. NBC hesaplanırken iki parametre(r,c) üzerinden işlem yapılır. Beklenenin üzerindeki gözlemler c parametresi ile ele alınır.

$$c = \frac{\mu}{\mu - \sigma^2} \quad (2.15)$$

$$r = \frac{\mu^2}{\mu - \sigma^2} \quad (2.16)$$

Eşitliklerdeki ortalama ve varyans hesapları yapılırken, bir hareketli pencere boyu belirlenir. Bu zaman diliminde kaydedilen gözlemlerden ortalama ve varyans hesabı yapılarak r ve c hesapları yapılır. r değişkeninin eşitlik (2.16) da yapıldığını varsayarsak, c değişkeninin denetimdeki c_0 dan denetim dışındaki c_1 düzeyine yükseldiğini izlemek için Cusum karar aralığı hesabı için:

$$NBC_t = maks(0, NBC_{t-1} + x_t - k) \quad (2.17)$$

Eşitlik (2.17) eşitliği kullanılır. *NBC* için başlangıç değeri olarak 0 kullanılır. Eşitlikte yer alan k değeri *Cusum*'dan farklı olarak eşitlik (2.18)'de gösterildiği gibi hesaplanır:

$$k = r \cdot \ln \left[\frac{c_0(1+c_1)}{c_1(1+c_0)} \right] / \ln \left[\frac{1+c_0}{1+c_1} \right] \quad (2.18)$$

Denetim dışındaki c_1 'in, denetimden iki standart sapma kadar uzaklaşmasından elde edilen seviye olarak kullanılması önerilmiştir [40].

HLM [24], geçmiş gözlemleri *Cusum* yöntemleri aksine, mevsimsel etkiyi ele alabilecek şekilde tasarlanmıştır [45].

$$\frac{x_0}{\mu} > 1 + \frac{2\sigma_x}{\mu} \quad (2.19)$$

$x_0 = \text{Gözlem sayısı}$

μ ve $\sigma_x = \text{Tarihsel verinin ortalaması ve standart sapması}$

Eşitlik (2.19)'da kullanılan tarihsel veri, bir haftalık bir pencereyi ele alırsa, geçmiş yıllara ait bir hafta öncesinde, yaşanan haftasında ve takip eden haftasında, toplam 3 haftalık süreçleri hesaba katarak ortalama ve standart sapma hesaplanır. Eğer pencereyi 2 hafta olarak varsayarsak, toplam 5 haftalık bir pencere elde edilir. Ancak anlaşılacağı üzere, *HLM* ciddi bir veri geçmişine sahip olmayı gerektirir. Bu *EARS* ile kıyaslanmasının pek de doğru olmadığını ve iki yöntemin çok farklı durumlarda tercih edileceğini gösterir. Ne var ki, yapılan araştırma *HLM*'nin başarımının daha iyi olduğu sonucunu işaret etmektedir [45].

2.6. Başarım Değerlendirme Ölçütleri

Günlük toplanan verilerin çözümlendiği ve salgın durumlarının araştırıldığı bir araştırmanın neticesinde varılmak istenen sonuç, hangi yöntemin talep edilen zaman serisi üzerinden daha etkin olduğunun belirlenebilmesidir. Bu da ancak doğru tanımlanmış ölçüm yöntemleri ile gerçekleştirilir. Biyosürveyans araştırmalarında, en yaygın olarak kullanılan ve standartlaştığı öngörülen hesaplamalar "*sensitivity*", "*specificity*" ve "*timeliness*" ölçütleridir.

Sensitivity, bir yöntemin hangi oranda salgını tespit ettiğini ölçen bir ölçüttür. Doğru olarak tespit edilen salgın sayısının gerçekleşen toplam salgın vakasına oranı bu ölçütü verir. Ne kadar yüksek olursa o kadar başarılı bir sonuçtan söz edilir.

Specificity, bir yöntemin uyarı üretmediği durumlar için, bu durumun doğruluk oranını verir. Bu ölçüm ne kadar yüksekse, yöntemin başarımı bu ölçüt için o kadar başarılıdır.

Timeliness, bir yöntemin salgın dönemleri için ürettiği uyarıların ortalama zamanının ölçüldüğü ölçüttür. Yani gerçek bir salgın döneminde üretilen uyarının, salgının başlangıcından itibaren ne kadar süre geçtiği belirlenir ve tüm gecikmelerin ortalaması alınarak *timeliness* belirlenir. Fricker tarafından karşılaştırma ölçütü olarak tanımlanan [42] *ATFOS*(*Average Time to First Outbreak Signal*) ölçütü ile şekilde hesaplanır. Bu ölçüt için, düşük değerler yüksek başarımı gösterir.

Çizelge 2-1’de verilen ilişkiler kullanılarak hesaplanabilecek olan ölçütler:

$$sensitivity = \frac{DP}{DP+YN} \quad (2.20)$$

$$specificity = \frac{DN}{DN+YP} \quad (2.21)$$

$$Pozitif Kestirim Değeri = \frac{DP}{DP+YP} \quad (2.22)$$

$$Negatif Kestirim Değeri = \frac{DN}{DN+YN} \quad (2.23)$$

Çizelge 2-1 Gerçek sonuç ile üretilen sonuç ilişki çizelgesi

	Gerçek Sonuç Pozitif	Gerçek Sonuç Negatif
Test Sonucu Pozitif	DP	YP (1. Tür Hata)
Test Sonucu Negatif	YN (2. Tür Hata)	DN

Eşitlik (2.20),(2.21),(2.22) ve (2.23) çizelge 2-1’de yer verilen dört değer ile hesaplanabilmektedir. Bunların yanı sıra, 1. Tür Hata $1 - sensitivity$ ve 2. Tür Hata $1 - specificity$ ile hesaplanabilen değerlerdir. Çizelge 2-1’de yer alan değerlerin hesaplanmasına katkı sağlamayacağı bir ölçüt olan “*timeliness*” ise eşitlik (2.24) ile hesaplanır:

$$timeliness = \frac{t_1+t_2+\dots+t_n}{n} \quad (2.24)$$

Eşitlik (2.24)'te verilen t_i , tespit edilen salgın dönemine ait, salgının gerçek seride başladığı an ile yöntem tarafından tespit edildiği an arasında geçen süredir. Bir diğer anlamıyla salgın tespitine kadar geçen süreden meydana gelen gecikmedir. Bu gecikmelerin ortalaması “*timeliness*” ölçütünü vermektedir.

ARL(Average Run Length), SPC için gelenekselleşmiş bir ölçüm yöntemidir. Biyosürveyans araştırmaları için *ARL*, “*in-control ARL*” ve “*out-of-control ARL*” olarak iki ayrı biçimiyle kullanılmaktadır. “*in-control ARL*”, aykırılık içermeyen bir dönemde, yöntemin zaman serisi için ürettiği uyarılar arasında geçen ortalama süredir. “Average Time to False Signal”(ATFS), “*in-control ARL*” Fricker tarafından önerilen adıdır [42]. Aynı araştırmada “*out-of-control ARL*”, *ATFOS* adıyla anılmıştır. *ATFOS*, salgın gerçekleştiğinde, yöntemin zaman serisinde uyarı üretmesi için geçmesi beklenen süredir. Geleneksel ölçütlerden farklı olarak ortaya atılan bu ölçütler, aslında bahsi geçen yöntemler ile benzerlik göstermektedir. Ancak salgın, bir anda değil dönemde gerçekleşir. Tespit edilen salgın, yalnızca tespit edildiği anı değil, tespit edildiği salgın döneminin tamamını etkiler. Örneğin, “*sensitivity*” ölçütü salgın tespitinde başarıyı ölçen ölçüttür. Bir salgın döneminin 4. Gününde yapılan tespit, “*sensitivity*” ölçütünü 0.25'lik bir ölçüme götürür. Ancak salgının dönem içinde tespit edildiği düşünülürde “*sensitivity*” 1.0 olarak hesaplanmalıdır.

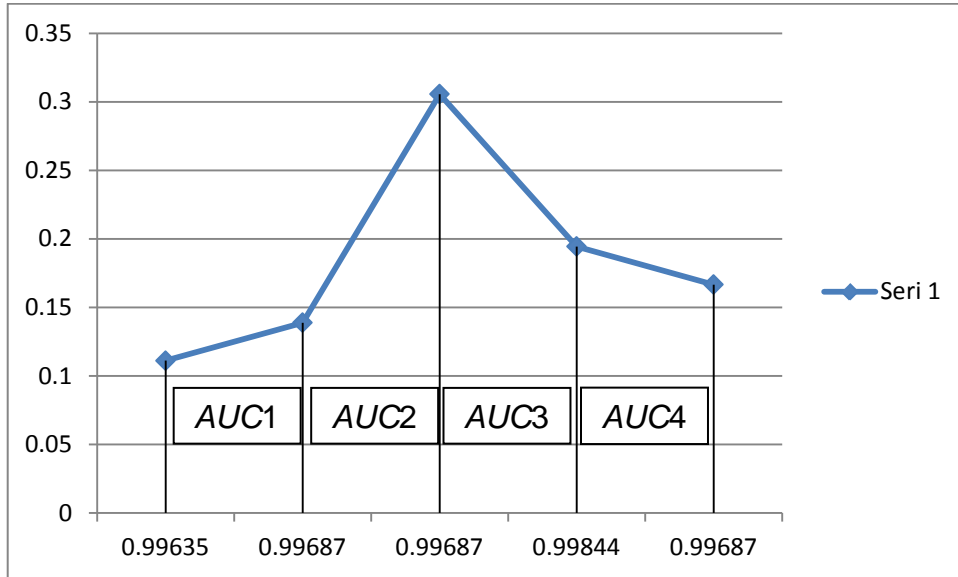
Görüldüğü gibi, bir yöntemin başarı ölçümü için birden çok ölçüt mevcuttur. Bu durumda, bu ölçütlerin karma olarak karşılaştırılması, bir ölçütün eşitlenerek, bu eşit durum için diğer ölçümlerden alınan sonuçların karşılaştırılması gibi başarı değerlendirilmesini daha nesnel hale getirecek yollar aranmalıdır. Bu amaçla Fricker, yöntemlerin *ATFS* değerlerini eşitleyen eşik değerlerini hesaplayarak, yaptığı diğer hesaplamaları çizimlerinde kullanmayı tercih etmiştir [42]. Birçok araştırma ise “Receiver Operating Characteristic”(ROC) eğrileri ile başarı çözümlemesini önermiştir [2] [7]. ROC eğrilerinde kullanılacak olan iki ölçüm “*sensitivity*” ve “*specificity*” olabilir. Klasik bir ROC eğrisinde, “*sensitivity*” ile “*false positive rate*” değerleri kullanılarak ROC eğrisi çizilir. Lotze, bu doğrultuda $1/ATFS$ 'yi ROC eğrilerinde kullanmıştır [2]. Aslında *AUC* hesaplamasında kullanılan eşitlik (2.25) için ROC eğrisinde *specificity* ya da *false positive rate* tercihi arasında bir fark yoktur. Yalnızca eğri eğimi eksi yönde değişir. ROC eğrileri

gecikmeleri doğrudan ele alamadığı için “Activity Monitoring Operating Characteristic”(AMOC) eğrilerinin kullanılması önerilir. AMOC eğrilerinde “timeliness” ve “false positive rate” ölçümleri kullanılabilir.

ROC eğrilerinden anlamlı sonuç elde edebilmek için, belirtilen aralıkta, eğrinin altında kalan alanın(AUC) hesaplanması beklenir. AUC, ROC eğrileri ile birlikte, bu tür kalite denetim yöntemlerinin standartlaşan bir algoritma ölçme yöntemleridir. AUC değeri ne kadar yüksek çıkarsa, kullanılan yöntemin belirtilen hata aralığı için o kadar başarımının yüksek olduğu anlaşılır. AUC hesaplanmasının farklı yöntemleri mevcuttur. Bu çalışma kapsamında AUC hesabı için kullanılacak olan yöntem “Trapezoidal Rule” integral alma tekniğidir. Bu teknikte, kesikli değerler içeren ROC eğrisi altında kalan alanın yamuklardan oluştuğu varsayılmıştır.

$$\int_a^b f(x)dx \approx (b - a) \frac{f(a)+f(b)}{2} \quad (2.25)$$

Eşitlik (2.25)'te verilen a ve b değerleri ROC eğrisindeki iki ardışık noktanın değerlerini ifade etmektedir. Her iki a, b ikili değeri arasında kalan alanlar toplanarak toplam AUC hesabı yapılmış olur. Şekil 2-1'de yer alan örnek grafik için toplam AUC için AUC1, AUC2, AUC3 ve AUC4'ün toplanması gerekir.



Şekil 2-1 AUC'nin Trapezoidal ile hesaplanması

3. VERİ ÖN İŞLEME YÖNTEMLERİ

Denetim grafikleri, zaman serilerindeki aykırılıkları ve özel durumları ortaya koymak için kullanılmaktadır. Salgınlar da zaman serilerinde ortaya çıkan özel durumlardır. Bu yüzden denetim grafiklerinin salgın tespiti için kullanılması oldukça olağandır. Ne var ki, denetim grafikleri bağımsız ve özdeşçe dağılmış (*independently distributed(iid)*) zaman serilerinde oldukça etkili bir süreç takibi sağlarken, bunu, varsayımları ihlal eden veriler için güvenilir şekilde üretmeleri mümkün olmamaktadır [46]. Başka bir deyişle, geleneksel denetim grafiklerinin sürveyans verilerinde de etkili başarımlar ortaya koyabilmesi verideki eğilim (*trend*), haftanın-günü etkisi (*day-of the week effect*), mevsimsel etki (*seasonal effect*) gibi açıklanabilir örüntülerin temizlenerek bağımsız ve eşit şekilde dağılmış hale getirilmesi gerekir [9].

Açıklanabilir örüntüler, hastalıkla ilişkisi olmayan ancak kayda geçen hastalık sayısını etkileyen sebeplerden oluşurlar [2]. Örneğin grip benzeri hastalıkları gözlemleyen bir sistemin, kış aylarında vaka sayısının arttığını göz önünde bulundurması ve bu artışı bir aykırı durum olarak algılamaması sağlanmalıdır. Aslında mevsimsel etki akla ilk gelen örüntülerden biri olarak kabul edilebilir. Sürveyans sistemlerinde ele alınan bir diğer örüntü kaynağı ise haftanın-günü etkisidir. Eğer bir bölgede hastaneler hafta sonu sadece acil servis hizmetleri veriyorsa, bu bölgeye ait hastanedeki Pazartesi gözlemlerinin Cumartesi ve Pazar'a göre çok daha fazla olması mümkün olabilir. Ayrıca eğilimler, tatiller, tatil dönüşleri gibi sistematik etkenler söz konusu olabilmektedir.

Zaman serileri üzerinde en çok yapılan işlemlerin başında tahmin (*forecast*) ve düzleştirme (*smoothing*) gelmektedir. Denetim grafiklerinin sürveyans verileri üzerinde doğru sonuçlar üretebilmesi için, bu verilerin düzleştirme yöntemleri ile işlenmesi kaçınılmazdır. Bu yöntemlerin başarımlarının değerlendirilmesi için Hata Kareler ortalaması (*MSE*), Hata Kareler ortalaması karekökü (*RMSE*), Mutlak Ortanca Yüzde Hatası (*MdAPE*) ve Ortanca Mutlak Sapma (*MdAD*) yöntemleri kullanılmıştır [7] [9].

Tahmin yöntemleri, uygulamada farklı gereksinimlere ihtiyaç duyarlar. Bazı tahmin yöntemleri eski verilerden çıkarım yaptıkları için uzun ya kısa süreli tarihe kadar

süren verilere ihtiyaç duyabilirken, doğrusal regresyon vb. bazıları başka veri kaynaklarına ihtiyaç duyabilirler.

3.1. Regresyon

Regresyon çözümlenmesi, bir bağımlı değişkenin, bir ya da daha çok bağımsız değişkenle olan ilişkisini ölçmek için kullanılan bir çözümlenme yöntemidir [47]. Başka bir söylemle, regresyon çözümlenmesi ile bağımsız bir değişkendeki değişimin bağımlı değişkeni nasıl etkilediğinin anlaşılması sağlanır. Pek çok regresyon çözümlenmesi tekniği geliştirilmiştir ancak standart bir gösterimle bir regresyon modeli:

$$Y \approx f(X, \beta) \quad (3.1)$$

Eşitlik (3.1) gibi gösterilir. Bu gösterimde, Y bağımlı değişkeni, X bağımsız değişkenleri ve β bilinmeyen parametreleri ifade etmektedir.

Regresyon modelleri ile zaman serilerinde tahmin yapmak oldukça yaygın bir şekilde kullanılmaktadır. Günlük tutulan sağlık serilerinde de tahmin için en çok kullanılan yöntem olduğu bilinmektedir [2].

3.1.1. Basit Doğrusal Regresyon

Basit doğrusal regresyon, bağımlı değişken ile yalnızca bir bağımsız değişken arasındaki doğrusal ilişkiyi anlamak için kullanılan bir doğrusal regresyon modelidir.

$$y = \beta_0 + \beta_1 x + \varepsilon \quad (3.2)$$

y bağımlı değişken, x ise bağımsız değişken, β_0 kesişim sabiti (*intercept*) ve β_1 ise eğimi (*slope*) ifade etmektedir. ε ise hata parametresidir. Örneğin hava sıcaklıklarının acil servis başvurularına etkisi çıkarılmak istendiğinde başvurulabilecek bir modeldir. Bu durumda acil servis başvuruları bağımlı değişken, hava sıcaklıkları ise bağımsız değişken olacaktır.

3.1.2. Çoklu Doğrusal Regresyon

Çoklu doğrusal regresyon, basit doğrusal regresyondan farklı olarak, bir veya birden çok bağımsız değişken ile bağımlı değişken arasındaki doğrusal ilişkiyi çözümlenmektedir.

$$y_t = \beta_0 + \beta_1 x_{t1} + \dots + \beta_n x_{tn} + \varepsilon_t \quad (3.3)$$

$y_t = t$ anındaki gözlemlenen bağımlı değişken

$x_{ti} = t$ anındaki i . bağımsız değişken

$\beta_0 =$ regresyon katsayısı

$\beta_i = i$. bağımsız değişken katsayısı

$n =$ bağımsız değişken sayısı

$\varepsilon_t = t$ anındaki hata parametresi

Çoklu doğrusal regresyon, biyosürveyans kapsamında, özellikle haftanın günü etkisi, mevsimsel etki, tatil etkisi gibi girdilerin yapay(*dummy*) değişkenlerin de yardımıyla tek modelde ele alınmasını sağlamaktadır. 3.6.1 ve 3.6.2 kesimlerinde bu işlemin nasıl yapıldığı da daha ayrıntılı bir biçimde ele alınmaktadır.

Doğrusal regresyonda, bağımsız değişkenlerin bağımlı değişken ile toplama işlemine göre ilişkisi ele alınmaktadır. Kestiricilerin etkisinin toplanır(*additive*) yerine çarpımsal(*multiplicative*) olduğunu varsayıldığı durumlar için ise, bağımlı değişken y_t 'nin yerine $\log(y_t)$ kullanılması da önerilmiştir [48].

3.1.3. Adaptive Regression

“*Adaptive regression*” biyosürveyans için ilk olarak Burkom’un çalışmasında [9] önerilmiş ve Cusum ile birlikte uygulandığında başarılı olduğu başka çalışmalara da konu olmuştur [42].

“*Adaptive regresyon*” kayan bir pencere ile hesaplanır. Bu pencere, ele alındığı döneme ait mevsim ve eğilim örüntülerine dair çıkarımın yapılmasında kullanılır. Pencere uzunluğu için Burkom her ne kadar 8 haftalık bir sabit pencereyi önermişse de, Burkom’un bu yöntemini öneren Fricker, 15 gün ile 90 gün arasında değişen en iyi başarımları not etmiştir [42]. Pencere büyüklüğünün nasıl belirleneceği ile ilgili ayrıntılı bir inceleme için Dunfee ve Hegler’in çalışmasına bakabilirsiniz [30].

$$y(i) = \beta_0 + \beta_1 \times (i - t + n + 1) + \beta_2 I_{Pzt} + \beta_3 I_{Sal} + \beta_4 I_{Çrs} + \beta_5 I_{Prş} + \beta_6 I_{cum} + \beta_7 I_{cts} + \varepsilon \quad (3.4)$$

y_t , gözlemlenen vaka sayısını, $\beta_2 - \beta_7$ gün etkisi katsayılarını, β_0 sabit kesişimi, β_1 doğrusal eğilimin eğim katsayısını vermektedir. Burkom’un önerdiği “*adaptive*

regresyon” modelinde, $y_t^* = \log(Y_t + 1)$ olarak kullanılmıştır. Çoklu doğrusal regresyonda da belirttiğimiz gibi, logaritması alınarak etkenlerin çarpımsal olduğu öngörülmüştür.

“*Adaptive regresyon*” ın, doğrusal eğilimleri ortaya çıkarmaktaki başarımı yüksek olsa da, kademeli bir artışa sahip olan bir biyolojik terör saldırısına karşı zaafılarını göz ardı etmemek gerekir. Bunun önüne geçmek adına, *EARS* yöntemlerinde önerilen iki günlük gecikme(*lag*), “*adaptive regression*” için de önerilmektedir [9]. Bu sayede, son iki gün içinde gerçekleşen kademeli yükseliş yerine, daha büyük bir eğime sahip yükseliş gözlemlenecektir. “*Adaptive regression*” ile *EARS*’ın tek benzerliği bu değildir. Aynı zamanda, her iki yöntemde de kayan bir pencere kullanılmaktadır. Eşitlik (3.4)’teki t değeri, kayan pencerenin boyutudur.

3.2. 7 Gün Farkı

7 gün farkı [49] oldukça yalın bir tahmin yöntemidir. Doğrudan 7 gün öncesindeki gözlem sayısını tahmin olarak kullanmaktadır.

$$y_t = x_{t-7} \quad (3.5)$$

Eşitlik (3.5)’te hesaplanan 7 gün farkı eşitliğinde y_t , öngörülen gözlem miktarıdır. Tahmini yapılan günün bir önceki haftadaki gözlem sayısı doğrudan öngörülen gözlemdir. Biyosürveyans kapsamında, diğer yöntemlerde olduğu gibi artık (*residual*) değer çıkartılarak düzleştirme işlemine tabi tutularak kullanılabilir. Böylece bir önceki haftanın aynı gününe göre gözlem miktarındaki artış oranı incelenmiş olur.

3.3. Hareketli Ortalama

Hareketli ortalama(*moving average*), veri kümesine ait alt kümelerden yeni bir veri kümesi yaratmaya yarayan bir tahmin yöntemidir. Bu alt kümelerin uzunluğu, tahminde kullanılacak olan dönemin uzunluğudur.

3.3.1. Basit Hareketli Ortalama

Basit hareketli ortalama(*Simple Moving Average*), uzunluğu verilen dönemlerin ortalamasından tahmin çıkaran bir yöntemdir. En basit hareketli ortalama yöntemidir.

$$y_t = \frac{1}{n} \sum_{i=1}^n x_{t-i} \quad (3.6)$$

3.3.2. Ağırlıklı Hareketli Ortalama

Biyosürveyans kapsamında yaygın değildir. Basit hareketli ortalama, yükselişleri gözlemlemek zaman alabilmekteyken, ağırlıklı hareketli ortalama (*Weighted Moving Average*) daha çabuk tepki vermektedir [50].

$$y_t = \frac{nx_{t-1} + (n-1)x_{t-2} + \dots + 2x_{t-n-1} + x_{t-n}}{n + (n-1) + \dots + 2 + 1} \quad (3.7)$$

Tahminde bulunulan güne en yakın günün çarpanı alt küme uzunluğu iken, geçmişe doğru gidildikçe çarpan küçülmekte ve en uzaktaki elemanın çarpanı 1'e düşmektedir. Bu sayede, yakın zamana ait gözlemlerin tahmin üzerindeki etkinliği artırılmaktadır.

3.3.3. Üstel Ağırlıklı Hareketli Ortalama

Üstel ağırlıkla hareketli ortalama (*EWMA*), 2.2. kesimde ele aldığımız üzere, aynı zamanda bir *SPC* yöntemidir. Zaman serilerinde tahmin amaçlı kullanılan popüler yöntemlerden bir tanesidir.

$$y_t = \alpha x_t + (1 - \alpha)y_{t-1} \quad (3.8)$$

Geçmiş verilerin etkinliği üstel olarak azalır. Ayrıca hesaplaması kolay ve asgari veri ihtiyacı oldukça düşüktür.

3.4. Holt Yöntemi

Charles H. Holt'un önerdiği bir üstel düzleştirme yöntemidir [51]. *Holt*, *EWMA*'ya ek olarak, eğilimleri ortaya çıkarmak için eğilim etmenini de hesaplamaya dâhil etmeyi önermiştir.

$$\begin{aligned} L_t &= \alpha x_{t-1} + (1 - \alpha)(F_{t-1} + T_{t-1}) \\ T_t &= \beta(F_t - F_{t-1}) + (1 - \beta)T_{t-1} \\ H_{t+m} &= F_t + mT_{t+m} \end{aligned} \quad (3.9)$$

$$\alpha = \text{Düzgünleştirme Sabiti } (0 \leq \alpha \leq 1)$$

$$\beta = \text{Doğrusal eğilim kestirimi sabiti } (0 \leq \beta \leq 1)$$

$$L_t = t \text{ anında, eğilimin dahil edilmediği öngörülen gözlem sayısı}$$

$$T_t = t \text{ anındaki eğilim}$$

$$H_{t+m} = t + m \text{ günü için öngörülen gözlem sayısı}$$

Esnek ve basit bir yöntem olmasına karşın eğilimleri verinin içinden sıyırmakta oldukça başarılıdır [52]. Sürveyans sistemlerinde, $m=1$ olarak düşünülmelidir. Çünkü salgın güncel verilerde aranır ve gelecekte olacağına dair bir tahminde bulunmak mümkün olmayabilir.

Holt yönteminde başlangıç değerleri $L_1 = x_1$ olarak atanır. T_1 ise 3 farklı şekilde ilklendirilebilir.

$$T_1 = x_2 - x_1 \text{ veya } T_1 = \frac{(x_p - x_1)}{p-1} \text{ veya } T_1 = 0 \quad (3.10)$$

$$p = \text{Döngü süresi (günlük veriler için haftalık döngü için } p = 7)$$

3.5. Holt ve Winters Yöntemi

Holt'un öğrencisi olan Peter Winters tarafından, Holt'un önerdiği yöntemden genişletilerek önerilmiştir [53]. Holt'un düzleştirme yöntemine ilave olarak, dönemsel bileşenlerin de ele alınması sağlanmıştır. Dönemler, belirlenen uzunluklarda ele alınarak, dönemlik etkiler ortaya çıkarılmaktadır.

Zaman serilerinde etkiler toplanır ve çarpımsal model olmak üzere iki farklı şekilde etki yaratır. Kesim 3.6'da bu etkiler daha ayrıntılı bir şekilde anlatılmaktadır. Bu modeller Holt&Winters düzleştirme yönteminde, her iki etkide de modellenenmektedir.

Toplanır HW Modeli

$$L_t = \alpha(x_t - S_{t-s}) + (1 - \alpha)(L_{t-1} + T_{t-1})$$

$$T_t = \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1}$$

$$S_t = \gamma(x_t - L_t) + (1 - \gamma)S_{t-s}$$

$$HW_{t+m} = L_t + mT_t + S_{t-s+m} \quad (3.11)$$

$$L_t = t \text{ anındaki temel düzey değeri}$$

$$T_t = t \text{ anındaki eğilim değeri}$$

$$S_t = t \text{ anındaki mevsim etkisi değeri}$$

$$\alpha = \text{Düzgünleştirme sabiti } (0 \leq \alpha \leq 1)$$

$$\beta = \text{Doğrusal Eğilim sabiti } (0 \leq \beta \leq 1)$$

$$\gamma = \text{Mevsimsel Değişim sabiti } (0 \leq \gamma \leq 1)$$

$$s = \text{Dönem uzunluğu}$$

$$HW_{t+m} = t + m \text{ anındaki öngörülen gözlem değeri}$$

Holt yönteminde olduğu gibi, (3.10) eşitliğinde verilen L_t , T_t ve S_t değişkenleri de başlangıç değerleri ile ilklendirmek gerekir. Önerilen [54] ilklendirmeler:

$$\begin{aligned} L_s &= \frac{1}{s}(x_1 + x_2 + \dots + x_s) \\ T_s &= [(x_{s+1} + x_{s+2} + \dots + x_{s+s}) - (x_1 + x_2 + \dots + x_s)]/m^2 \\ S_i &= x_i - L_s \quad (1 \leq i \leq s) \end{aligned} \quad (3.12)$$

Anlaşılabacağı üzere, Holt&Winters düzleştirme yöntemi ilk dönemde tahmin üretmez. Çünkü eşitlik (3.11)'de görüleceği gibi, mevsimsel bileşen, bir dönem önceki mevsimsel etkiye ihtiyaç duymaktadır.

Çarpımsal HW Modeli

Çarpımsal model ile toplanır modelde, temel düzey değeri ile eğilim değerlerinin hesaplanması aynı şekilde yapılır. Ancak mevsimsel etkinin çarpımsal etki yaratması için, mevsimsel etki ve tahmin:

$$\begin{aligned} S_t &= \gamma \frac{x_t}{L_t} + (1 - \gamma)S_{t-s} \\ HW_{t+m} &= (L_t + mT_t)S_{t-s+m} \end{aligned} \quad (3.13)$$

Eşitlik (3.13)'te gösterildiği gibi hesaplanmaktadır. Yine aynı şekilde başlangıç ilklendirmelerini yaparken, temel düzey ile eğilim başlangıç değerli toplanır modelde olduğu gibi hesaplanır. Mevsimsel etkiler ise eşitlik (3.14) ile ilklendirilmektedir [54].

$$S_i = \frac{x_i}{L_s} \quad (1 \leq i \leq s) \quad (3.14)$$

3.6. Ayrışım (Decomposition)

Zaman serileri genel olarak dört bileşenden oluşmaktadır. Bunlar eğilim(T) bileşeni, döngüsel(C) bileşen, mevsimsel(S) bileşen ve düzensiz(I) bileşendir. Eğilim; doğrusal, karesel(*quadratic*) ya da eğrisel(*curvilinear*) olabilir. Döngüsel bileşenler çok daha uzun süreli eğilimler için kullanılır. Uzun bir zaman serisi gerektirir [55, 56]. Tespitinin zor ya da imkânsız olduğu durumlarda yok sayılmak durumunda kalır [55]. Eğilim bileşeni ile döngüsel bileşenin tek bileşen olarak kullanılması da yaygın bir kullanımdır. Bu durumda ana bileşen sayısının üç olduğunu da düşünebiliriz. Mevsimsel etki, zaman serilerindeki en çok ele alınan bileşenlerin başında gelmektedir. Genel olarak, yıllık döngülerde, aylık ya da

mevsimlik etkilerin ele alındığı bileşendir. Son olarak, düzensiz bileşenler, açıklanamayan tüm diğer etkenlerin toplandığı bileşendir. Bazı seriler bu bileşenlerin tamamını içerebilirken, bazı seriler bu bileşenlerin bazılarını içermeyebilir [56]. Bu temel bileşenler dışında, bir zaman serisinde etken olarak yer alan pek çok örüntü yer alabilir. Bir zaman serisi üzerinde, haftanın hangi günü olduğunun, maaş gününün, hava sıcaklığının vb. bazı etkenlerin açıklanabilir örüntülere sebep olduğu görülebilir.

Zaman serilerinde örüntüler, veri kümesi üzerindeki etkileri toplanır ve çarpımsal olarak iki farklı şekilde olabilir. Toplanır model, bir bileşenin, veriye aynı koşullarda aynı *büyükükte* bir örüntüye sebep olduğu modeldir. Örneğin, her yıl Haziran ayında otomobil satışları yaklaşık 1000 adet artıyorsa, bu, Haziran ayının satışlar üzerinde toplanır bir etki yarattığını gösterir.

$$x_t = T_t + C_t + S_t + I_t \quad (3.15)$$

Eşitlik (3.15)'te x_t gözlem miktarını, T_t eğilim etkisini, C_t döngüsel etkiyi, S_t mevsimsel etkiyi ve I_t düzensiz, tanımlanmamış etkiyi ifade etmektedir. Yukarıda bahsedildiği gibi, bazı zaman serilerinde T_t ve C_t birlikte varsayılmaktadır. Bu zaman serilerini eşitlik (3.16) ile ifade edebiliriz.

$$x_t = T_t + S_t + I_t \quad (3.16)$$

Çarpımsal model, bir bileşenin, veriye aynı koşullarda aynı *oranda* bir örüntüye sebep olduğu modeldir. Örneğin, her yıl Haziran ayında otomobil satışları %20 artıyorsa, bu, Haziran ayının satışlar üzerinde çarpımsal bir etki yarattığını gösterir.

$$x_t = T_t \times C_t \times S_t \times I_t \quad (3.17)$$

$$\log(x_t) = \log(T_t) + \log(C_t) + \log(S_t) + \log(I_t) \quad (3.18)$$

Eşitlik (3.17) ve (3.18) çarpımsal modeli karşılayan eşitliklerdir. Ancak her zaman bir model, tamamı toplanır ya da tamamı çarpımsal bileşenlerin oluşturduğu örüntülerden oluşmaz. Bu durumlar için geliştirilmiş olan bir model de her bileşen türünü de eşitlik (3.19) ile ifade edebiliriz.

$$x_t = (T_t + I_t) \times S_t \quad (3.19)$$

Winters'in hocası Holt'tan [51] esinlenerek geliştirdiği Holt&Winters yönteminin [53] kesim 3.5'te anlatılan çarpımsal modeli de eşitlik (3.19)'un bir örneğidir.

Ayrışım(*decomposition*), veriyi oluşturan ve tanımlanabilir olan örüntülerin, düzleştirme(*smoothing*) maksadıyla veriden çıkarılması işlemidir. Düzleştirme işlemi, verilerin temizlenerek çözümlenebilirliğini artırmak amacıyla yapılmaktadır. Düzleştirme, her çalışma alanı için farklı bir anlama gelebilir. Örneğin bir istatistikçi bunu eğri-uydurma(*curve fitting*) olarak düşünürken, bir elektronik mühendisine göre sinyali korumak için yapılan bir gürültü önleme işlemi olarak görülebilir [57]. SPC yöntemlerinde güvenilir bir tespit başarımı elde etmek için oldukça uzunca bir geçmişi olan bağımlılık çıkarma işlemi kullanılır [58]. Eşitlik (3.15), (3.16), (3.17), (3.18) ve (3.19)'da gözlemi oluşturan bileşenlerin, gözlemden çıkarılması için yapılması gereken, toplanır örüntülerin çıkarılması, çarpımsal örüntülerin ise bölünmesidir.

$$d_t = x_t - T_t = S_t + I_t \quad (3.20)$$

$$d_t = \frac{x_t}{S_t} = T_t \times C_t \times I_t \quad (3.21)$$

Eşitlik (3.20) ve (3.21)'de bulunan d_t , veri kümesinin t anındaki(dönemindeki) ayrıştırılmış ya da düzleştirilmiş gözlem miktarıdır. (3.16) eşitliğindeki gözlemden eğilimi ayrıştırırsak (3.20) eşitliği elde edilir. (3.17) eşitliğinden mevsimsel etkiyi ayrıştırırsak (3.21) eşitliği elde edilecektir.

3.6.1. Haftanın Günü Etkisi Ayrışımı

Haftanın günü etkisi(*day of the week effect-DOW*), zaman serilerinin bilinen örüntülerindedir. Dünyada ve ülkemizde, geçmişten bugüne özellikle borsalar üzerinde haftanın günü etkisi araştırmaları ön plana çıkmaktadır [59] [60] [61] [62]. Sürveyans sistemlerinde de haftanın günü etkisi yoğun olarak işlenmektedir [7] [63]. Hatta sürveyans verisi benzetiminde yapılan araştırmalarda da bu etkinin göz ardı edilmediğini görebiliyoruz [64] [63]. Bu araştırmalarda genel olarak haftanın her günü ayrı ayrı ele alınıyor olsa da, hafta içi/hafta sonu etkisi şeklinde ele alındığı da görülmektedir [65].

$$x_t = (\beta_0 + \beta_1 \times t) + \beta_{pzt}I_{pzt} + \beta_{sal}I_{sal} \dots + \beta_{pzt}I_{pzt} \quad (3.22)$$

$\beta_0 + \beta_1 \times t =$ uzun dönemsel doğrusal eğilim

$\beta_{pzt} \dots \beta_{pzt} =$ kestirim sabitleri

$I_{pzt} \dots I_{pzt} =$ yapay değişkenler

Haftanın günü etkisi araştırılırken yapay değişkenler kullanılır. Bu yapay değişkenler sayesinde zaman serisini günlere göre kategorize ederek inceleme imkânı olur. Eşitlik (3.22)'de verilen yapay değişkenleri şu şekilde kullanırız; eğer günlerden pazartesi ise I_{pzt} 1 ile işaretlenir, diğer yapay değişkenlere 0 atanarak kestirim sabitleri devre dışı bırakılır. Böyle çoklu değişken içeren etkileri araştırmak için çoklu doğrusal regresyon uygulamak doğru bir yaklaşım olabilir [38]. Eşitlik (3.22) ile haftanın gününün gözlem üzerindeki ilişkisini araştırmaya yarayan çoklu doğrusal regresyon eşitliği ifade edilmektedir. Yani, haftanın günü etkisi normal bir çoklu doğrusal regresyon modeli ile ifade edilmiştir. İstenirse hafta içi/hafta sonu şeklinde bir sınıflandırma 7 yapay değişken yerine 2 yapay değişken kullanılarak da ele almak mümkündür. İlişkiler hesaplandıktan sonra ayrıştırma yöntemleri ile haftanın günü etkisi zaman serisinden çıkartılarak veri düzleştirilmesi sağlanabilir.

3.6.2. Mevsimlik Etki Ayrışımı

Mevsimsel etki(*seasonal effect*), yıl içerisinde, mevsime bağlı olarak oluşan örüntülerin ifadesi için kullanılmaktadır. Mevsim, otomotiv piyasası [66], turizm sektörü [67] veya enflasyon [68] gibi pek çok ekonomiyi ilgilendiren alanı doğrudan etkilemektedir. Mevsimsel etkinin yarattığı örüntünün ayrıştırılması daha tutarlı karşılaştırılmaların yapılabilmesi için önemli bir veri işleme adımıdır [69].

Halk sağlığında mevsimsel etki yaygın olarak ele alınmaktadır. Bu mevsimsel etkinin yarattığı düzensizliği ortadan kaldırmak için Serfling [70] sinüs ve kosinüs kestiriciler kullanarak çözmeyi önermiştir. Bu yöntem sürveyans sistemlerinde, veri üretme ve salgın tespit etme konularında geniş bir kabul görerek yaygın olarak kullanılmaktadır [64].

$$x_t = (\beta_0 + \beta_1 \times t) + \beta_3 \sin\left(\frac{2\pi}{365.25}\right)t + \beta_4 \cos\left(\frac{2\pi}{365.25}\right)t \quad (3.23)$$

Burada, haftanın günü etkisindeki hesaplama benzer bir hesaplama yapılmaktadır. 365.25'e bölünme işleminin yapılmasını sebebi, bir yılın ortalama

gün uzunluğundan kaynaklanır. Eğer mevsimsel etkiyi haftanın günü etkisinde olduğu gibi yapay değişkenlerle çözümlenmek istersek eşitlik (3.24)'ü kullanabiliriz.

$$x_t = (\beta_0 + \beta_1 \times t) + \beta_{ock}I_{ock} + \beta_{şbt}I_{şbt} \dots + \beta_{arl}I_{arl} \quad (3.24)$$

$$\beta_{ock} \dots \beta_{arl} = \text{kestirim sabitleri (ocak ... aralık)}$$

$$I_{ock} \dots I_{arl} = \text{yapay değişkenler(ocak ... aralık)}$$

Böylece mevsimsel etkiyi aylar bazında, daha ayrıntılı bir şekilde, ele alma imkânı elde ederiz. Ancak bu şekilde ele alındığında da, tüm Mart ayının aynı etkiyi yarattığı varsayımını gözden kaçırmamak gerekmektedir. Bunun yanı sıra, hem (3.23) eşitliği hem de (3.24) eşitliği toplanır bir mevsimsel ilişki araştırmaktadır. Ne var ki, kaydedilen araştırmalar [9] [63] [71] gösteriyor ki mevsimsel örüntü toplanır yerine çarpımsal bir etki ile oluşmaktadır. Daha önceki kesimlerde değinildiği gibi, gözlem sayısının ve mevsimsel parametrelerin doğal logaritmasını alarak çoklu doğrusal regresyona tabi tutarsak mevsimsel etkiyi çarpımsal olarak ele alabiliriz. Bunun yanı sıra Winters'ın yöntemi de mevsimsel etkiyi hem toplanır hem de çarpımsal olarak ele alabilecek şekilde ayarlanabilmektedir.

3.7. Veri Ön İşleme Başarım Ölçütleri

Yukarıda açıklanan ön işleme yöntemleri, kendine özgü durumlarda kullanılması daha doğru sonuçlar üreteceği öngörülen, esnek ve farklı parametrelerle, zaman serileri üzerinde farklı başarımlar elde ederler. Bu ölçümler, istatistiği araç olarak kullanan araçlarda yaygın olarak yapılmaktadır. Biyosürveyans verileri üzerinde işlem yapıldığını öngören araştırmalar da kullandıkları ölçütleri zaman zaman paylaşmışlardır. Bu tez kapsamında gerçekleştirilen araçta kullanıcıya sunulan veri işleme başarım ölçütleri Hata Kareler ortalaması(MSE), Hata Kareler ortalaması karekökü(RMSE), Mutlak Ortanca Yüzde Hatası(MdAPE) ve Ortanca Mutlak Sapma(MdAD)'dır. Bu ölçütlerle yapılan ölçümler kullanılarak, düzleştirme yönteminde tercih edilen parametrelerin tayini tekrar düzenlenerek başarımın artırılması hedeflenebilir.

4. GERÇEKLEŞTİRİM

Sürveyans sistemlerinin doğru ön işleme ve çözümlene yöntemleri ile topladığı verilerden salgın tespiti yapması yalnızca bir çalışma alanı olmaktan çıkmış, çözülmesi zorunlu olan bir hal almıştır. Bu bağlamda, gelişmiş ülkelerde birçok sürveyans merkezi oluşturulup, biyolojik terör tehdidinin yarattığı tedirginliğe çözüm aranmaktadır.

Salgın tespitinde kullanılan tespit yöntemleri ile girdi olarak kullanılan veri kümelerinin ön işleme işlemlerinde yaygın olarak kullanılan yöntemler kesim 2 ve 3'te açıklanmıştır. Salgın tespitini araştıran araştırmacılar, yeni yöntemler önermekten ziyade, *SPC* yöntemleri başta olmak üzere, var olan yöntemleri, farklı parametreler ve eşik değerleri ile karşılaştırmışlardır. Örnek bir karşılaştırma çalışması olarak Fricker'ın yaptığı araştırmada [42], *Cusum* ile *EARS* yöntemleri kıyaslanmıştır. Kıyaslanma yapılırken, *EARS* yöntemleri için pencere boylarını ve eşik değerlerini değişik parametrelerle test etmiş, *Cusum* için ise veriyi "*adaptive regression*" ile ön işleme tabi tutmuş ve daha sonra elde edilen düzleştirilmiş veriye *Cusum* uygulamıştır. "*Adaptive Regression*" ise, farklı pencere boyları ile uygulanarak farklı sonuçlar gözlemlenerek karşılaştırmaya konu edilmiştir.

Yapılan araştırmalar genel olarak veriye özel olarak, kullanılan yöntemlerin başarımlarının kıyaslanması üzerine yoğunlaşmaktadır. Bu amaçla yapılan çalışmalarda, bir yöntemin bir diğerine üstünlüğünden söz etmekten ziyade, kullanılan yöntemlerin yalnızca kullanılan veri kümesi için yorumlanması tercih edilmiştir [15]. Her veri kaynağının kendine özgü örüntüler barındırdığı gerçeğinin yadsınamaz olduğu daha önceki kesimlerde de belirtilmiştir.

Bu tez kapsamında, araştırmacıların ellerindeki verilere, literatürde sık sık kullanılan ön işleme ve tespit yöntemlerini uygulayarak hem bilimsel çalışmalarında kullanabilecekleri, hem de gerçek zamanlı sistemlerinde hangi ön işleme ve tespit yöntemi ikilisinin tercih edilmesi gerektiğine karar vermelerine destek olacak bir uygulama gerçekleştirilmiştir. Uygulama, günümüz şartlarına uygun bir şekilde web tabanlı olarak gerçekleştirilmiş, kullanıcılara araştırmalarını ve verilerini kayıt altına alarak tekrar kullanma imkânı sağlayacak şekilde tasarlanmıştır. Ayrıca elde edilen sonuçların, tam olarak hangi ön işleme ve tespit

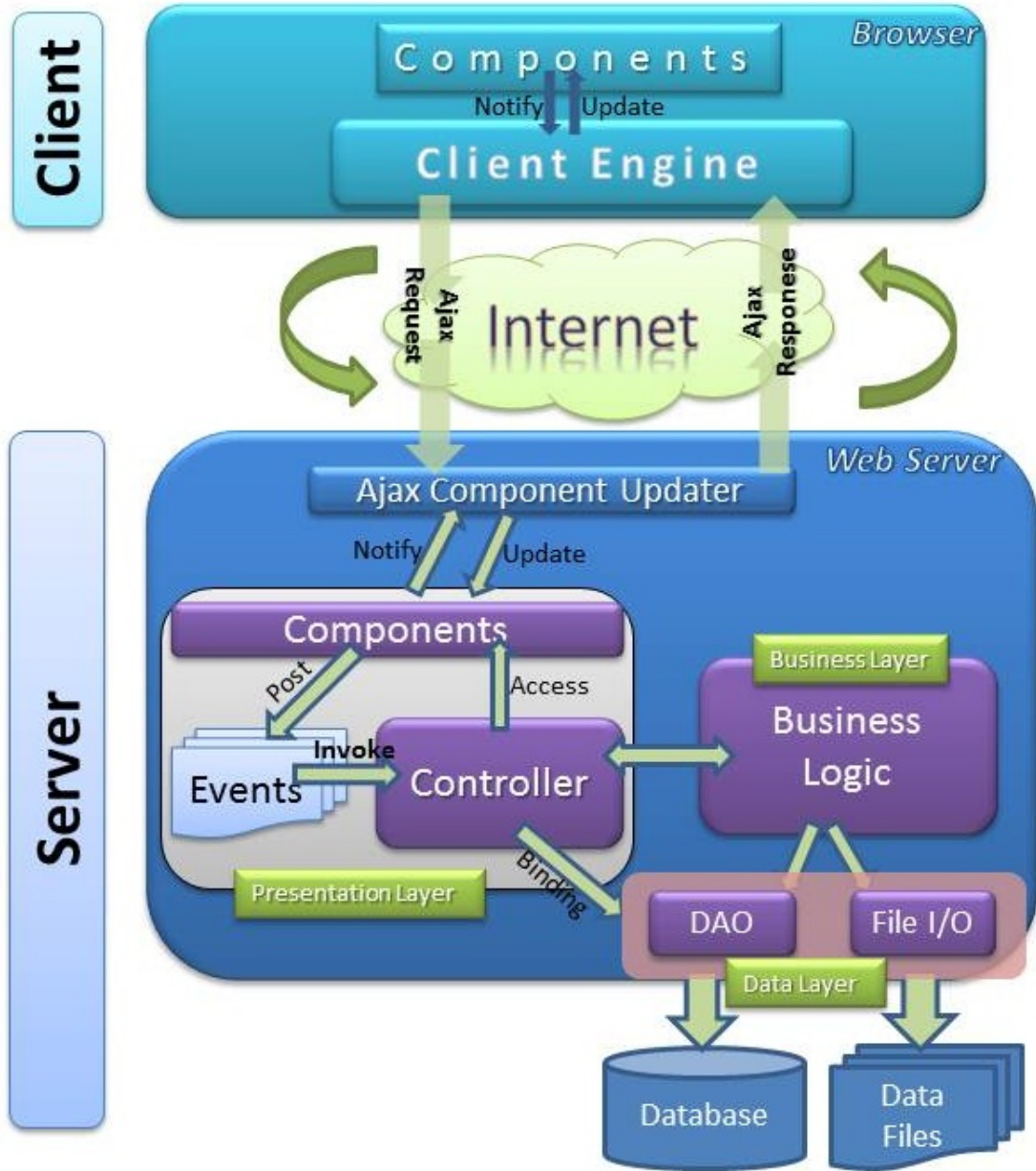
yöntemlerinin hangi parametreler ile elde edildiği de gözlemlenebildiği “*Outbreak Detection Tool*” geliştirilmiştir(bundan sonra *araç* olarak anılacaktır).

4.1. Sistem Mimarisi ve Gereksinimler

Bu kesimde kullanılan yöntemlerden bahsedilmiş, ancak kullanılan teknolojilerin isimlerinden olabildiğince kaçınılmıştır. Çünkü tercih edilen teknolojilerin yerine, bu yetenekleri sağlayan başka pek çok seçenek bulunabilir ve yöntemin anlaşılması kullanılan teknolojiye daha önemlidir. İlerleyen kesimlerde kullanılan teknolojileri, açıklamalar ve neden tercih edildiklerine dair bilgilerle birlikte bulabilirsiniz.

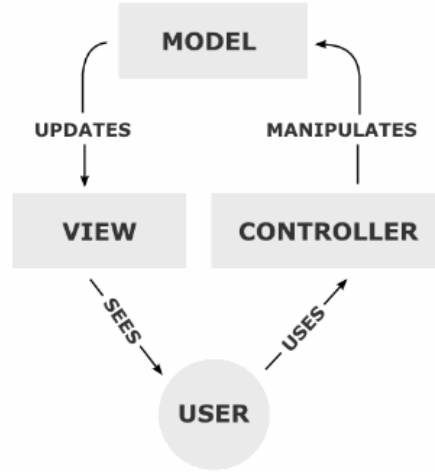
Tez kapsamında geliştirilen araç sunucu-istemci mimarisi üzerinde koşmaktadır. Bu mimaride, sunucu ile istemciler ayrı uçlarda bulunur ve istemciler bir uygulama aracılığı ile sunucuya isteklerini gönderirler. Aracın bu mimariye uyarlanması, en yaygın sunucu-istemci uygulama yöntemi olan, internet üzerinden iletişimin kurulduğu web uygulaması tercih edilmiştir. Web uygulamalarında, istemci uygulamaları tarayıcılardır. Araç, etkileşimli bir web uygulaması geliştirme tekniği olan *Ajax* tekniği ile geliştirilmiştir. *Ajax* ile bir web sayfasındaki alanlardan tümünün değil, yalnızca gerekli kısım için istem yaratılabildiği için çok daha hızlı çalışan web uygulamaları geliştirilmesi mümkün olabilmektedir. *Ajax* tekniğinin Java ile uygulanabilmesi için bu yeteneği sunan bir uygulama çatısı kullanılmalıdır. İstemciden gelen ilk *istek* ile birlikte istemciye bir seferlik, *Ajax* desteği sağlayacak olan motor(*engine*) yüklenmektedir. Motor ile hem güvenilir hem de hızlı bir etkileşim sağlanabilir. İstemciden sunucuya gelen tüm *Ajax* talepleri *http* portundan gelmektedir. Sunucuya gelen istekler ilk olarak, uygulama çatısı tarafından sağlanan *Ajax Component Updater* tarafından ele alınır ve sunucuda koşan uygulamadaki form öğelerinin ele alındığı bileşen sınıflarına bildirilir. Sınıflar, kendilerine gelen bu isteklerle olayları tetikler ve *Events* veri yapısında kayıtlı olaylar da ilgili *Controller* nesnelere tetiklenmesini sağlarlar. *Controller* nesnelere, bazen iş mantığının ele alındığı *Business Logic*'e ihtiyaç duyduğu parametreleri sağlarken, bazen de, bizim aracımızda da uyguladığımız gibi, doğrudan veri tabanından elde edilen verilerden oluşturulan veri yapılarına bağlanarak görsel öğeler ile gerçek veriler arasında devingen bir bağlantı oluştururlar. Veri katmanı, kullanıcının veri tabanı üzerinde gerçekleştirmek istediği ilk işlem ile açılan veri tabanı oturumunu kullanarak gerçekleştirilir. Oturumun sürekliliği Oracle tarafından

Java standardına dâhil edilen JPA (*Java Persistence API*) ile sağlanmaktadır. Bağlantı ayarları (port bilgileri, veri tabanı adresi, kullanıcı adı, şifre, oturum süresi, bağlantı türü...) ve veri tabanındaki çizelgeler ile uygulamadaki sınıflar arasındaki ilişki bir XML ile uygulamaya tanımlanmakta ve sunucunun başlatılması ile birlikte uygulama veri tabanının şablon olarak durumu hakkında bilgi sahibi olmaktadır. Veri tabanından çekilen bir satırın uygulamadaki ilgili sınıftan yaratılan bir nesne ile bağlantısı kurulmakta ve ilgili alanlar yaratılan nesnedeki değişkenlere yazılmaktadır.



Şekil 4-1 ODT* Uygulama Mimarisi(**Outbreak Detection Tool*)

Şekil 4-1’de görüldüğü gibi, araç 3 katmandan oluşmaktadır. Bu katmanlar dışında bir de araca kullanıcılardan gelen istekleri bildiren ve yanıtları kullanıcıların görmesini sağlayan istemciye(tarayıcıya) gönderen *Ajax Component Updater* bileşeni bulunmaktadır. Kullanıcıdan gelen istekleri ayrıştırıp gerekli çağrılarını yapıldığı katman sunum katmanı (*presentation layer*), kullanıcının elde etmek istediği sonuçların üretilmesini sağlayan istatistiksel hesaplamaları gerçekleştirildiği katman iş katmanı (*business layer*) ve son olarak da veri tabanı ve kütük işlemlerinin ele alındığı katman veri katmanıdır (*data layer*). Belirtilen katmanlar Şekil 4-2’de genel mimarisi gösterilen MVC¹⁰ tasarım örüntüsünde de belirtilen yazılım mühendisliği bağlamında kabul görmüş katmanlardır. Katmanların birbirinden bağımsız tasarlanıp gerçekleştirilmesi, öngörülen ya da beklenmedik değişikliklere uyum sağlanması için çok önemlidir. Biz de tez kapsamında geliştirdiğimiz aracımızda bu gereksinimlere yanıt verecek bir yol izledik. İstemcide bulunan katmanlara gelince, kullanıcı arayüzlerindeki form bileşenlerini ifade eden *Components* katmanı ile bu bileşenlerin etkileşimini organize eden *Client Engine* motoru istemcide koşan tarayıcılar tarafından ele alınmaktadır.



Şekil 4-2 MVC mimarisi

Hem sunucu hem de istemcide yer alan katmanların pek çoğu araç geliştirilmesi sırasında kodlanmış olsa da, bazı katmanlar kullanılan uygulama çatısı tarafından sunulan ve doğrulanmış katmanlardır. İstemci tarafındaki form bileşenleri tarafımızca kodlanmış, ancak bu bileşenlerden elde edilen istekler ve bu isteklerin sonuçlarının görüntülenmesini sağlayan motor uygulama çatısı tarafından

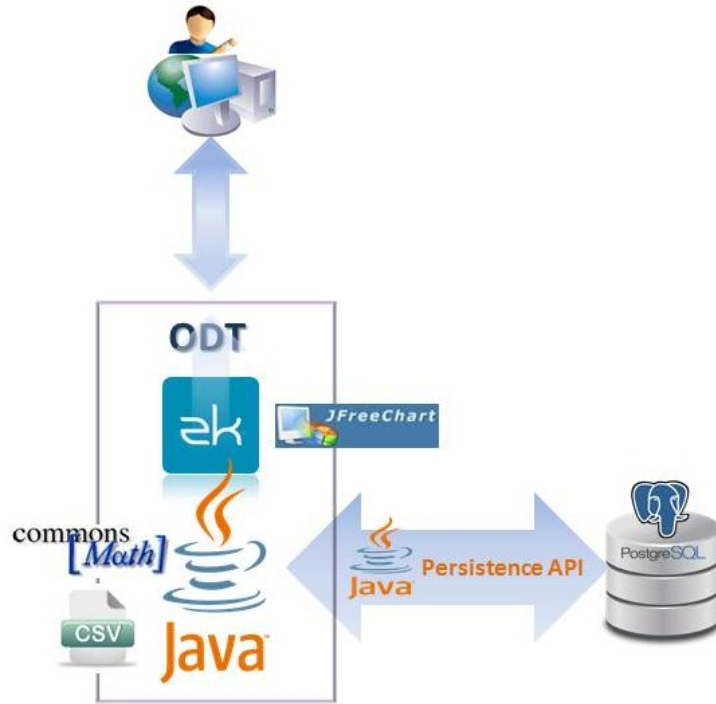
¹⁰ <http://en.wikipedia.org/wiki/Model%E2%80%93view%E2%80%93controller>

sağlanmıştır. Sunucu uygulamasında yer alan *Ajax Component Updater* ve olayları dinleyerek yöneten *Events* kesimi de uygulama çatısı tarafından sağlanan hizmetlerdir. Bunun dışında kalan katmanlar tez kapsamında tasarlanmış ve gerçekleştirilmiştir.

Kullanıcı, aracı kullanabilmek için internet bağlantısına, bir web tarayıcısına ve aracı kullanabilmek için bir kullanıcı adı, şifre ikilisine ihtiyaç duyar. Aracın işletim sistemi, tarayıcı vb. herhangi bir ortam bağımlılığı yoktur.

4.2. Kullanılan Teknolojiler

Araç, günümüz programlama dillerinin en popülerlerinden olan Java¹¹ programlama dili ile geliştirilmiştir. Aracın yeteneklerini tek tek ele aldığımızda,



Şekil 4-3 ODT Teknoloji-Kütüphane Tercihleri

arayüz kaygılarını bir kenara bırakırsak, tüm işlemlerin MATLAB¹² ya da R¹³ ile gerçekleştirilmesi de mümkündür. Ancak hem yazarın Java'ya olan yakınlığı hem de Java'nın birçok konuda daha esnek bir yapı sunmasından ötürü Java ile

¹¹ <http://www.oracle.com/us/technologies/java/overview/index.html>

¹² <http://www.mathworks.com/products/matlab/>

¹³ <http://www.r-project.org/>

geliştirilmesi uygun bulunmuştur. Araç geliştiriminde kullanılan teknolojilerin gene bir mimarisi Şekil 4-3 ile gösterilmiştir.

4.2.1. ZKoss (ZK)

Araç geliştirilirken, MVC mimarisi için uygun olan ZK¹⁴ uygulama çatısı tercih edilmiştir. ZK hem ücretli hem ücretsiz lisanslamaya sahiptir. Ücretsiz LPGL¹⁵ lisansı ile sunulan *Community Edition* yeterli olduğu için bu sürüm tercih edilmiştir.

ZK, Kore’li ZKoss firması tarafından geliştirilmekte olan, 2005 yılından günümüze *JavaScript* kodu yazmadan *Ajax* uygulamaları geliştirmeyi sağlayan, Java tabanlı bir uygulama çatısıdır. ZK’nın muadilleri olan *JSF*, *PrimeFaces* gibi Java tabanlı uygulama çatıları da yaygın bir biçimde kullanılmaktadır. Ancak yazarın ZK ile olan deneyimlerinden ötürü ZK tercih edilmiştir.

4.2.2. PostgreSQL

Araç için küçük de olsa bir veri tabanı desteği gerekmektedir. Bu destek, yine ücretsiz bir VTYS olan *PostgreSQL*¹⁶ ile sağlanmıştır. VTYS ile iş mantığı arasındaki iletişim *Java Persistence API*¹⁷ ile kurulmuştur.

PostgreSQL, ilişkisel veri tabanı standartlarını SQL dili desteği ile sunan ücretsiz ve açık kaynak bir VTYS’dir. 1995 yılında *Postgres95* olarak piyasaya çıkmış olsa da, geçmişi 1982 yılında Berkeley tarafından geliştirilen *Ingres* projesine kadar uzanmaktadır¹⁸. 1996 yılında *PostgreSQL* olarak ismi değiştirilmiştir. Ses, video, resim gibi çoklu ortam değişkenlerinin de saklanabildiği *PostgreSQL*’in geliştirilmesine günümüzde de devam edilmektedir.

4.2.3. Hibernate (Java Persistence API)

Hibernate, 2001 yılında Gavin King tarafından açık kaynak olarak başlatılmış, VTYS ile iletişim kuran ve veri tabanının yönetilmesine dair pratik çözümler sunan bir Java uygulama çatısıdır. Kısa sürede *JBoss* tarafından desteklenmeye

¹⁴ <http://www.zkoss.org/>

¹⁵ <http://www.gnu.org/copyleft/lesser.html>

¹⁶ www.postgresql.org/

¹⁷ <http://www.oracle.com/technetwork/java/javaee/tech/persistence-jsp-140049.html>

¹⁸ <http://en.wikipedia.org/wiki/PostgreSQL>

başlanmıştır. 2010 yılı itibari ile *Oracle* tarafından Java'nın standart kütüphanelerinden *Java Persistence API*'nin kapsamına dâhil edilerek Java standardı olarak kabul görmüştür. *Hibernate* sayesinde, geleneksel bir ilişkisel veri tabanı, nesnel ilişkisel bir iş mantığının koşulduğu uygulamada haritalanarak kullanılabilir.

4.2.4. Kullanılan Kütüphaneler

4.2.4.1. Apache Commons Math

"*Commons Math*"(CM)¹⁹, *Apache* tarafından, açık kaynak olarak sunulan, Java destekli bir matematik kütüphanesidir. CM kütüphanesi ile Veri Üretimi, Doğrusal Cebir, Nümerik Çözümleme, Karmaşık Sayılar, Diferansiyel Denklemler, Olasılık Dağılımı, Dönüşüm Yöntemleri, Geometri, Genetik Algoritmalar, İstatistiksel araçlar geliştiricilere sunulmaktadır. Bu tez kapsamında, basit doğrusal regresyon ve çoklu doğrusal regresyon yeteneklerinden yararlanılmıştır.

4.2.4.2. OpenCsv

Tez kapsamında, sürveyans sistemlerinden sağlanacak olan veriler, günlük gözlemlenen vaka sayısı şeklinde elde edileceği için, ücretsiz ve yaygın olarak kullanılan bir format olan csv tercih edilmiştir.

OpenCsv, açık kaynak erişimi bulunan, ücretsiz, basit bir csv ayrıştırıcıdır²⁰. Csv uzantılı kütüklerin, virgül tabanlı olarak okunması ve yazılmasını sağlayan, kullanımı kolay bir Java kütüphanesidir.

4.2.4.3. JFreeChart

Yapılan bir bilimsel çalışmanın en büyük değer taşıyan faydalarından biri, araştırma sonucunda elde edilen sonuçlardır. Sonuçların daha etkin ifade edilmesi için ise görsel öğeleri kullanmak yaygın bir tercihtir. Bu tez kapsamında, elde edilen sonuçlar görsel olarak çizelgelerle desteklenerek gerekli raporlama ihtiyacının giderilmesi amaçlanmıştır. Görsel çizimlerin sunulabilmesi için Java tabanlı *JFreeChart* çatısı kullanılmıştır. *JFreeChart*, isminden de anlaşılacağı üzere ücretsiz bir çizim aracıdır. *JFreeChart* ile yapılan çizimler standart resim biçimlerinde kaydedilebilmektedir. Java uygulamalarında, *JFreeChart* dışında, *JasperReports* gibi başarılı başka çizim araçları da mevcuttur. *JasperReports* da

¹⁹ <http://commons.apache.org/math/userguide/index.html>

²⁰ <http://opencsv.sourceforge.net>

ücretsiz bir açık kaynak uygulama çatısıdır ve doğrudan JFreeChart yerine kullanılabilir, ancak *JFreeChart*'in yeteneklerinin yeterli olmasından dolayı *JasperReports*'a ihtiyaç duyulmamıştır. Uygulamanın Ajax tabanlı geliştirilmesinde kullanılan ZK çatısının da barındırdığı çizim araçları mevcuttur. Ücretli sürümünde yer almasından ötürü tercih edilmemiştir.

4.3. Outbreak Detection Tool ve Örnek Süreç

Bu kesimde geliştirilen salgın tespit aracının yetenekleri ve örnek bir sürecin araç üzerinde uygulanması tartışılmıştır. Uygulanan örnekte, yapılan gerçek araştırmalardan esinlenerek gerçekçiliğinin artması hedeflenmiştir. Aracın arayüzünde, çalışmaların yurt dışında daha ağırlıklı yapıldığı gerçeğinden dolayı İngilizce kullanılmıştır.

4.3.1. Veri Kümesi İşlemleri

Veri kümeleri, salgın araştırmalarından elde edilen bulguları doğrudan etkileyen etmenlerdendir. Bunun yanında, farklı yöntemleri kıyaslayabilmek için, o yöntemlerin aynı veri kümelerinde gösterdiği başarımları kullanmak, yapılan araştırmanın daha sağlıklı olmasını sağlayacaktır. Bu tez kapsamında, veri kullanımını kolaylaştırmak ve esnetmek için veri ekleme/silme imkânı veren özelliklere de yer verilmiştir.

4.3.1.1. Veri Kaynağı Ekleme

Veri bulmak ve bu veriyi sisteme yüklemek, araştırma sürecinin başlangıç noktasıdır. Aracın desteklemiş olduğu kütük yapısı Şekil 4-4 'de gösterilmiştir. Bu biçimsel yapıda olmayan veri kaynaklarının araç tarafından doğru olarak ele alınması mümkün olamayacaktır.

Kullanıcı doğru bir biçime sahip kaynak verisini ön işlemeden geçirdikten sonra, araç hem işlenmemiş veri kütüğünü, hem de işlenmiş veri kütüğünü yukarıda belirtilen yapıda saklamaktadır. Bu saklama sırasında, salgın durumu ve gözlem tarihlerinde herhangi bir değişiklik olmamakta, yalnızca gözlem miktarının işlenmiş hali işlenmemiş miktarın yerini almaktadır.

25	01/24/1994,137,0	
26	01/25/1994,201,0	
27	01/26/1994,194,0	
28	01/27/1994,255,0	
29	01/28/1994,140,0	
30	01/29/1994,148,0	
31	01/30/1994,157,0	
32	01/31/1994,219,0	
33	02/01/1994,146,0	
34	02/02/1994,212,0	
35	02/03/1994,164,0	
36	02/04/1994,192,0	
37	02/05/1994,237,0	Gözlem Tarihi
38	02/06/1994,160,0	
39	02/07/1994,139,0	Salgın Durumu (0-yok / 1-var)
40	02/08/1994,203,0	
41	02/09/1994,148,0	Gözlem Miktarı
42	02/10/1994,196,0	
43	02/11/1994,222,0	
44	02/12/1994,170,1	

Şekil 4-4 .csv uzantılı girdi biçimi

Veri ekleme ile yalnızca sürveyans sistemlerince sağlanan ya da sonradan benzetimle üretilen veriler değil, aynı zamanda ön işleme amacıyla regresyonda kullanılacak olan veri kümeleri de yüklenebilir. Kullanıcı gözlem verilerinin sıcaklık, tatil ve sonrası vb. ile ilişkisini sorgulamak isteyebilir. Şekil 4-4 'de belirtilen girdi biçimine uyduğu sürece, her türlü veri kümesi araç tarafından kabul edilir.

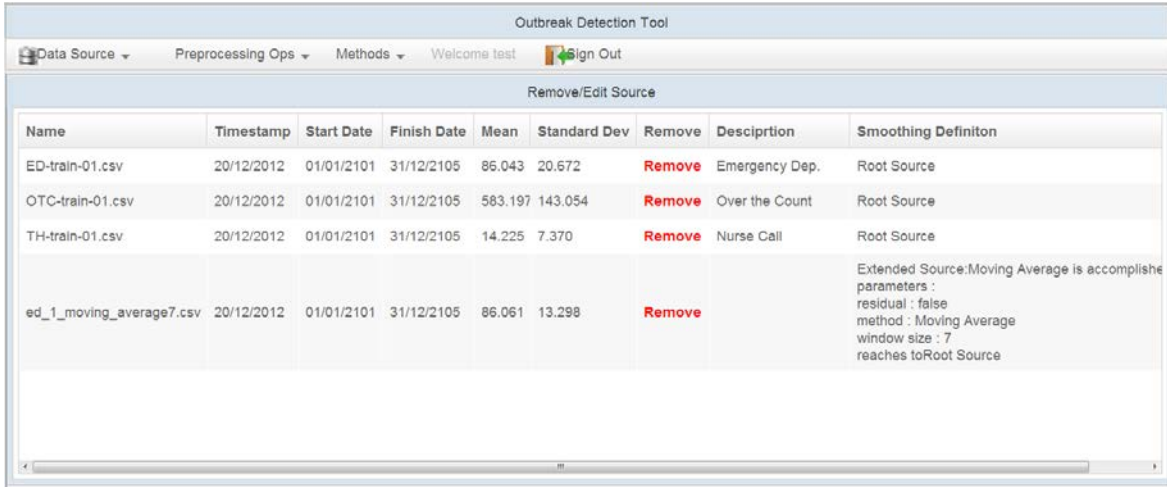
Outbreak Detection Tool			
Data Source	Preprocessing Ops	Methods	Welcome test Sign Out
Browse to upload		Save All	
RECENTLY ADDED DATA SOURCES			
File Name	Description (recommending : fill it)	Timestamp	Remove
ED-train-01.csv	Emergency Dep.	20/12/2012	Remove
OTC-train-01.csv	Over the Count	20/12/2012	Remove
TH-train-01.csv	Nurse Call	20/12/2012	Remove

Şekil 4-5 Veri Ekleme Ekranı

4.3.1.2. Veri Kaynakları Listeleme / Silme

Kullanıcı, yüklemiş olduğu ham veriler ile işlenmiş verileri görüntüleyerek, arzu ederse araçta yer almasını anlamlı bulmadığı verileri silebilir. Ancak işlenmiş verilere kaynak olarak kullanılan veri kaynaklarının silinmesine izin

verilmemektedir. Bu sayede, araştırmanın ilerleyen aşamalarında karşılaşılabilecek diğer sorunlar engellenmiş olacaktır.



The screenshot shows the 'Remove/Edit Source' window of the Outbreak Detection Tool. The window has a menu bar with 'Data Source', 'Preprocessing Ops', 'Methods', 'Welcome test', and 'Sign Out'. Below the menu bar is a table with the following columns: Name, Timestamp, Start Date, Finish Date, Mean, Standard Dev, Remove, Description, and Smoothing Definition. The table contains four rows of data sources, each with a 'Remove' button in red text.

Name	Timestamp	Start Date	Finish Date	Mean	Standard Dev	Remove	Description	Smoothing Definition
ED-train-01.csv	20/12/2012	01/01/2101	31/12/2105	86.043	20.672	Remove	Emergency Dep.	Root Source
OTC-train-01.csv	20/12/2012	01/01/2101	31/12/2105	583.197	143.054	Remove	Over the Count	Root Source
TH-train-01.csv	20/12/2012	01/01/2101	31/12/2105	14.225	7.370	Remove	Nurse Call	Root Source
ed_1_moving_average7.csv	20/12/2012	01/01/2101	31/12/2105	86.061	13.298	Remove		Extended Source:Moving Average is accomplish parameters : residual : false method : Moving Average window size : 7 reaches toRoot Source

Şekil 4-6 Veri Listeleme / Silme Ekranı

4.3.2. Veri Ön İşleme İşlemleri

Yapılan çalışmalardan elde edilen bulgular göstermiştir ki, süveyans sistemlerinde toplanan verilerin, *SPC* yöntemlerinin tamamında doğrudan kullanılabilir olması mümkün değildir. Hasta davranışlarının getirdiği bazı tanımlanabilir ve tanımlanamaz belirtiler söz konusu olabilmektedir. Konu ile ilgili daha ayrıntılı bir inceleme kesim 3.'te yer almaktadır. İncelenen yöntemler, bu tezin kapsamı olan salgın tespitini hedefleyen araştırmalarda yaygın olarak kullanılan yöntemlerdir. Bu yöntemlerden elde edilen veriler iki şekilde değerlendirilir. Birincisi düzleştirme ile elde edilen verinin tespit yöntemlerine sokulması, bir diğeri ise düzleştirme işleminden elde edilen tahmin değerinin gözlemden çıkarılarak, artık değerlerin içinde salgın tespitinin araştırılmasıdır. Bu tez kapsamında her iki değerlendirmeye de imkân sağlanırken, kullanıcı dilediği yolu izleyerek bir sonraki aşama olan tespit aşamasında ürettiği yeni kaynağı kullanabilir. Kullanıcı, ön işleme yönteminin başarımını, kesim 3.7'de anlatılan ölçütlere göre karşılaştırarak, sonuçları yorumlayabilecektir.

4.3.2.1. Basit Doğrusal Regresyon Ekranı

Basit doğrusal regresyon çözümlemesinin yapılabilmesi için, iki farklı veri kümesine ihtiyaç duyulur. Bunlardan birincisi, regresyonun uygulanacağı beklenen süveyans verisidir. İkinci veri kümesi ise, süveyans veri kümesine olan etkisi araştırılan regresör veri kümesidir. Kullanıcı, belirlediği tarih aralığında(veri

kümeleri izin verdiği sürece), regresyon ilişkisini toplanır/çarpımsal olarak araştırabilir. Elde edilen sonuçlar kendisi için anlamlı ise, sürveyans veri kümesinin işlenmesi için bu regresyon çözümlemesinin sonucunu kullanır.

Outbreak Detection Tool

Data Source ▾ Preprocessing Ops ▾ Methods ▾ Welcome fest Sign Out

Simple Linear Regression

Select Data Source - Regressand y.csv First Day of Data : 01.Oca.2012
Last Day of Data : 16.Oca.2012

Define day interval : 01.Oca.2012 16.Oca.2012

Select Data Source - Regressor b1.csv First Day of Data : 01.Oca.2012
Last Day of Data : 16.Oca.2012

REGRESSION Summary:

Slope: 0.005	Slope Std Error: 0.000	Slope Confidence Interval: 0.001
Intercept: 10.591	Intercept Std Error 0.032	Sum Squared Error 0.002

Multiplicative Run Regression

Smoothing Step (Run Regression Before)

Extended Source Name simpleReg_y_b1

Extended Source Description

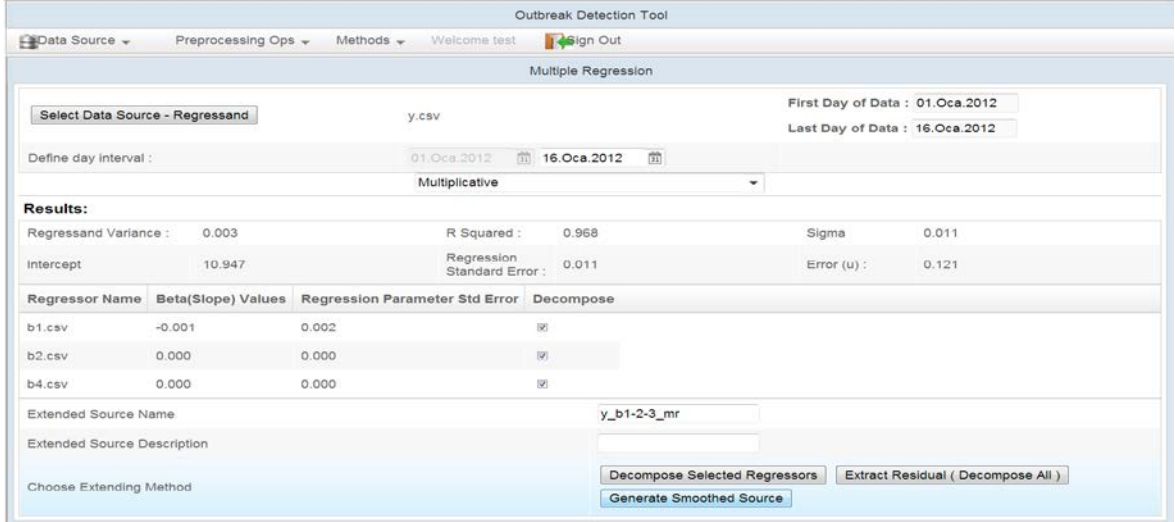
Choose Extending Method (Create new data source bounced with concurred days) Extract Residual Generate Smoothed Source

Şekil 4-7 Basit Doğrusal Regresyon Uygulama Ekranı

4.3.2.2. Çoklu Doğrusal Regresyon Ekranı

Basit doğrusal regresyonda yalnızca bir regresörün sürveyans verisi üzerindeki etkisi araştırılabilmektedirken, çoklu doğrusal regresyon ile birden çok regresör kaynağının tek sürveyans verisini nasıl etkilediği incelenebilir. Kullanıcı, basit doğrusal regresyonda izlediği yolları izleyerek çoklu doğrusal regresyon çözümlemesini yapabilir ve anlamlı bulduğu ilişkilerin sürveyans veri kümesinden çıkarılmasını sağlayabilir.

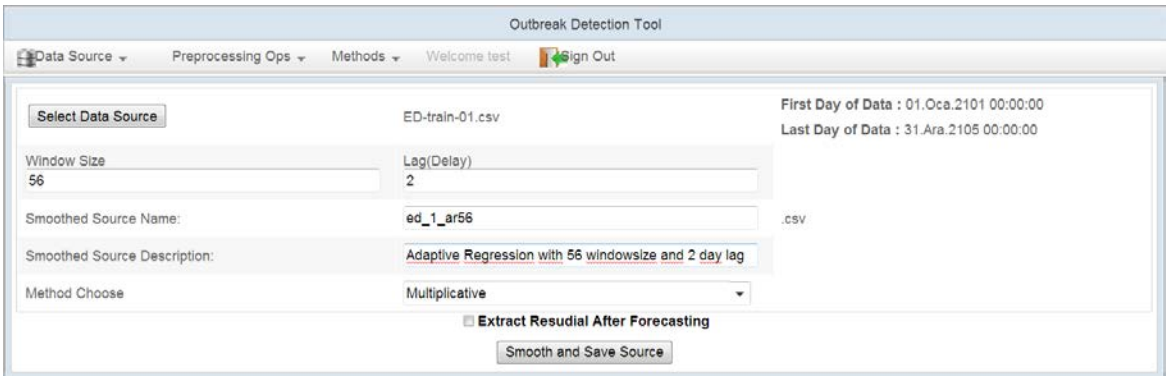
Çoklu doğrusal regresyon, araç ile gerçekleştirilebilen haftanın günü etkisi ayrıştırma, mevsimsel döngü etkisi ayrıştırma ya da “adaptive regression” ön işleme yöntemi uygulanırken süreç gereği çağırılırken, kullanıcının kendi araştırmak istediği ilişkileri çıkarmak için de kullanılabilir.



Şekil 4-8 Çoklu Doğrusal Regresyon Uygulama Ekranı

4.3.2.3. Adaptive Regression Ekranı

"Adaptive Regression", Burkom'un biyosürveyans için ortaya attığı bir yöntemdir. Bu yöntem, verilen pencere boyu sınırları içinde kalacak olan gözlem sayısının zamanla olan ilişkisini ortaya koyar ve her hesaplamada yeni bir çözümlene yapar. Bu da, defalarca tekrar hesaplanan küçük regresyon çözümleneleri demektir. Bu tez kapsamında da, "adaptive regression" ele alınmış ve ihtiyaç duyulan esneklik değişkenler ile arayüzden sağlanmıştır. Bu değişkenlerden birincisi pencere boyu, ikincisi ise göz ardı edilmek istenen gecikmedir.



Şekil 4-9 "Adaptive Regression" Uygulama Ekranı

Eğer bir salgın yavaş yayılıyorsa, bunun tespit edilmesi oldukça güçtür. Bir de kullanılan veri işleme yöntemi bu yükselişi zamanla ilişkilendirirse, salgını verinin içinde tamamen kaybetmiş olacaktır. Bunun önüne geçmek için bir gecikme/tampon belirlenmiştir. EARS yöntemlerinden C2'de aynı sebepten iki

günlük bir gecikme öngörüldüğü için bu araçta da varsayılan geciktirme iki olarak belirlenmiş olup, kullanıcı tarafından değiştirilmesi mümkündür.

4.3.2.4. 7 Gün Farkı Ekranı

7 gün farkı, uygulaması ve kullanımı oldukça kolay bir veri işleme yöntemidir. Bu yöntem ile 7 gün önce gerçekleşen gözlemden bugüne ne kadarlık bir artışın olduğu ortaya konulmaktadır. Kullanıcı, ön işleme tabi tutmak istediği kaynağı seçerek işlemi gerçekleştirir.

The screenshot shows the 'Outbreak Detection Tool' interface. The top navigation bar includes 'Data Source', 'Preprocessing Ops', 'Methods', 'Welcome test', and 'Sign Out'. The main content area is divided into several sections: a 'Select Data Source' button, a text field containing 'ED-train-01.csv', and two date fields: 'First Day of Data : 01.Oca.2101 00:00:00' and 'Last Day of Data : 31.Ara.2105 00:00:00'. Below these is a 'Smoothed Source Name' field with 'ed_1_7dif' and a '.csv' extension. A 'Smoothed Source Description' field is empty. A checkbox labeled 'Extract Residual After Forecasting' is checked. At the bottom, there is a 'Smooth and Save Source' button.

Şekil 4-10 7 Gün Farkı Uygulama Ekranı

4.3.2.5. Hareketli Ortalama Ekranı

Hareketli ortalama ile verilen pencere boyu büyüklüğünde bir serinin ortalamasının alınması sağlanır. Hareketli ortalama pencerede yer alan en yakın zamana ait veri ile en uzun zaman önce gözlenmiş değerlerin eşit ağırlıkta olması gerçeğinin göz ardı edilmemesi gerekmektedir. Bu durum, hareketli ortalama ile işlenmiş verilerin tepki zamanını uzatacak bir etki yaratır.

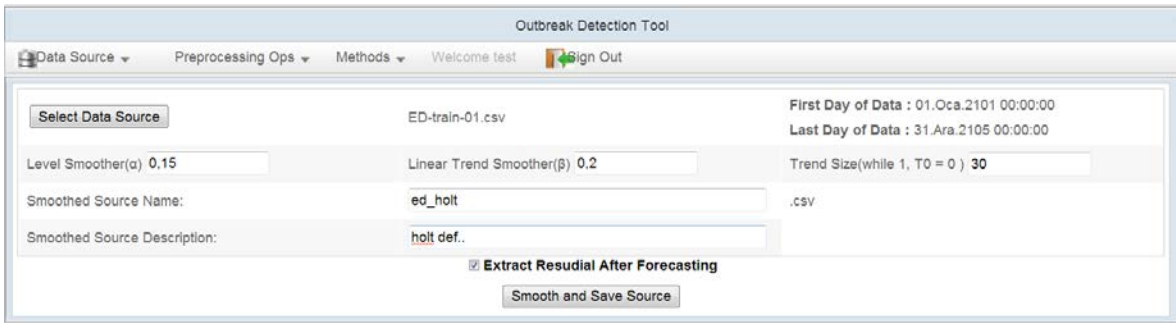
The screenshot shows the 'Outbreak Detection Tool' interface. The top navigation bar is the same as in the previous screenshot. The main content area includes the same 'Select Data Source' button and 'ED-train-01.csv' text field. The date fields are also present. The 'Smoothed Source Name' field now contains 'ed_1_wma'. The 'Moving Average Method' dropdown menu is set to 'Weighted Moving Average'. The 'Moving Average Window Size' field contains the number '7'. The 'Smoothed Source Description' field is empty. The 'Extract Residual After Forecasting' checkbox is unchecked. The 'Smooth and Save Source' button is highlighted in blue.

Şekil 4-11 Hareketli Ortalama Uygulama Ekranı

Ağırlıklı hareketli ortalama, hareketli ortalama gözlemlenen bu eksiği gideren özellikleri barındırır. Ağırlıklı hareketli ortalama ile yakın zamana ait ölçütlerin katsayıları artırılarak değişikliklerin daha hızlı geri bildirim yaratması sağlanır.

4.3.2.6. Holt Ekranı

Bu yöntem ile kullanıcı bir zaman serisinden, gözlemin yapıldığını döneme ait eğilimleri ayıklayabilir. Böylece verideki anlamlı eğilimler, tarif edilemeyen diğer etkenlerden ayırt edilir. Kullanıcı, doğrusal eğilim ile düzgünleştirme katsayılarını (α , β) ve doğrusal eğilimin aranacağı pencere boyunu ekrandan girerek veri düzleştirme işlemini gerçekleştirebilir.



The screenshot shows the 'Outbreak Detection Tool' interface. The main window has a menu bar with 'Data Source', 'Preprocessing Ops', 'Methods', 'Welcome test', and 'Sign Out'. Below the menu bar, there is a 'Select Data Source' button and a text field containing 'ED-train-01.csv'. To the right, there are two text fields: 'First Day of Data : 01.Oca.2101 00:00:00' and 'Last Day of Data : 31.Ara.2105 00:00:00'. Below these, there are three input fields: 'Level Smoother(α) 0,15', 'Linear Trend Smoother(β) 0,2', and 'Trend Size(while 1, T0 = 0) 30'. There are also two text fields for 'Smoothed Source Name:' (containing 'ed_holt') and 'Smoothed Source Description:' (containing 'holt def.'). A checkbox labeled 'Extract Residual After Forecasting' is checked. At the bottom, there is a 'Smooth and Save Source' button.

Şekil 4-12 Holt Uygulama Ekranı

4.3.2.7. Holt&Winters Ekranı

Holt&Winters düzleştirme yöntemi, Holt ile araştırılan etkilere ek olarak, döngüsel olarak tekrar edildiği varsayılan etkileri ayıklamak için kullanılır. Holt yöntemi için istenen iki düzleştirme katsayısı ile birlikte mevsimsel etki katsayısı da kullanıcı tarafından tanımlanır(γ). Kullanıcı eğer haftalık, aylık ya da yıllık döngüler ile döngü içindeki yerinin zaman serisine etkisini incelemek isterse, sırası ile 7, 30 veya 365 günlük pencere boyu ile HW'yi veri kümesine uygulayabilir. Bunun yanı sıra, kullanıcı oluşan dönemsel etkinin(mevsimsel etkinin), çarpımsal ya da toplanır olarak incelemesini, geliştirilen araç ile yapabilir. Lotze, sürveyans verileri için çarpımsal etkinin daha doğru bir seçim olduğunu belirtmiştir [71].

Her veri kümesinin farklı öznelikleri olduğu düşünülürse, seçilecek olan veri kümesi için düzleştirme katsayıları, uygulanan yöntemin başarımında önemli bir yere sahiptir. Burkom, düzleştirme katsayıları için $\alpha = 0,4$; $\beta=0$; $\gamma=0,15$ ve $\alpha = 0,1$; $\beta=0$; $\gamma=0,15$ katsayılarını tercih etmiştir [9]. Görüldüğü gibi, her iki tercih için de eğilimin olmadığını öngörmüştür. Ancak her zaman serisinin kendi karakteri vardır

ve bu yüzden daha doğru ve genel çözümlerden söz etmek doğru olmayabilir. HW parametrelerinin seçilmesiyle ilgili Chatfield'in çalışması incelenebilir [72].

The screenshot displays the 'Outbreak Detection Tool' interface. At the top, there are navigation tabs: 'Data Source', 'Preprocessing Ops', 'Methods', 'Welcome test', and 'Sign Out'. The main area is a configuration form for the Holt-Winter method. It includes a 'Select Data Source' button, a text field for 'ED-train-01.csv', and two date fields: 'First Day of Data : 01.Oca.2101 00:00:00' and 'Last Day of Data : 31.Ara.2105 00:00:00'. Below these are several input fields and dropdown menus: 'Complete Cycle(Period) Size(365 for year)' set to 365, 'Level Initialization' set to 'Average', 'Seasonal Indices Initialization' set to 'Regression Based', and 'Method Choose' set to 'Multiplicative'. There are also three numerical input fields: 'Level Smoother(α) 0.2', 'Linear Trend Smoother(β) 0.15', and 'Seasonal Smoother(γ) 0.2'. A 'Smoothed Source Name' field contains 'ed_1_hw_365' and a file extension '.csv'. A 'Smoothed Source Description' field is empty. A checkbox labeled 'Extract Residual After Forecasting' is checked. At the bottom, there is a 'Smooth and Save Source' button.

Şekil 4-13 Holt&Winter Düzleştirme Yöntemi Uygulama Ekranı

İklendirme yöntemi de ön işlemenin başarımını doğrudan etkilemektedir. Bu yüzden kullanıcı kendisi için doğru olanı tercih etmenin yanında, farklı iklendirme seçenekleri ile başarımları hem test edebilir, hem de daha doğru iklendirme parametrelerini belirleyerek, tespit yöntemlerine daha iyi ayıklanmış bir zaman serisi sunabilir.

4.3.3. Salgın Tespit İşlemleri

Araç ile salgın tespit yöntemlerinden herhangi biri bir veri kümesine uygulandığında, elde edilen sonuç da, csv uzantılı olarak kayıt altına alınmaktadır. Bu sayede, araştırmacıların sonuçların gün gün nasıl değiştiğini görüp yorumlayabilmesi hedeflenmektedir.

Yöntemlerden *HLM* dışında tamamı için otomatik olarak eşik değeri belirleme imkânı araç tarafından kullanıcılara sağlanmaktadır. Bu özelliği sayesinde, kullanıcı hata kabul edilebilirliğini eşitlediği eşik değerlerine sahip yöntemlerin kıyaslanmasını daha kolay yapabilir.

150	06/28/1994,86.241,0	
151	06/29/1994,86.493,0	
152	06/30/1994,86.857,0	
153	07/01/1994,87.484,0	
154	07/02/1994,88.375,0	
155	07/03/1994,89.503,0	
156	07/04/1994,90.894,0	
157	07/05/1994,92.582,0	
158	07/06/1994,94.607,0	
159	07/07/1994,96.933,0	Salgın araştırılan tarih
160	07/08/1994,99.543,0	
161	07/09/1994,102.438,0	Salgın Durumu (0-yok / 1-var)
162	07/10/1994,105.579,0	
163	07/11/1994,108.995,0	Yöntemden çıkan sonuç
164	07/12/1994,112.587,0	
165	07/13/1994,116.290,1	
166	07/14/1994,119.907,1	

Şekil 4-14 Salgın Tespit yönteminden elde edilen sonuçların kaydedildiği csv uzantılı kütük biçimi

Kullanıcı, herhangi bir salgın dönemi içermeyen veri kümesini, ister yöntemi değerlendirirken kullanacağı veri kümesinin salgın içermeyen bir bölümünü kullanarak, ister yeni ve salgın içermeyen bir veri kümesi seçerek, belirlediği *ATFS* değeri ile tölare ederek eşik değeri tanımlar. Eşik değeri belirlenirken eğer kullanıcı salgın içeren bir veri kümesi tercih etmişse, eşik değer yine hesaplanır ancak belirlenen *ATFS*'nin güvenilirliği tartışılır bir hal alır.

Auto Threshold x

ATFS(Average Time to First Signal) :

Source: Use same source Select another source to train

Enter a number for days of data per cent of data

Şekil 4-15 Otomatik Eşik Değeri Oluşturma Ekranı

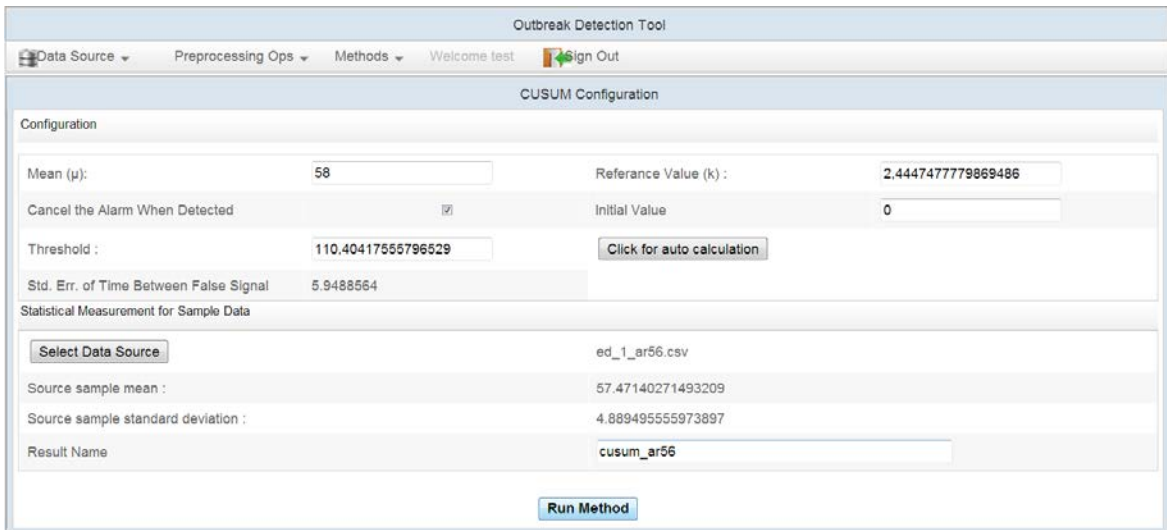
ATFS hesaplanırken, eşik değeri 1'den başlayarak, girilen parametreler ile yöntem test edilir. Test sonucunda, eğer eşik değeri kabul edilebilir hata oranından küçük bir hataya sebep oluyorsa daha küçük bir eşik değeri, büyük bir hata oranına sebep oluyorsa daha büyük bir eşik değeri ile tekrar test edilir. Bu şekilde *ATFS* istenen seviyeye gelene kadar ayarlama yapılır. *ATFS* değerinin eşit çıkacağı bir aralık meydana gelecektir. Bu durumda da, uyarıların oluşması için geçen sürenin standart hatası alınır. En düşük seviyede tutulan bu değere göre eşik belirlenir ve

eşitlenen şartlardan emin olunduktan sonra, yöntemler için karşılaştırma kolaylaşmış olur. Eğer standart hata 1'in altına düşerse ya da tanımlanan yeni eşik ile eski eşik arasında 1/1.000'den daha küçük bir fark kalırsa, yani artık bir iyileşme mümkün olmayacaksa, araç bu eşik değerini belirlenen *ATFS* için en uygun olarak tavsiye eder. Ayrıca, tavsiye edilen *ATFS* için, uyarılar arasında geçen zamanların standart hatası da bilgi olarak kullanıcıya sunulur.

Kullanıcı eşik değerini otomatik olarak yarattıktan sonra eğer bir parametrede değişiklik yapmak isterse, araç tarafından önerilen eşik değerini yeniden hesaplatmalıdır.

4.3.3.1. CUSUM Yöntemi Ekranı

Salgın tespit yöntemlerinden yaygın olarak tercih edilen yöntemlerinden biri olan Cusum, birikimli olarak büyüyen bir sonuç üretir. Bu sonuç tanımlanan eşığı aştığı anda salgın uyarısı oluşur. Bu andan itibaren, eğer kullanıcı yöntemin başlangıç konumuna dönmesi için "*Cancel the Alarm When Detected*" seçeneğini işaretlememişse uyarının kendi kendine ortadan kalkması oldukça güçtür. Bu yüzden bu tercihin seçili olması daha doğru bir kullanım olacaktır. Eğer uyarı üretilmişse ve başlangıç konumuna dönmesi için gerekli seçim yapılmışsa, *Cusum* sonucu tekrar 0'a döndürülür.



The screenshot shows the 'Outbreak Detection Tool' interface with the 'CUSUM Configuration' window open. The window is divided into several sections:

- Configuration:**
 - Mean (μ): 58
 - Reference Value (k): 2.4447477779869486
 - Cancel the Alarm When Detected:
 - Initial Value: 0
 - Threshold: 110.40417555796529 (with a 'Click for auto calculation' button)
 - Std. Err. of Time Between False Signal: 5.9488564
- Statistical Measurement for Sample Data:**
 - Select Data Source: ed_1_ar56.csv
 - Source sample mean: 57.47140271493209
 - Source sample standard deviation: 4.889495555973897
 - Result Name: cusum_ar56
- Run Method:** A button at the bottom of the window.

Şekil 4-16 Cusum Test Ekranı

Cusum ile birlikte "*adaptive regression*"ın kullanıldığı araştırmalar [73] [42], bu birlikteliğin başarımının oldukça yüksek olduğu göstermiştir.

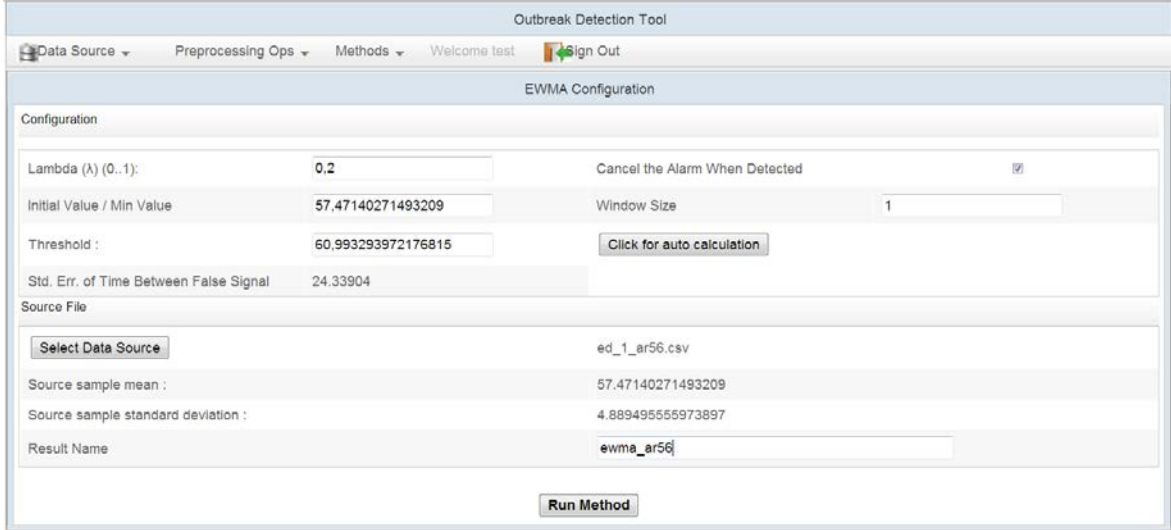
Çizelge 4-1, Fricker [38] tarafından Hawkins&Olwell'ın araştırmasından [23] uyarlanmış olan, seçilecek olan kaynak değer ve eşik değerinde oluşacak *ATFS* değerlerini göstermektedir. Şekil 4-16'te gösterilen *Cusum* ekranı ile kullanıcının yapacağı kaynak değer(k) tercihi ve *ATFS* seçimine göre eşik değerinin belirlenmesi araç tarafından sağlanmaktadır. Kullanıcı tarafından değiştirilmediği sürece *kaynak değer*, seçilen veri kümesinin standart sapmasının yarısı olarak atanır.

Çizelge 4-1 k ve h değerleri ile $ATFS$ arasındaki ilişkiyi gösterir çizelge. Fricker [38] tarafından Hawkins&Olwell'in araştırmasından [23] uyarlanmıştır.

k	0.00	0.25	0.50	0.75	1.00	1.25	1.50
h							
1.000	4.75	7.0	11.2	19.2	35.3	68.9	142.2
1.125	5.27	8.0	13.2	23.4	44.8	91.4	196.8
1.250	5.84	9.1	15.4	28.6	57.2	122.1	274.9
1.375	6.44	10.3	18.0	34.9	73.1	164.0	387.2
1.500	7.09	11.6	21.1	42.6	93.8	221.5	549.7
1.625	7.76	13.0	24.6	52.0	120.7	300.5	786.0
1.750	8.48	14.6	28.6	63.5	155.5	409.4	1,130.8
1.875	9.22	16.3	33.3	77.4	200.5	559.4	
2.000	10.00	18.2	38.5	94.3	258.7	766.2	
2.125	10.81	20.2	44.6	114.9	333.8	1,051.0	
2.250	11.66	22.4	51.5	139.7	430.7		
2.375	12.53	24.7	59.3	169.7	555.5		
2.500	13.43	27.3	68.2	206.0	716.0		
2.625	14.37	30.0	78.3	249.7	922.2		
2.750	15.33	32.9	89.8	302.5	1,187.0		
2.875	16.32	36.1	102.8	366.1			
3.000	17.35	39.5	117.6	442.8			
3.125	18.41	43.1	134.4	535.3			
3.250	19.50	47.0	153.4	646.9			
3.375	20.61	51.2	175.0	781.4			
3.500	21.76	55.7	199.6	943.7			
3.625	22.95	60.5	227.4	1,139.4			
3.750	24.16	65.7	259.0				
3.875	25.40	71.2	294.8				
4.000	26.68	77.1	335.4				
4.125	27.99	83.4	381.4				
4.250	29.32	90.2	433.6				
4.375	30.69	97.4	492.8				
4.500	32.09	105.1	559.9				
4.625	33.53	113.4	636.0				
4.750	34.99	122.2	722.3				
4.875	36.48	131.6	820.1				
5.000	38.01	141.7	930.9				
5.125	39.57	152.4	1,056.5				
5.250	41.15	163.9					
5.375	42.77	176.1					
5.500	44.42	189.2					
5.625	46.11	203.2					
5.750	47.82	218.0					
5.875	49.56	233.9					
6.000	51.34	250.8					

4.3.3.2. EWMA Yöntemi Ekranı

EWMA hem *SPC* yöntemi, hem de ön işleme yöntemi olarak çeşitli araştırmalarca ele alınmaktayken, biyosürveyans araştırmalarında salgın tespit yöntemi olarak kullanıldığı gözlemlenmiştir. Bu tez kapsamında geliştirilen araçta *EWMA*, bir salgın tespit yöntemi olarak kullanılmaktadır.



The screenshot displays the 'EWMA Configuration' window of the 'Outbreak Detection Tool'. The window has a title bar with 'Outbreak Detection Tool' and a menu bar with 'Data Source', 'Preprocessing Ops', 'Methods', 'Welcome test', and 'Sign Out'. The main content area is titled 'EWMA Configuration' and contains a 'Configuration' section with the following fields: 'Lambda (λ) (0..1):' with value '0.2', 'Initial Value / Min Value' with value '57.47140271493209', 'Threshold :' with value '60.993293972176815', and 'Std. Err. of Time Between False Signal' with value '24.33904'. There is a 'Cancel the Alarm When Detected' checkbox which is checked, and a 'Window Size' field with value '1'. A 'Click for auto calculation' button is located next to the 'Threshold' field. Below the configuration fields is a 'Source File' section with a 'Select Data Source' button and the file name 'ed_1_ar56.csv'. The 'Source sample mean :' is '57.47140271493209' and the 'Source sample standard deviation :' is '4.889495555973897'. The 'Result Name' field contains 'ewma_ar56'. A 'Run Method' button is located at the bottom center of the window.

Şekil 4-17 EWMA Test Ekranı

EWMA parametreleri belirlenirken, işlenmiş ve işlenirken de artık değeri çıkarılmış veri kümesi kullanıldığında, başlangıç değerinin doğru seçilmesi gerekir. Eğer kullanılan veri kümesi ham ise, zaman serisinin salgın içermeyen yeterli büyüklükte bir dönemine ait ortalaması kullanılabilir. Aracın seçilen veri kümesine ait sunduğu hesaplar, salgın içerse de içermese de, tüm gözlemlerin ortalama ve standart sapmasıdır. Bu yüzden doğrudan bu değeri kullanmak sonuçları olumsuz yönde etkileyebilir. Buna karşın, eğer kullanıcı kullanmak istediği veri kümesini bir ön işlemeye tabi tutmuş ve artık değerlerin tutulduğu bu zaman serisini kullanıyorsa, başlangıç değerini 0 olarak seçmesi daha doğru olacaktır.

Pencere boyu araç tarafından varsayılan değeri 1 olarak tanımlanmıştır. Ancak, kullanıcı pencere boyunu değiştirerek, bu pencerenin ortalamasını *EWMA* sonucunu hesaplarken kullanabilir. Bu, verinin işlenmeden doğrudan kullanılmasına imkân sağlayan bir parametre olarak düşünülebilir.

EWMA sonucunun belirlenmesinde en kritik parametre, *EWMA* katsayısı olan α 'nın seçimidir. Eğer kullanıcı geçmiş verilerin ağırlığını artırmak istiyorsa, α

katsayısını düşük tutabilir. Kullanıcı α katsayısını yüksek tutarak son gözlemin salgın tespiti için ağırlığını artırabilir.

4.3.3.3. EARS Yöntemleri Ekranı

CDC tarafından geliştirilen, araştırmalar çerçevesinde düşünüldüğünde en popüler biyosürveyans sistemi olarak göze çarpan EARS'ın, referans olarak yaygın kullanılan üç yöntemi C1,C2 ve C3 araç tarafından tez kapsamında ele alınmıştır. C1 ve C2 yöntemleri kendi aralarında, araç tarafından sunulan esnekliği ile geçişli bir duruma gelmektedir. Bu yöntemler kesim 2.4'te açıklanmıştır fakat özetle ifade etmek gerekirse, C1 ve C2 arasındaki tek fark, C2'nin kullandığı pencerede iki günlük bir gecikme ile çalışmasıdır. Bu sayede yavaş yayılan salgınların daha keskin bir hale getirileceği varsayılmaktadır.

EARS kapsamında ele alınan yöntemlerin parametreleri sabit olarak tanımlanmıştır. Bu parametrelerin içinde eşik değeri de vardır. Ancak zaman serisine özgü bazı durumlar bu eşik değerinin çok daha büyük değerlerde olmasını gerektirebilmektedir. Bu yüzden otomatik eşik değeri belirleme yöntemi ile eşğin belirlenmesi kullanıcı için daha doğru olabilmektedir.

Outbreak Detection Tool

Data Source Preprocessing Ops Methods Welcome test Sign Out

EARS Method : C1, C2, C3, W2

Ready To Use Methods

Configuration

Day Lag : 0 Window Size: 7

Minimum SD : 0.2 Threshold: 4,86455140288

Cancel the Alarm When Detected Click for auto calculation

Std. Err. of Time Between False Signal 23.508863

Source File

Select Data Source ED-train-01.csv

Source sample mean : 86.043

Source sample standard deviation : 20.672

Result Name c1_ed_1

Run Method

Şekil 4-18 EARS Yöntemlerinin Test Ekranı

Son olarak, 0'a bölünme hatalarına karşın, kullanılan penceredeki serinin standart sapması için bir asgari değer tanımlanması yöntemi uygulanabilir. 0'a bölünme hatası çalışmayı durdurabilen ya da çalışmanın sağlıksız yürümesine sebep

olabilen bir çalışma zamanlı hatadır. Bu yüzden kullanıcının, “Minimum SD” alanını doldurması uygulamanın sürekliliği için gerekli olabilmektedir.

4.3.3.4. Shewhart Yöntemi Ekranı

Shewhart, SPC'nin en eski ve temel alınan yöntemidir. Shewhart ile bir zaman serisinde meydana gelen ani yükselişler hızlıca tespit edilebilmektedir. Biyolojik terörün ani bir salgın yayılımına sebep olabileceği varsayıldığında, bu eski ve basit yöntemin *Cusum* ve *EWMA*'dan daha başarılı olabileceği durumlar söz konusudur. Zaten CDC tarafından geliştirilen *EARS* yöntemleri de temel olarak Shewhart'ı kullanmaktadır.

The screenshot shows the 'Shewhart Configuration' window in the 'Outbreak Detection Tool'. The window has a menu bar with 'Data Source', 'Preprocessing Ops', 'Methods', 'Welcome test', and 'Sign Out'. The main area is titled 'Shewhart Configuration' and contains the following fields and controls:

Mean :	86	Cancel the Alarm When Detected	<input checked="" type="checkbox"/>
Standard Deviation :	20	Threshold :	1.94998031599181
Window Size :	1	<input type="button" value="Click for threshold auto calculation"/>	
Std. Err. of Time Between False Signal	3.559026		

Below the configuration fields is a 'Source File' section with a 'Select Data Source' button and the following information:

Source File	ED-train-01.csv
Source sample mean :	86.04271631982475
Source sample standard deviation :	20.671847948450704
Result Name	shewhart_ed1

At the bottom of the window is a 'Run Method' button.

Şekil 4-19 Shewhart Test Ekranı

Kullanıcı, Shewhart yöntemini kullanırken, beklenen değer, standart sapma ve pencere boyutunu belirtmelidir. Beklenen değer, salgın görülmeyen bir dönemin ortalaması olarak kullanılması durumunda daha doğru sonuçlar üretecektir.

Pencere boyu ise, Shewhart'ın ani yükselişleri hızlıca bulmasını engelleyecek, ancak yöntemin içinde bir düzeltirmeye imkân sağlayacak olan bir parametredir. Kullanıcı bu alanın 1 olarak varsayılan değerde olduğunu görecektir. Eğer kullanıcı dilerse pencere boyunu değiştirerek Shewhart yöntemini test edebilir.

4.3.3.5. NBC Yöntemi Ekranı

“*Negative Binomial Cumulative Sum*”, kalite denetim süreçlerinde hatayı asgari düzeyde tutan bir yöntem olarak kullanılmaktadır. Her ne kadar *Cusum*, Shewhart ya da *EWMA* kadar olmasa da, kalite denetim süreçlerinde zaman zaman

karşılaşılmaktadır. Biyosürveyans arařtırmalarında da birkaç arařtırmaya konu olarak, olumlu görüřler kaydedilmiřtir [40].

The screenshot displays the 'Negative Binomial CUSUM Configuration' window of the 'Outbreak Detection Tool'. The interface includes a navigation bar at the top with 'Data Source', 'Preprocessing Ops', 'Methods', 'Welcome test', and 'Sign Out' options. The main configuration area is divided into two sections: 'Configuration' and 'Source File'. The 'Configuration' section contains several input fields: 'Shift Tolerance' (2), 'Lag' (2), 'Window Size' (7), 'Std. Err. of Time Between False Signal' (39.564995), and 'Minimum SD' (0.2). There is also a 'Threshold' field with the value 56.947226231052184 and a checkbox for 'Cancel the Alarm When Detected' which is checked. A button labeled 'Click for threshold auto calculation' is present. The 'Source File' section shows a 'Select Data Source' button and the file path 'ED-train-01.csv'. Below this, the 'Source sample mean' is 86.04271631982475 and the 'Source sample standard deviation' is 20.671847948450704. The 'Result Name' field contains 'nbc_ed_1'. A 'Run Method' button is located at the bottom center of the window.

řekil 4-20 NBC Test Ekranı

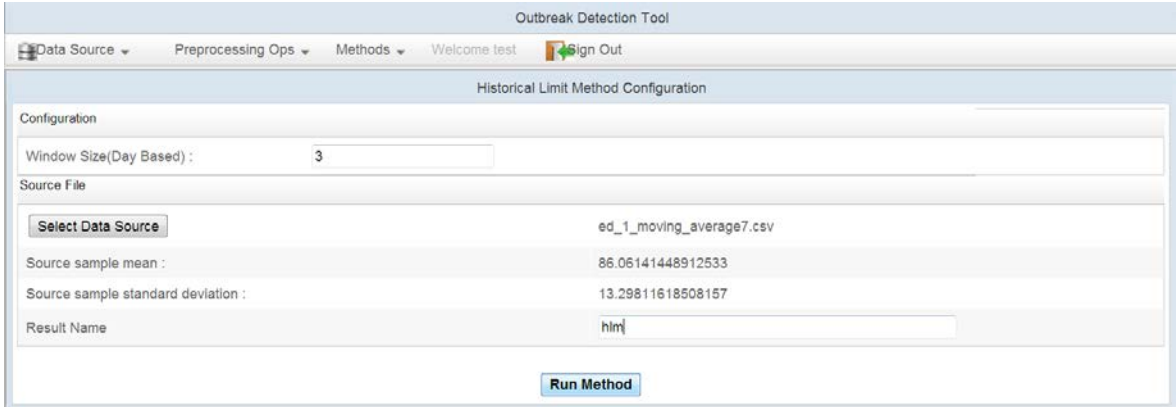
Kullanıcı, *NBC* yöntemi ile tespit yapmak için, ne kadarlık bir sapmaya göz yumabileceğini, gecikmeyi, pencere boyunu ve asgari standart sapma deęerini girerek *NBC* sonuçlarını gözlemleyebilir. Eřik için otomatik eřik belirleyici kullanması daha ölçülebilir sonuç alması için kullanıcının tercih etmesi gereken bir özelliktir. Yine de kullanıcı dilerse eřik deęerini doğrudan kendisi belirleyebilir.

Kullanıcının gireceęi sapma toleransı, gözlem miktarının ortalamadan kaç standart sapma kadar uzaklaşmasının kabul edilebileceğinin seçilmesidir. Kesim 2.5'te anlatılan c_1 deęeri bu parametreden hesaplanır.

4.3.3.6. HLM Yöntemi Ekranı

"*Historical Limits Method*", uzun bir veri geçmiři gerektiren bir tespit yöntemidir. Verilen pencere boyu için, yaşanan yılın ve önceki yılların istatistikleri kullanılır. *HLM*'yi yaptığı salgın arařtırmasında kullanan Pelecanos [45], haftalık olarak toplanmış verileri sabit olarak 5 yıllık bir geçmiři kullanarak hesaplamayı tercih etmiştir. Ancak günümüz sürveyans sistemleri, günlük verileri bile yeterince hızlı yanıt vermekte sorgular hale geldięi için, bu tez kapsamında haftalık deęil günlük verilerdeki başarımların hesaplanmasına olanak sağlanmaktadır. Ayrıca 5 yıllık süreç ile elde ettięi başarımların ne kadar yüksek olsa da, daha az geçmiři olan verileri kullanarak da başarımların ölçülmesi araç ile mümkün olabilmektedir.

Gerçekleştirilen *HLM* yönteminde, geçmiş veriler sonucu birikimli olarak etkilemektedir.



Şekil 4-21 HLM Test Ekranı

Eşitlik (2.19)'da verilen değerler, belirlenen pencerenin içinde kalan günlerin ortalaması ve standart sapması olarak düşünülmelidir. Örneğin 20 Ocak'ta yapılan gözlemin 5'lik bir pencere ile çözümlenmesi yapılırsa, 16 Ocak ile 20 Ocak arasında gerçekleşmiş olan gözlemlerin ortalaması x_0 , geçmiş yıllarda gerçekleşen 15 Ocak ile 25 Ocak arasında gerçekleşen gözlemlerin ortalama ve standart sapması μ ve σ_x olarak kullanılır.

Son olarak, *HLM* yönteminde eşik değeri kullanılmamaktadır. *HLM* yöntemi ile bir sonuç hesabı yapılmamakta, bir eşitlik test edilmektedir. Bu yüzden, herhangi bir eşik belirleme söz konusu olmamaktadır.

4.3.4. Örnek Süreç ve Sonuçları Görüntüleme

Bu kesimde, bir zaman serisinin sisteme yüklenmesinden, bu seri üzerinde yapılacak işlemlerin sonunda elde edilen başarımların görüntülenmesine kadar geçen sürecin bir örneği verilecektir. Şekil 4-22 ile bir sürecin araç ile nasıl modelleneceğine ait akış çizeneği verilmiştir. Kullanıcı sisteme giriş yaptıktan sonra eğer sınamak istediği veri kümesini daha önce yüklememişse veri ekleme ekranını kullanarak yükler. Seçtiği veri kümesine uygulayacağı algoritmaya göre, eğer model ön işleme ihtiyacı duyuyorsa, tanımlanmış olan yöntemlerden birini ya da birkaçını veri kümesine uygulayabilir. Her uygulamadan elde ettiği yeni veri kümesini, üzerindeki değişimi ve ön işleme yönteminin başarımlarını raporlayabilir (*Report#1*). Kullanıcı eğer veride araştırılan örüntünün çarpımsal olmasını dilerse,

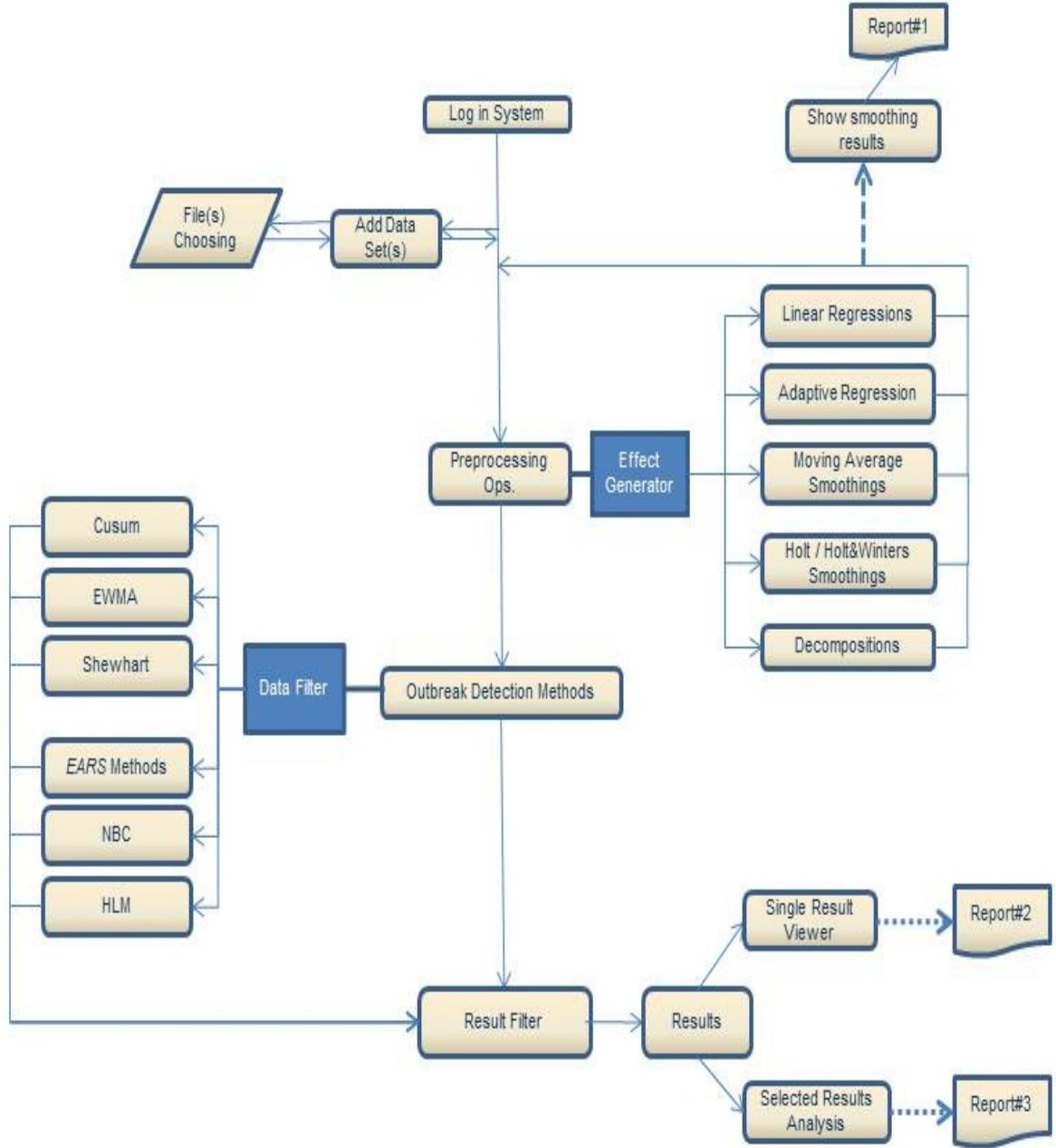
araç veri kümesini işlemeden önce *Effect Generator* ile doğal logaritmasını alır ve hesaplamaların ardından tersine işleme tabi tutar. Kullanıcı bir sonraki aşamada salgın tespit yöntemini seçerek modelinin ikinci aşamasını da kurgular. Bu noktada kullanıcı dilerse farklı parametrelerle aynı yöntemi sınavabilir ya da benzer parametrelerle farklı yöntemleri sınavabilir. 0'a bölünme durumunun önüne geçmek için kullanıcının onayı ile asgari standart sapma tanımlama, gözlem sayısı 0 olan değerleri tamamlama gibi seçenekler işleme alınır. Gözlem sayısının 0 olarak kaydedildiği günler için 7 gün farkı düzleştirme yöntemi uygulanır ve tam 1 hafta önce kaydedilen gözlem miktarı tekrar edilmiş gibi işlem yapılır. Kullanıcı salgın tespit yönteminin veriye uygulanması ile ilgilendiği sonuçları elde eder. Elde ettiği eski ve yeni sonuçları süzgeçten geçirerek, filtrelenen verilerin kıyaslanmasını görüntüler. Bu noktada kullanıcı iki farklı rapor alabilir. Bunlardan birincisi, uygulanan algorithmadan elde edilen sonuçların, zaman serisi olarak, eşik değeri ile bir çizelgede gösterimi (*Report#2*), diğeri ise seçilen sonuçların XY çizelgesidir (*Report#3*). XY çizelgesinde X ve Y koordinatlarına *sensitivity*, *specificity*, *false alert rate* ölçütleri ile *window size* tercihinden ikisi yerleştirilebilir. Daha önceki kesimlerde anlatılan ROC, AMOC gibi eğriler ve *AUC* hesabı bu seçimlere göre oluşur. Kullanıcı elde edilen XY eğrisinin, tanımladığı aralık için, altında kalan alanını da ekranda görüntüleyebilecek ve kişisel raporunda kullanabilecektir.

Örnek süreçte iki farklı veri kümesi üzerinde çalışılacaktır. Bunlardan birincisi *CDC* tarafından yayınlanan bir veri kümesidir. Hem salgın içermeyen hali ile hem de salgın içeren hali ile bu veri kümesi birlikte kullanılacaktır. Salgın içermeyen hali, *ATFS* ile eşik belirleme maksatlı kullanılacakken, salgın içeren hali ile tespit yöntemleri test edilecektir. Bir diğeri veri kümesi ise, *ISDS*²¹ tarafından sağlanan veri kümelerindedir.

Öncelikle, *CDC* veri kümeleri salgın tespit aracımıza uygun kütük biçimine dönüştürülmüştür. *ISDS* verisi ise uygun biçimde olduğu için herhangi bir yükleme öncesi düzenleme yapılmamıştır. Veriler uygun şekilde düzenlendikten sonra araç veri ekleme ekranı kullanılarak yükleme yapılmıştır. Yükleme işleminde veriler farklı ön işleme yöntemleri ile işlenerek araştırılan örüntülerin veriden çıkarılması

²¹ The International Society for Disease Surveillance

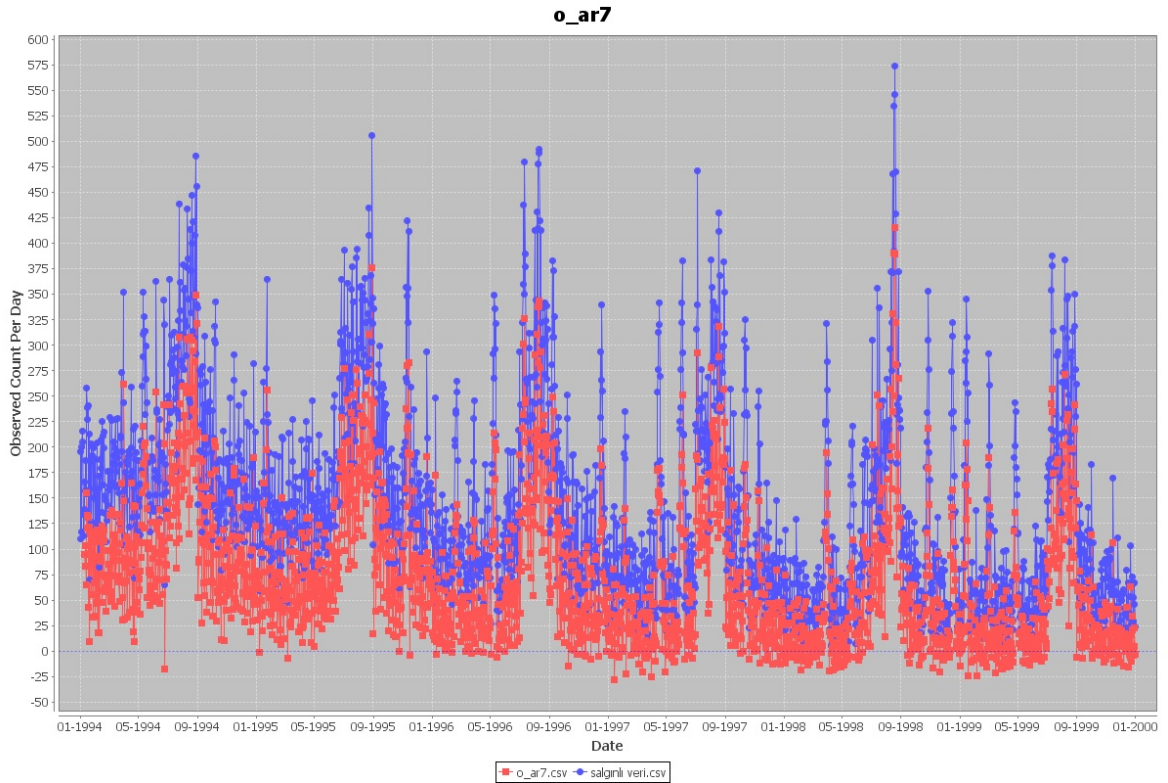
sağlanmaya çalışılmıştır. Veri ön işleme ile verinin bir örüntüden tamamen kurtarılması sağlanabilir ya da örüntüler toplamı düzleştirilebilir. Veri ön işleme yöntemlerinin başarımını belirleyen ölçütler düzleştirme yöntemleri için daha anlamlı sonuçlar verirler. Ancak eğer örüntü veriden çıkartılırsa, bu ölçütlerden elde edilecek ölçümler gerçek değerlerden uzaklaşacağı için sağlanan başarımları doğrudan yansıtamayabilirler.



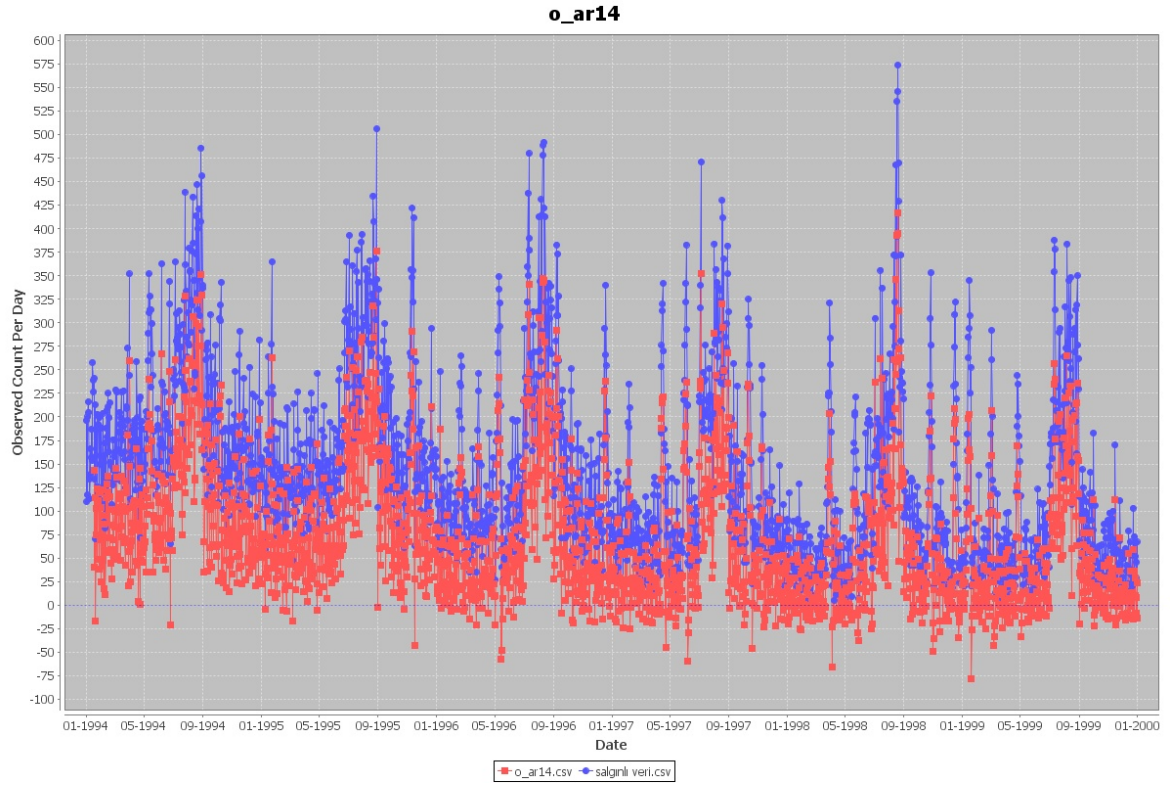
Şekil 4-22 Araç ile bir Sürecin Modellenmesi

Bu kesimde verilen çözümler, temsili olarak seçilmiş ön işleme ve teslim yöntemlerinin bir bütünlük içinde ele alınmasını kapsar. Elde edilen sonuçlardan daha iyileri, farklı seçimlerle sağlanabilecekken, örnek çalışmanın amacı, bir araştırmacının sırasıyla, bir zaman serisini düzleştirme yöntemi ile nasıl işleyeceğinin anlatılması ve bu işlenmiş yeni zaman serisini kullanarak yöntemlerde test edilerek sonuçların görüntülenmesidir. Araştırmacı, elde ettiği sonuçlar ile ara sonuçların istatistiklerini ve görsellerini yorumlayabilecektir.

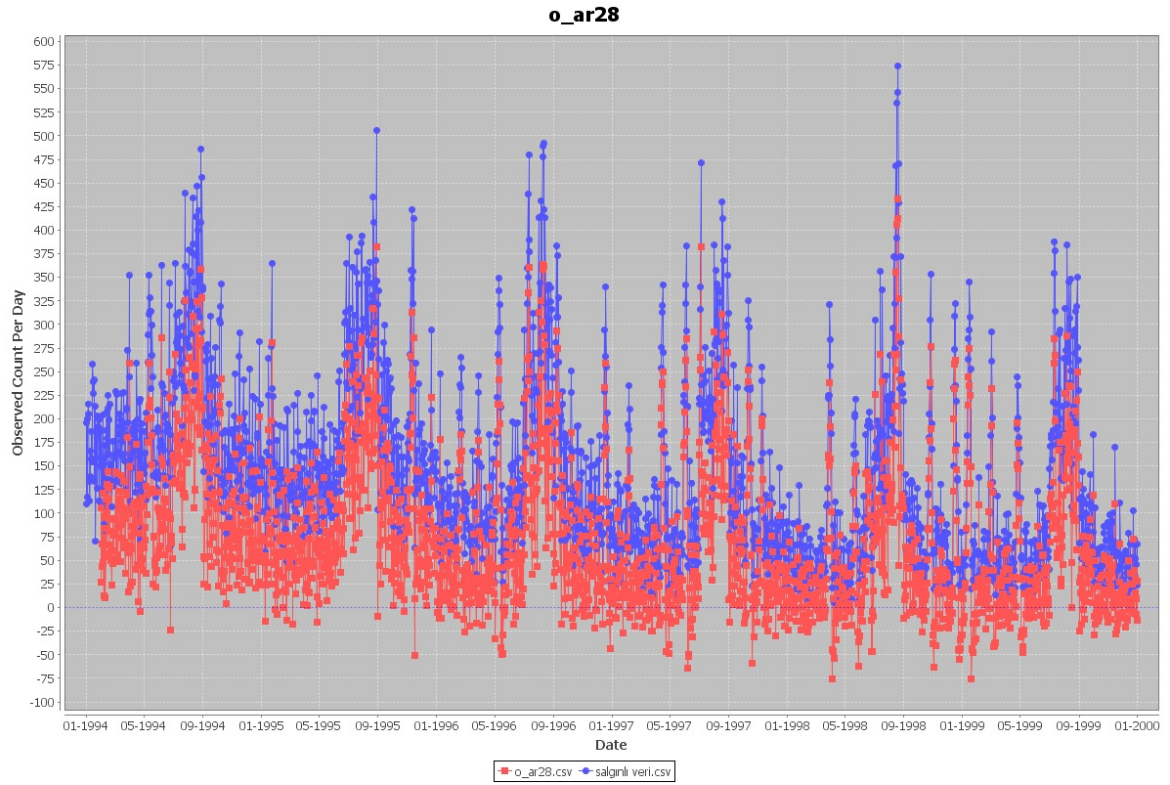
Verilerin yüklenmesinin ardından, "adaptive regression" veri ön işleme yöntemi ile yüklenen her üç zaman serisi de işlenmiştir. Pencere boyu olarak, 7, 14, 28, 56 ve 84 belirlenmiştir. Bunun yanı sıra, bu zaman serilerine haftanın günü etkisi ile mevsimsel etkinin ayrışımı, HW düzleştirme yöntemiyle mevsimsel etkinin ayrıştırılması işlemleri uygulanmıştır. Örnek sürecin verildiği bu kesimde yalnızca salgın içeren CDC kaynağından elde edilen sonuçlar verilecektir. Kullanılan kaynağın ortalaması 144,53, standart sapması ise 96,04'tür.



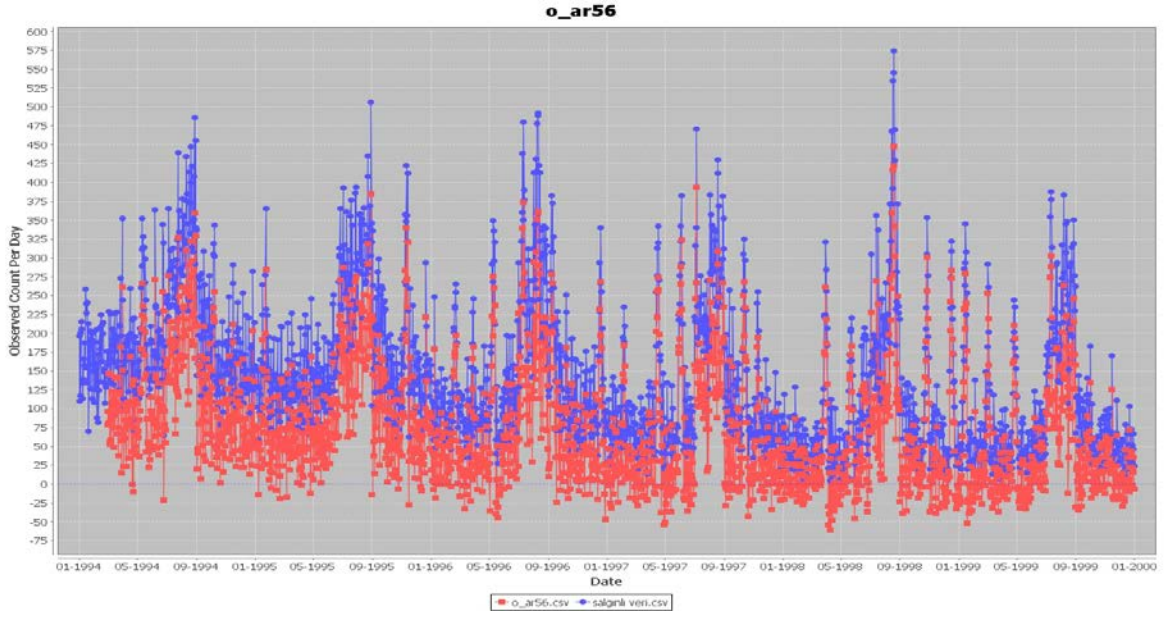
Şekil 4-23 CDC tarafından sağlanan veriye pencere boyu 7 olan "adaptive regression" uygulandığında elde edilen sonuç grafiği.



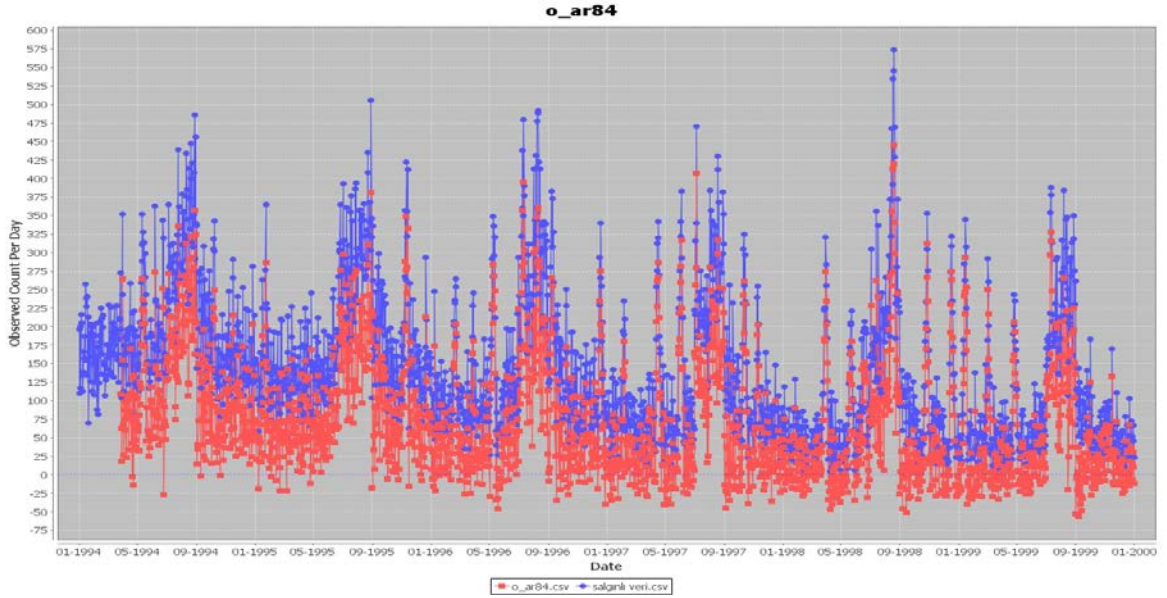
Şekil 4-24 CDC tarafından sağlanan veriye pencere boyu 14 olan "adaptive regression" uygulandığında elde edilen sonuç grafiği



Şekil 4-25 CDC tarafından sağlanan veriye pencere boyu 28 olan "adaptive regression" uygulandığında elde edilen sonuç grafiği



Şekil 4-26 CDC tarafından sağlanan veriye pencere boyu 56 olan "adaptive regression" uygulandığında elde edilen sonuç grafiği



Şekil 4-27 CDC tarafından sağlanan veriye pencere boyu 84 olan "adaptive regression" uygulandığında elde edilen sonuç grafiği

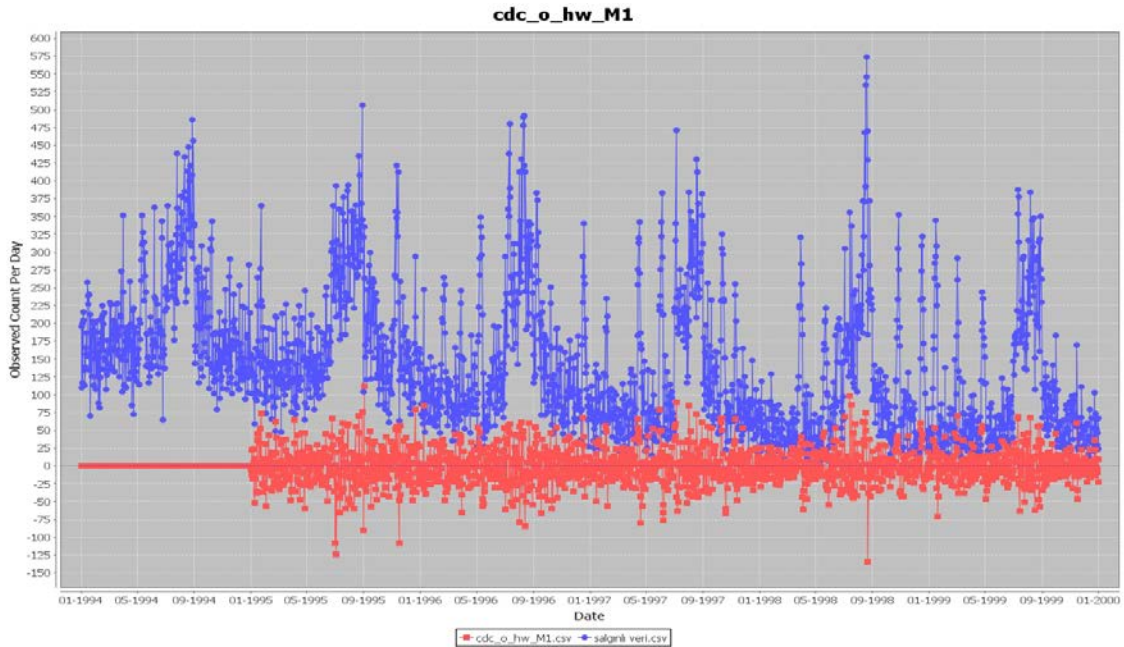
Şekil 4-23, Şekil 4-24, Şekil 4-25, Şekil 4-26 ve Şekil 4-27'te verilen grafiklerde kaynak veri kümesi mavi, işlenmiş veri kümesi ise kırmızı ile gösterilmiştir. Düzleştirme işlemi *çarpımsal* yöntem kullanılarak yapılmıştır. Bu şartlarda elde edilen istatistikler Çizelge 4-2'de verilmektedir. Buradaki değerler tercih edilen yöntemin iyi bir tahmin yöntemi olmadığını işaret etmektedir. Ancak burada amaç

en iyi tahmin yöntemini seçmek değil, bir sonraki adımda uygulanacak olan tespit yöntemi ile birlikte en iyi sonucu verecek olan düzleştirme yöntemini seçmektir. Tahmin yapmakta başarılı olan bir yöntemin, doğrudan salgın tespitinde kullanılmasının doğru bir yol olduğuna dair herhangi bir çalışma bulunmamaktadır. Bu belirsizlik, ele alınması gereken bir çalışma alanı olarak düşünülebilir.

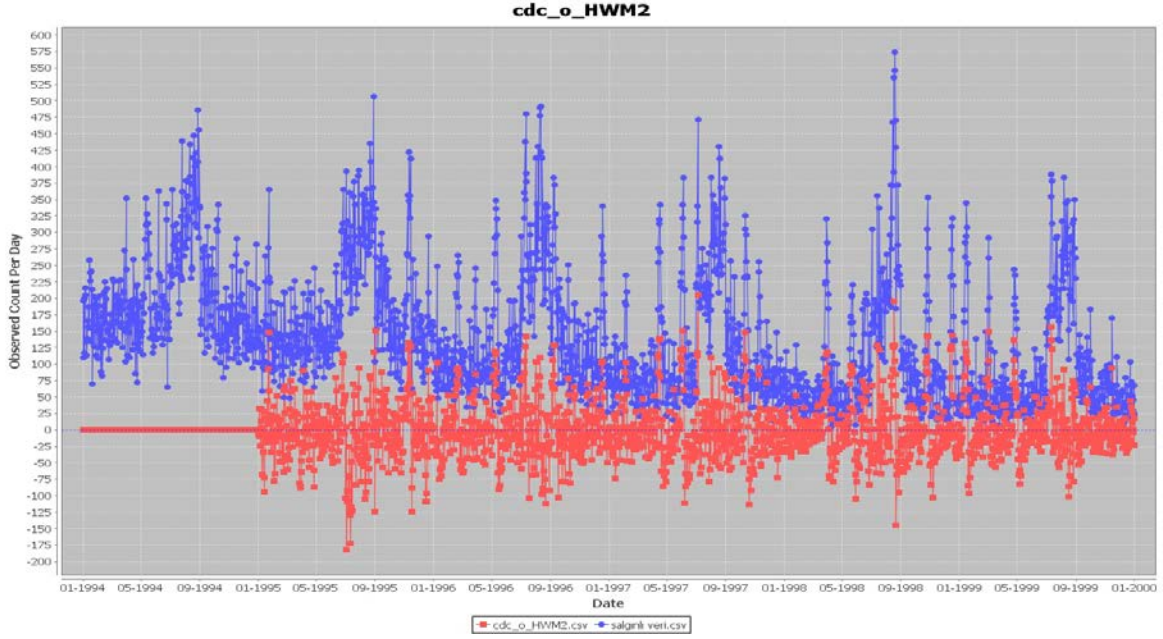
Çizelge 4-2 "Adaptive Regression" Ön işleme sonuçları

İstatistik	Ortalama	Standart Sapma	MSE	RMSE	MdAD	MdAPE
Pencere Boyu						
7	72,35	30,27	10063,24	100,32	41,31	-79,81
14	71,93	29,09	10563,38	102,78	42,51	-80,53
28	71,21	26,69	11245,81	106,04	44,43	-78,49
56	70,78	25,84	11460,74	107,05	46,24	-74,86
84	70,65	26,45	11469,51	107,1	46,49	-72,34

"Adaptive Regression" uygulandıktan ve sonuçlar alındıktan sonra, aynı ham veri kümesine HW yöntemi Burkom tarafından önerilen [9] her iki düzleştirme parametreleri kümesi ile de uygulanmıştır. Çarpımsal olarak uygulanan HW işleminden elde edilen sonuçlar Şekil 4-28 ve Şekil 4-29 ile gösterilmiştir.



Şekil 4-28 CDC tarafından sağlanan veriye HW $\alpha = 0,4$; $\beta=0$; $\gamma=0,15$ düzleştirme katsayıları ile uygulandığında elde edilen sonuç grafiği



Şekil 4-29 CDC tarafından sağlanan veriye HW $\alpha = 0,1$; $\beta=0$; $\gamma=0,15$ düzleştirme katsayıları ile uygulandığında elde edilen sonuç grafiği

Her iki düzleştirme uygulamasında da kullanılan başlangıç değerleri için aynı parametreler tercih edilmiştir. Temel düzey belirlenmesi için, belirlenen dönemin ortalaması alınmış, *mevsimsel etkiler* ise regresyon ile hesaplanmıştır.

Görüldüğü gibi, HW ilk bir yıl herhangi bir düzleştirme uygulamamış, işlenmiş veri sürekli olarak 0 kaydedilmiştir. HW yönteminde mevsimsel etkiyi ortadan kaldırmak için belirlenen pencere boyu, günlük bir zaman serisinde, bir yılda kaydedilen gözlem sayısı olan 365'tir. HW zaman serisinin ilk dönemini öğrenme için kullandığı için, bu süreçte herhangi bir gözlem düzleştirmesi yapılamaz. Dönemsel etkiyi tanımlamada başarılı bir yöntem olan HW'yi eğer haftalık dönemleri incelemek için kullansaydık, 7 günlük bir öğrenme penceresi tercih edecektik ve yalnızca 7 günlük kayıt düzleştirme/ayırışım sonucu 0 olacaktı.

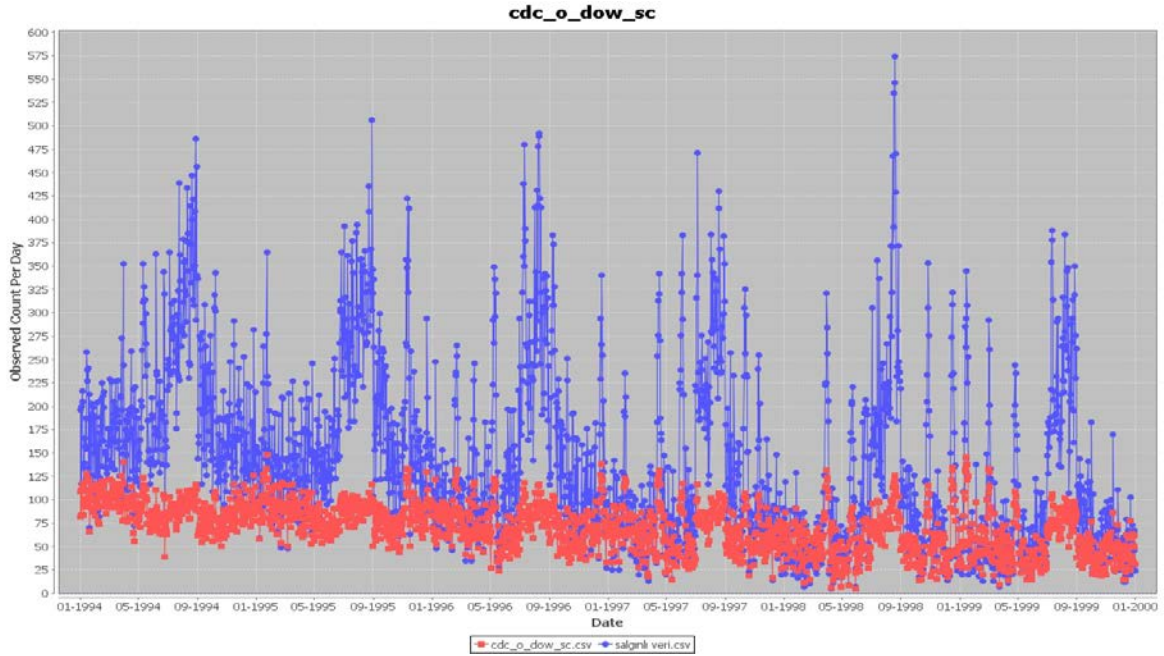
Çizelge 4-3 Holt&Winters Düzleştirme Yöntemleri Başarım İstatistikleri

İstatistik	Ortalama	Standart Sapma	MSE	RMSE	MdAD	MdAPE
$\alpha;\beta;\gamma$						
0,4;0;0,15	-0,18	23,18	537,18	23,18	10,62	6,24
0,1;0;0,15	-0,61	41,48	1720,53	41,48	18,95	15,13

Çizelge 4-3'te, iki farklı düzleştirme kümesi ile elde edilen başarım sonuçları gösterilmiştir. Yalnızca düzey değişkeni olan α için yapılan değişikliğin sonuçları

nasıl etkilediği gözlenebilmektedir. α için 0,4 test edildiğinde ortaya çıkan istatistikler, bu seçimin daha doğru olduğunu göstermektedir.

Örnek sürecin son ön işleme yöntemi olarak seçilen ayrıştırma işleminin sonuçları Şekil 4-30'de gösterilmektedir. Ayrıştırma işlemi, hem haftanın günü etkisi için hem de mevsimsel etki ayrıştırması için çarpımsal olarak uygulanmıştır. Haftanın günü etkisinin belirlenmesinde Pazar referans gün olarak tercih edilirken, mevsimsel etkinin kestirimde hem kosinüs hem de sinüs trigonometrik değişkenlerden yararlanılmıştır. Araştırmacı dilerse, daha dengeli bir düzleştirme için, ortalamaya daha yakın olduğunu varsaydığı bir başka günü referans gün olarak seçebilir. Referans gün, gözlem miktarlarına meydana gelen değişimlerde, haftanın günü etkisi tayini için normal kabul edilen gündür. Yani eğer Pazar günü referans olarak seçilmişse, diğer günlerin Pazar gününe göre zaman serisinde yol açtıkları değişim araştırılmış olur. Mevsimsel etki ayrıştırmasında ise, 1 Ekim orijin noktası olarak seçilmiştir. Yani, bir gözlemden meydana gelen mevsimsel etkilenmenin kestiriminin yapıldığı Eşitlik (2.33)'te yer alan t değeri, gözlem yapılan günün geriye doğru 1 Ekim'e olan uzaklığını vermektedir.



Şekil 4-30 CDC tarafından sağlanan veriye Haftanın Günü Etkisi Ayrıştırma ve Mevsimsel Ayrıştırma uygulandığında elde edilen sonuç grafiği

Şekil 4-30'te görüldüğü gibi, ayrışımı yapılmış olan verideki dalgalanmalar, ham veri kaynağından daha yumuşaktır. Ayrıştırma işlemine dair başarımlar istatistikleri

tutulmamaktadır. Çünkü ayrışım tekniğinin tahmin başarımlarını ölçütleri ile ölçülmesi anlamlı değildir.

Başlığın kalan kısmında, yukarıda tanımlanan ön işleme tabii tutulmuş zaman serisi ile işlenmemiş ham veriye uygulanan çeşitli tespit yöntemleri ve sonuçları elde edilen bulgular verilecektir. Elde edilen sonuçlar, hangi yöntemin bir diğerinden daha iyi olduğu sonucuna varmamıza değil, kullanılan zaman serisine uygulandığında hangi ön işleme ve tespit yöntemi ikilisinin daha iyi sonuç verdiğini ölçmemize yarayacaktır.

EARS yöntemleri sırasıyla 7, 14, 28, 56 ve 84 büyüklüğünde pencerelerle ve *ATFS* değerlerinin %0,5, %1, %2, %5 hata oranı sağladığı eşik değerleri ile test edilmiştir. *ATFS* değerlerini sabitleyerek eşik belirleme işlemleri yapılırken, hata kabul oranı büyüdükçe *ATFS* standart hatasının küçülmekte olduğu gözlemlenmiştir. Yani hata kabul oranı düştükçe, önerilen eşik değerinin güvenilirliği de düşmektedir. Salgın içeren *CDC* veri kümesi üzerinde tespit işlemi yapılırken, aynı veri kümesinden salgınlar çıkarılarak elde edilen veri kümesi *ATFS*'nin belirlenmesi için kullanılmıştır. Ayrıca, *ATFS*'ye göre belirlenen eşik değerlerinin *EARS*'in önerdiği eşiklerden daha büyük olduğu gözlemlenmiştir. Elbette ki *EARS*'in önerdiği daha düşük eşik değeri, tespit edilen salgın sayısını artıracaktır ancak yanlış oluşan uyarı sayısının da artacağı unutulmamalıdır. Amacımız bu örnek süreçte, eşik değerlerini sabitleyerek bir karşılaştırmayı örneklemek değil, *ATFS* değerlerini sabitleyerek karşılaştırmalar yapmaktır, bu yüzden *EARS* tarafından önerilen eşik değerleri bu örnek kapsamında değerlendirilmeye alınmamıştır.

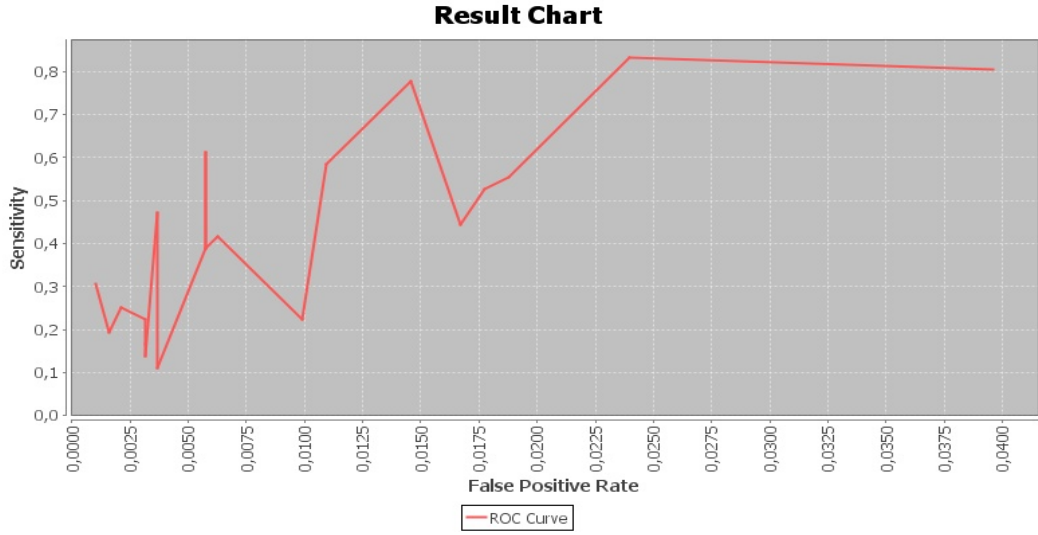
Çizelge 4-4'te *EARS* yöntemlerinin her biri için, *CDC*'den sağlanan ham veriye uygulandığında elde edilen toplam 20 koşul ve 3 yöntem için elde edilen 60 adet sonuca yer verilmiştir. Örnek olarak, pencere boyu ile *sensitivity* arasındaki ilişkiyi incelersek, *ATFS*'nin 20 olduğu durum hariç, tüm *ATFS* değerleri için *sensitivity* ölçütünün en iyi değerleri 28 pencere boyunda ölçülmüştür. Bir araştırmacı, *sensitivity* ölçümlerine göre tercih edecek olsaydı, pencere boyunun 28 olması gerektiği sonucuna varırdı. Bunun yanı sıra, C3 yönteminin genel olarak diğer iki yöntemden daha geç uyarı verdiğini bu çizelgeden söyleyebiliriz. C1 ve C2 yöntemleri yalnızca verinin düzleşmesine yol açacak olan büyük pencere

boyarında ge uyarı retirken, C3 yntemi hibir koşulda 1'in altına dşmemektedir.

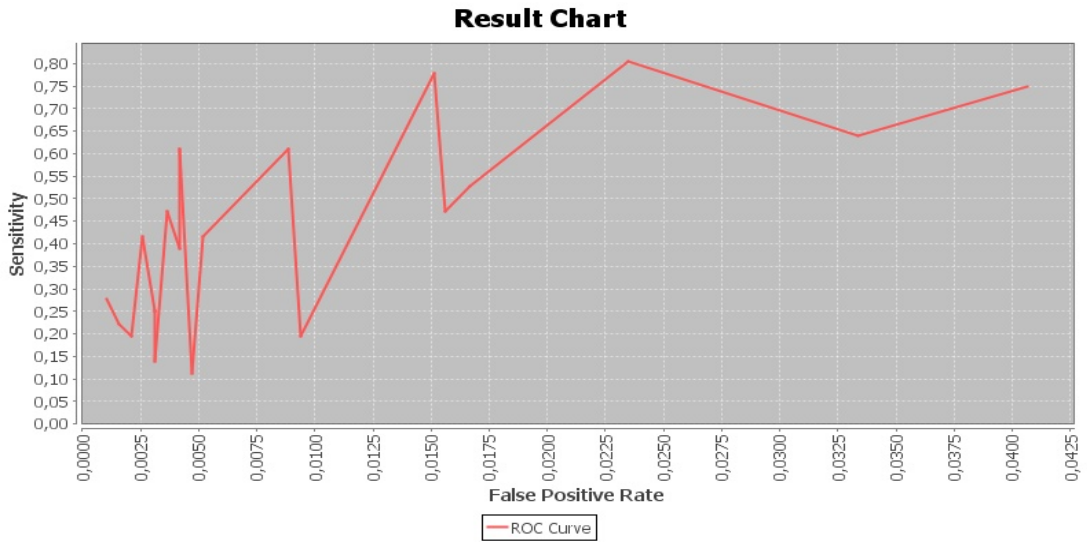
izelge 4-4 EARS Yntemlerinin Farklı Pencere Boyu ve ATFS deęerleriyle elde ettikleri bařarımları

ATFS (Gn)	Pencere (Gn)	Sensitivity (%)			Specificity (%)			Timeliness (Gn)		
		C1	C2	C3	C1	C2	C3	C1	C2	C3
200	7	11,1	11,1	27,8	99,6	99,5	99,9	0,3	0,3	1,8
	14	13,9	22,2	30,6	99,7	99,8	99,8	0,4	0,9	2,4
	28	30,6	27,8	27,8	99,7	99,8	99,8	1,5	1,7	2,4
	56	19,4	19,4	13,9	99,8	99,8	99,6	1,7	2,0	3,0
	84	16,7	13,9	13,9	99,7	99,7	99,6	1,5	1,2	3,6
100	7	22,2	19,4	44,4	99,0	99,1	99,8	0,1	0,1	1,3
	14	38,9	47,2	36,1	99,4	99,6	99,8	0,6	0,9	1,7
	28	47,2	41,7	38,9	99,6	99,7	99,8	1,8	1,6	2,7
	56	25,0	25,0	22,2	100,0	99,7	99,6	2,2	2,1	3,0
	84	22,2	25,0	19,4	99,7	99,7	99,5	2,0	2,0	3,6
50	7	44,4	47,2	55,6	98,3	98,4	99,3	0,6	0,4	1,2
	14	58,3	61,1	52,8	98,9	99,1	99,6	0,6	1,0	1,7
	28	61,1	61,1	50,0	99,4	99,6	99,7	1,3	1,5	2,4
	56	38,9	38,9	30,6	99,4	99,6	99,4	1,9	1,9	3,6
	84	41,7	41,7	33,3	99,4	99,5	99,4	2,3	2,3	3,4
20	7	80,6	75,0	75,0	96,0	95,9	97,8	0,4	0,4	1,2
	14	83,3	80,6	69,4	97,6	97,7	99,2	0,6	0,7	1,5
	28	77,8	77,8	66,7	98,5	98,5	99,3	1,0	1,0	2,2
	56	55,6	52,8	52,8	98,1	98,3	98,8	1,7	1,7	2,8
	84	52,8	63,9	50,0	98,2	96,7	98,9	1,9	1,3	2,6

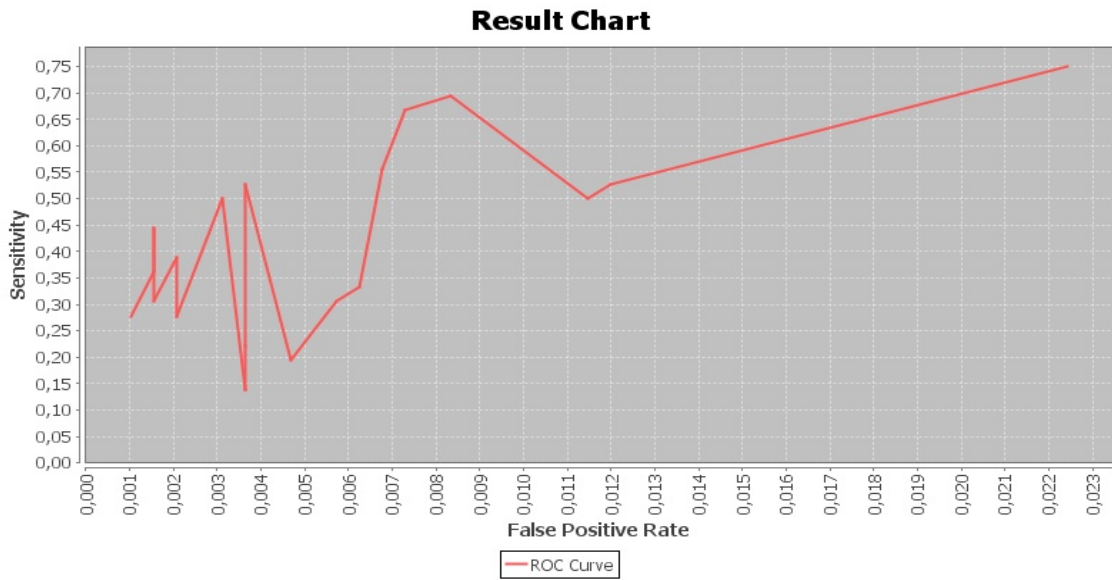
Őekil 4-31, Őekil 4-32 ve Őekil 4-33 C1, C2 ve C3'n ROC eęrileridir. Bu eęrilerin altında kalan alanlar bařarım SPC iin geleneksel bir bařarım ltdr. zellikle C1 ve C2 yntemleri birbirlerine ok yakın sonular vermiŐlerdir. Ancak C1 yntemi az da olsa daha iyi bir genel bařarım saęlamıŐtır. C3 ise her iki yntemin de gerisinde kalan bařarım gstermiŐtir. C3 ynteminin *timeliness* ltndeki lmlerinin de dięer iki yntemden daha geride olduęu dŐnldęnde, kullanılan zaman serisi iin C3 ynteminin dięer iki yntemden daha bařarısız olduęu sylenbilir.



Şekil 4-31 C1 yöntemine ait ROC eğrisi. $AUC=0,025$ olarak çıkmıştır



Şekil 4-32 C2 yöntemine ait ROC eğrisi. $AUC=0,024$ olarak çıkmıştır



Şekil 4-33 C3 yöntemine ait ROC eğrisi. $AUC=0,012$ olarak çıkmıştır

Araştırmacı dilerse, araç ile *sensitivity* ve *specificity* ölçütleri dışında *timeliness* ölçütü ya da pencere boyunu da istenen eksene yerleştirebilir. Pencere boyu seçildiğinde, araç tarafından tespit yönteminde pencere aranır. Eğer tespit yönteminde pencere boyu yoksa kullanılan zaman serisine uygulanan ön işleme yönteminde pencere boyu aranır. Eğer zaman serisine herhangi bir ön işleme uygulanmadıysa ya da kullanılan düzleştirme yönteminde pencere kullanılmadıysa, seçilen sonuç nesnesini grafiğe dâhil etmez.

Yukarıda *EARS* yöntemleri için uygulanan pencere boyu ve *ATFS* değişkenlerinin, aynı zaman serisi üzerinde *Cusum* için de uygulanmasıyla Çizelge 4-5'teki sonuçlar elde edilmiştir. *Cusum*'da kullanılan pencere boyu, Fricker'in araştırmasında [42] kullandığı gibi, Burkom'un önerdiği [9] "*Adaptive Regression*" ön işleme yönteminin pencere boyudur. *Cusum*'un "*adaptive regression*" ile test edilmesi sırasında, *EARS* yöntemlerinde test edilen pencere boyları ham veri kümesine "*Adaptive Regression*" uygulanırken, test edilen *ATFS* değerleri ise test istatistiği olan eşik değerleri hesaplanırken kullanılmıştır.

Cusum parametreleri belirlenirken, k değeri için $\frac{1}{2}\sigma_{\varepsilon} \frac{(n+1) \times (n+2)}{n(n-1)}$ formülünün verdiği sonuç kullanılmıştır. Eşitlik (2.1)'deki hedef değer olarak, salgın aranılan zaman serisi ile aynı ön işlemeden geçirilmiş olan, ancak salgın içermeyen halinin ortalaması kullanılmıştır. Başlangıç değeri olarak 0 kullanılmıştır. Ayrıca yöntem salgın tespit ettiğinde başlangıç değerine dönerek birikimli olarak test sonucunun yükselmesi engellenmiştir.

Bu şekilde test edildiğinde elde edilen sonuçlar gösteriyor ki, test edilen zaman serisi için, *Adaptive Regression* ön işleme yöntemi ve *Cusum* test sonuçları, 100 ve 200 gibi yüksek *ATFS* değerleri için *EARS* yöntemleri kadar başarılı değildir. *ATFS*'nin 20 olan değeri için birbirine yakın hata oranları çıkarken, doğru tespit miktarı daha kabul edilebilir değerlere yükselmektedir. Pencere boyunun 7 ve 14 olduğu durumlarda, *sensitivity* ölçümleri oldukça düşük bir başarımla sergilenmektedir. Test edilen beş pencere boyu içinde ise, tüm *ATFS* değerlerini içeren *ROC* eğrilerini değerlendirmeye aldığımızda, 56 ve 84 ile elde edilen *AUC* değerlerinin, diğer 3 pencere boyu içinde ölçülen değerlerden daha başarılı olduğunu söyleyebiliriz. Hem 56 hem de 84 pencere boyları için *ROC* eğrileri

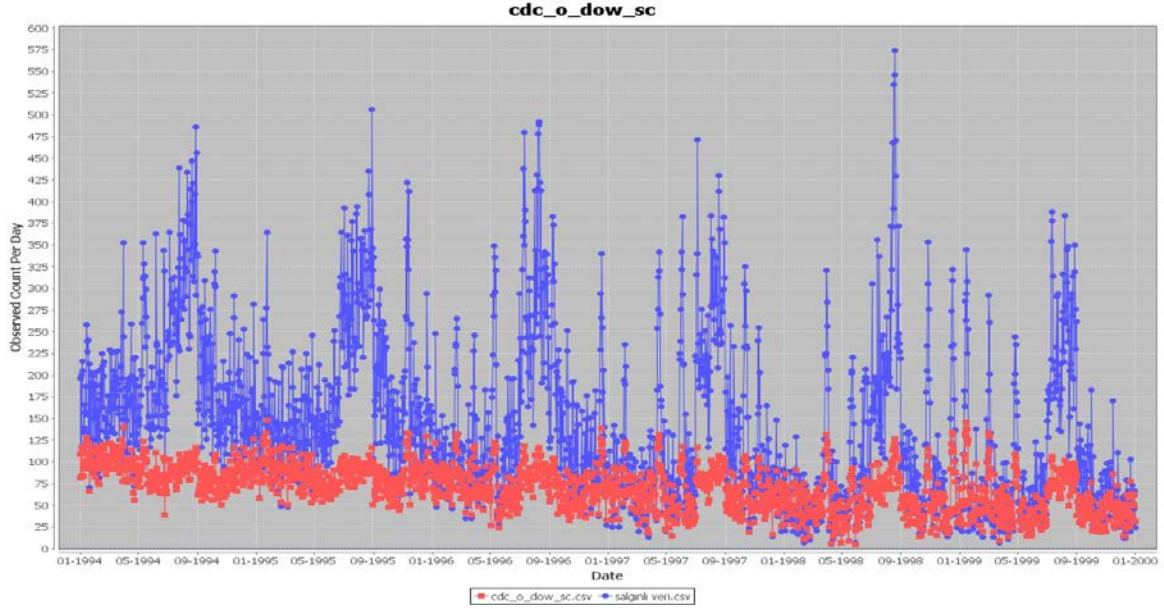
altında kalan alanlar(*AUC*) 0,01 olarak ölçülmüştür. Tüm koşulların *AUC* değerlerinin toplamı ise 0,012 olarak hesaplanmıştır.

Çizelge 4-5 *Cusum* Yönteminin Farklı Pencere Boyu ve *ATFS* değerleriyle elde ettikleri başarımları

ATFS (Gün)	Pencere (Gün)	Sensitivity (%)	Specificity (%)	Timeliness (Gün)
200	7	2,8	99,7	9,0
	14	5,6	99,7	5,5
	28	2,8	99,7	7,0
	56	2,8	99,7	6,0
	84	2,8	99,7	6,0
100	7	2,8	99,2	6,0
	14	11,1	99,4	7,3
	28	11,1	99,4	6,0
	56	8,3	99,4	5,3
	84	8,3	99,3	5,3
50	7	22,2	98,6	4,9
	14	16,7	98,6	3,5
	28	27,8	98,7	4,9
	56	41,7	98,9	5,7
	84	44,4	99,0	5,3
20	7	41,7	96,8	3,7
	14	52,8	97,0	3,0
	28	72,2	97,1	3,5
	56	75,0	97,4	3,3
	84	72,2	97,3	3,5

C1, C2, C3 ve *Cusum*'a ait *ROC* eğrilerinin altında kalan alanların toplamlarını kıyasladığımızda, C2'nin kullanılan veri kümesi için en iyi sonuçları verdiğini, onu C1, C3 ve *Cusum*'ün takip ettiğini söyleyebiliriz. *ATFS* değerlerinin yüksek olduğu durumlarda *Cusum*'ün başarımının *EARS* yöntemlerine nazaran çok daha kötü olduğunu gözlemlemiştik. Ölçümlerle de *Cusum*'ün genel başarımının *EARS* yöntemlerinden kötü olduğunu desteklemiş olduk. *ATFS* değerlerinin düşük, pencere boylarının ise 28, 56 ya da 84 olarak tercih edileceği durumlar için ölçümleri tekrarlarsak, dört yöntem için toplam altışar ölçümü hesaba katmamız gerekir. Bu durumda *AUC* değerleri C1, C2, C3 ve *Cusum* için sıralaması C2>*Cusum*>C1>C3 olmaktadır. Eski ölçümlerde oldukça gerilerde kalan *Cusum*'ün bu kez C1 ve C3'ten daha başarılı sonuçlar ürettiğini görmekteyiz.

Ancak C2, her iki kısıttaki kıyaslama için de en iyi sonuçları vermektedir. *Specificity* ölçümlerinin farklı sınırları arasındaki başarımlar alındığında farklı sonuçlar elde etmek mümkündür. Özellikle C3'ün *specificity* ölçümlerinin diğer yöntemlere göre daha yüksek çıkması onu daha güvenilir kılmakla birlikte, *AUC* hesaplarının düşük çıktığı gözlemlenmiştir.



Şekil 4-34 CDC tarafından sağlanan veriye DOW ve Mevsimsel Örüntü Ayırıştırması Uygulandığında elde edilen sonuç grafiği

Örnek sürecimizin son yöntemi olarak *EWMA* kullanılmıştır. *EWMA*'nın uygulandığı veri kümesi, *EARS* yöntemleri ve *Cusum* için kullanılan veri kümesidir. Zaman serisine ön işleme yöntemi olarak *DOW* etkisi ve mevsimsel etki ayırıştırma işlemi uygulanmıştır. Pazar günü referans gün olarak seçilmiş ve sinüs-kosinüs bileşenlerinin her biri ayırıştırmaya dâhil edilmiştir. Çarpımsal etki araştırılarak ayırıştırma yapıldığı için, yapılan ayırıştırma işleminin eşitliği $\log Y_t = (\beta_0 + \beta_1 \times t) + \beta_3 \sin\left(\frac{2\pi}{365.25}\right)t + \beta_4 \cos\left(\frac{2\pi}{365.25}\right)t + \beta_{pzt}I_{pzt} + \beta_{sal}I_{sal} \dots + \beta_{cts}I_{cts}$ olur. Uygulanan ayırıştırma işleminin ardından elde edilen düzleştirilmiş zaman serisi Şekil 4-34'de kırmızı çizimle gösterilmiştir. Mavi çizim ise ham veri kaynağının gösterimi içindir.

Elde edilen bu yeni işlenmiş veri kümesine farklı *ATFS* değerleri ile tanımlanan eşikler *EWMA* katsayısı $\alpha=0,3$ olacak şekilde uygulanmıştır. Çizelge 4-6 *EWMA* Yönteminin Farklı *ATFS* değerleriyle elde ettiği başarımlarıda, uygulanan *EWMA* tespit yönteminin değerlendirildiği veri kümesi üzerindeki farklı *ATFS*'ler için elde

ettiği istatistikler verilmiştir. Şekil 4-35'te ise bu *ATFS* değerlerinde elde edilen sonuçların *ROC* eğrisi verilmiştir. *AUC* yalnızca *Cusum*'dan daha iyi çıktığı gözlemlenmiştir. Ne var ki, diğer yöntemlerin 20 farklı durumla sınanmasına karşın *EWMA* yöntemi yalnızca 4 koşulda ölçüldüğü unutulmamalıdır.

Çizelge 4-6 *EWMA* Yönteminin Farklı *ATFS* değerleriyle elde ettiği başarımları

ATFS (Gün)	Sensitivity (%)	Specificity (%)	Timeliness (Gün)
200	22,2	99,9	5,5
100	36,1	99,8	4,7
50	44,4	99,2	3,7
20	63,9	97,8	3,0

Result Chart



Şekil 4-35 *EWMA* yöntemine ait *ROC* eğrisi. *AUC*=0,01 olarak çıkmıştır

EWMA hata oranının düşük tutulduğu durumlarda en iyi sonuçlara yakın bir ölçüm sunmaktadır. Ancak yine de mevcut parametreler ile ne yüksek hassasiyet, ne de güvenilir uyarı koşulları için en iyi sonucu verememektedir.

Örnek süreçlerde, *EARS* yöntemleri ile *Cusum* ve *EWMA* yöntemlerinin, *CDC* tarafından benzetimle üretilen tek bir zaman serisi üzerinde çözümlenmeleri yapılmıştır. Test edilen zaman serisi üzerinde, *CDC* tarafından geliştirilen *EARS* yöntemi olan *C2*, genel olarak diğer test edilen yöntemlerden daha başarılı sonuçlar elde etmiştir. Fakat yine de, farklı parametrelerin çok farklı sonuçlara sebep olduğu gözlemlenmiştir. Örneğin *EWMA* yönteminde ağırlık çarpanı olarak 0,3 değil de 0,2 test edilmiş olsaydı, *EWMA* için daha kötü sonuçlar alınacaktı. Eğer başarımları *timeliness* ölçütü üzerinden devam ettirseydik farklı sonuçlara varabilirdik. Bir çözümlene haritası belirleyip, araç ile bu çözümlenin uygulanmasını göstermek hedeflendiği için en iyi parametrelerin seçilmesi ikinci planda tutulmuştur. Elde edilen sonuçların *AUC* hesaplamaları yapılırken, bağımsız değişken(örnekte *specificity*) olarak kullanılan ölçütün sınırlandırılması karşılaştırmanın daha sağlıklı olmasını sağlayacaktır. Aksi durumda, bağımsız değişkende meydana gelen değişkenlik, *AUC* değerinin daha yüksek çıkmasına ve algoritmanın başarımlarının hatalı yorumlanmasına sebep olabilir.

Geliştirilen araç, tez konusu gereği, biyosürveyans araştırmaları için üretilmiş bir veri kaynağı ile örneklenmiştir. Ancak çalışma genelinde defalarca tekrar edildiği gibi, tanımlanan veri işleme yöntemleri ve aykırı durum tespit yöntemleri yalnızca salgın tespiti için kullanılan yöntemler değil, aynı zamanda süreç takibi, kalite denetimi gibi endüstriyel amaçlarla yapılan araştırmalar için de kullanılan yöntemlerdir. Bu sebeple, geliştirilen araç ile endüstriyel bir sürecin takibi ve kalite koşullarında meydana gelen aykırılıklar da test edilebilir ve sonuçlar raporlanabilir.

5. SONUÇ VE ÖNERİLER

Sürveyans sistemleri, 21.yy içinde hızla yayılan ve görev tanımları daha da belirginleşen bir duruma gelmiştir. *WHO*, *OECD* ve *CDC* gibi kuruluşlar biyolojik teröre ve nüfusun önlenemez hareketliliğinden kaynaklanan doğal salgınların yayılmasına karşı yönergeler, bildirimler ve standartlar tanımlamaya yönelmişlerdir. Ne var ki, henüz uluslararası düzeyde bir sürveyans sistemi geliştirilmemiştir. Başta ABD olmak üzere Kanada, İngiltere, Avustralya, Hollanda gibi gelişmiş ülkeler kendi ulusal sürveyans sistemlerine sahipken, bunların da ulusal anlamda bütün ülkeyi gözlem altında tutan bir örneği yoktur. Mevcut sürveyans sistemlerinden bazıları tanınların coğrafi konumlarla ilişkilendirilmesine kadar ayrıntılı bir işlem kapasitesine sahipken, bazı sürveyans sistemleri yalnızca veri toplama işlemlerini yürütmektedir. Ülkemizde de veri toplama işlemleri yapılırken, bu verilerin gerçek ya da gerçeğe yakın zamanlı olarak işlenmesiyle ilgili çalışan herhangi bir sistem yoktur. Ancak kalkınma planında yer alması, ilerleyen dönemlerde ülkemizin de salgın tespitinde kullanılacak yeterlilikte bir sürveyans sisteminin geliştirilmesi konusunda olumlu bir adımdır.

Ülkemizde, yakın geçmişe kadar RSHMB tarafından, yalnızca tanımlı bakterilerin suş incelemelerinin yapıldığı bir sürveyans sistemi söz konusuysen, günümüzde ICD 10 kodları ile toplanan Sağlık Bakanlığı'na ait bir veri merkezi mevcuttur. Faturalandırma kaygılarından dolayı, ICD 10 kodları sisteme bildirilmektedir. Ancak bu kodlamaların gecikmeli ya da özensiz olarak bildirilmesi durumunda dengeli bir erken uyarı sistemi geliştirmek mümkün değildir. ICD 10 kodlarının hasta bilgileri ile girilerek MERNİS²²'ten alınacak coğrafi bilgilerle eşleştirilmesi sonucu, coğrafi olarak salgın risk haritasının oluşturulması teknik olarak mümkündür. 1957 de Uzak Doğu'da ortaya çıkan Asya gribi yalnızca 65 yaş üstü insanlar arasında, pandemi oluşturmuştur [74]. Daha genç insanlarda daha nadir olarak karşılaşılmış, bir tehdit oluşturmamıştır. Bu gibi durumlar için, hastalıkların yaş gruplarındaki etkinliği de salgın tespitine girdi yapılarak, gerçek zamanlı, CBS²³ ile bütünleşik, yaş ve cinsiyet gibi kategorik koşullara uygun salgın risk haritasını oluşturacak bir sistemin teknik altyapısı ülkemizde mevcuttur. Bu tez çalışmasının hedeflerinden biri de, bu teknik altyapıyı kullanarak geliştirilecek olan

²² http://www.nvi.gov.tr/Hakkimizda/Projeler,Mernis_Genel.html

²³ Coğrafi Bilgi Sistemleri

bir srveyans sistemi iin ihtiya duyulacak bilimsel desteęin sunulmasıdır. Geliştirilen araç, sunucu-istemci mantığı ile alıřan, kk bir veri tabanı ile oturum tabanlı yerine kullanıcı tabanlı sreler modelleye imkan saęlayan bir araçtır. Kullanıcılar bir tarayıcı yardımı ile sunucuya baęlanarak profillerine eriřebilecekler ve kendilerine zg alıřmalarına daha nce bıraktıkları yerden devam edebileceklerdir. Ara ile, sına ma iin kullanılan verilerde hedeflenen sonuları reten algoritmanın bulunması, algoritmaların karřılařtırılması amalanmıřtır. Algoritmalar veri n iřleme ve aykırılık tespitinin ardıl kullanımından oluřmaktadır. Parametrik alanlarda yaratılan esneklikle algoritmaların iyileřtirilmesi saęlanabilmektedir.

Srveyans sistemlerinden saęlanan veriler, zaman serileridir. Modern srveyans sistemlerinin birincil amacı, salgın risklerinin erkenden tespit edilerek eylem planlarının devreye sokulmasının saęlanmasıdır. Bu yzden, yalnızca hastanelerden elde edilen verilerle yetinilmemektedir. Hastaların internet arama motorlarında arattıkları hastalık ile tanı anahtar szcklerinde meydana gelen artıř, eczanelerde gerekleřen ila satıřlarındaki artıřlar, okul ve iř yerlerindeki devamsızlık miktarındaki artıřların da srveyans sistemlerinde erken uyarı sistemleri iin kullanılması nerilen veri kaynaklarıdır.

Zaman serileri zerinde aykırılık tespiti iin makine ęrenmesi veya *SPC* arařtırmalarında kullanılan gzlem ve tespit yntemleri uygulanmaktadır. Salgın tespiti iin de *SPC* yntemleri yaygın olarak kullanılmaktadır. *Hidden Markov Model* gibi makine ęrenmesinde kullanılan yntemlerin de salgın tespitinde kullanıldıęı grlmřtr [44]. Salgın tespitinde yaygın olarak kullanılan yntemlerden bazıları 2. Kesimde ayrıntılı olarak ele alınmıřtır. Ne var ki, srveyans verilerine doęrudan bu tespit yntemlerini uygulamak uygun deęildir. Bu yzden zaman serisindeki aıklanabilir rntlerin dzleřtirilerek ıkarılması gerekir. Srveyans verilerindeki en yaygın rntler haftanın gn etkisi, mevsimsel dng etkisi, tatil ve tatil dnř etkileri gibi etkilerin yarattıęı rntlerdir. Veri n iřleme ve dzleřtirme yntemleri ayrıntılı olarak 3. Kesimde ele alınmıřtır.

Seilecek olan n iřleme yntemi de aykırı durum tespit yntemi de veriye ve sistemden beklenen hedefe zg olarak seilmelidir. Yani herhangi bir yntemin

bir diğ erinden daha iyi oldu ğ unu söylemek çok do ğ ru olmayabilir. Zira bir zaman serisine dü zleřtirme uygulandı ğ ında, o zaman serisinde olmayan bir örüntü arařtırılıp veriden ıkarılırsa, hata ile aykırı kayıtlar veriden ayıklanabilir. Tespit yöntemi seimi için iki soruya cevap bulmak gerekir. Bunlardan birincisi, sistemin hangi salgın türünü tespit edilmesinde kullanılacağıdır. Eğer bir sistemden verideki ani yükselişleri bulması bekleniyorsa Shewhart, ufak kaymaları tespit etmesi bekleniyorsa *Cusum* tercih edilebilir. Veri kaynağından sağlanan veriye ait bir geçmiş yoksa veya *timeliness* ölçütü için çabukluk bekleniyorsa *EARS* yöntemlerinden C2 tercih edilebilir. Bu koşullar benzeri şekilde çoğaltılabilir. Tespit yöntemi için cevap verilmesi gereken ikinci soru ise, kullanılacak yöntemin hatalı uyarıya sebep olma esnekliğinin ne olacağıdır. Eğer sistemden yanlış uyarıya sebep olmaması(en az uyarı ile çalışması) bekleniyorsa yüksek bir *ATFS* ile *specificity* ölçütünün 1'e yaklaşması sağlanabilir. Bu ancak eşik değerinin düşük tutulması ile olur. Eşik değeri düşürüldüğünde tespit edilen salgın miktarında da bir düşüş gözlemlenecektir. Yine aynı şekilde, do ğ ru uyarı oranını yükselterek *sensitivity* ölçütünden elde edilen ölçüm 1'e yaklařtırılmak isteniyorsa, daha düşük *ATFS* değ erleri ile hesaplanan eşik değ eri düşük tanımlanabilir. Bu durumda da *specificity* ölçütü için yapılan ölçümlerin başarımı düşecektir. Basit bir hesaplama, sistemin ülke genelinde 1.000 hastaneden veri sağladığı varsayılırsa ve bir tanı için yılda 5 yanlış uyarıya açık bir yöntem ve parametreler bütünü tanımlanmışsa, sistem yılda yaklaşık olarak 5.000 hatalı uyarı verir. Sistemin 10 farklı hastalık için bu ölçümleri yaptığı varsayılırsa ve her hastalık için ortalama 5.000 uyarı oluşursa her hastalık için, 50.000 hatalı uyarı demektir ki bu durumun ne kadar kabul edilebilir ya da ölçüklenebilir olduğu da tartışılabilir. Her iki ölçüt birleştirilerek en ideal koşul elde edilmek isteniyorsa, yöntemlerin *ROC* eğrileri çizilir ve *AUC* hesaplarından en iyi başarıma sahip yöntem belirlenebilir. Burada yöntemden kasıt, ön işleme yöntemi tercihi, ön işleme yönteminde seçilen parametreler, tespit yöntemi tercihi ve bu yöntemin seçilen parametrelerinin bütünüdür.

Bu çok seçimli ön işleme ve tespit yöntemlerinin parametrelerinin otomatikleştirilmesi sağlanarak daha iyi bir sürveyans sistemi oluşturulabilir. Bu teknik olarak mümkündür. İdeal bir sürveyans sisteminden beklenen yetenekler ile bu yeteneklerin gerçekleştirilmesi için ihtiyaç duyulan gereksinimler araştırılmaya açık olarak durmaktadır. Modern bir sürveyans sistemi herşeyden önce, veri

kaynaklarını doğru organize etmiş olmalıdır. Verilerin güvenilir, doğru ve zamanında elde edilmesi son derece önemlidir. Verilerin doğru ve güvenilir girilmesi, tabiplerin ICD-10 kodlarını her hasta için tam ve doğru olarak sisteme işlemesi anlamına gelmektedir ki, bir srveyans sisteminin başarımının ilk aaması, hastanın doktor tarafından sisteme belirtilen koulları saėlayacak Őekilde ilenmesidir. Verilerin sisteme doğru olarak girilmesinin ardından, sistem tarafından hastalığın sebep olabileceėi salgın trne gre; zaman, mekan, ya, cinsiyet gibi deėiŐkenlere uygun biŐimde sınıflandırılması gerekir. Yapılan czmlerlerin ise doğru ekip tarafından tanımlanan doğru lçtlere gre iletilmesi gerekir. Salgın, tanımı gereėi, belli bir zaman ve meknda aranılır. Ayrıca salgının anlamlı bir istatistiksel artışa sebep olduėu bilinmelidir. Bir srveyans sisteminin lçtlerinin belirlenmesinde halk saėlıėı uzmanları, istatistikçiler, biyoistatistikçiler bulunmalıdır. Ayrıca sistemin czmlenmesi ve gerçekteştirilmesi için bu bilirkiŐilerin tecrbelerini modelleyerek gerçekte zamanlı sisteme dnŐtrecek olan bir yazılım ekibi olmalıdır. Bunun dıŐında, nfus hareketliliėinin incelenmesi ve hangi koullarda gzlemlerin artmasının doėal karŐılanacaėının belirlenmesi gerekir. Yani anlamlı rntler tanımlanmalıdır.

Sonuç olarak, gnmzde biyolojik terr tehdidi ile baŐlayan, salgınların getirdiėi ekonomik yk ve saėlık sorunlarının yarattıėı huzursuzluk ile devam eden sebeplerden dolayı biyosrveyans sistemlerinin nemi artmıŐtır. Farklı calıŐmalarla farklı yntemler test edilmiŐ ve kullanılan veriye zg sonuçlar yayınlanmıŐtır. Uluslararası dzeyde bir srveyans sistemi gerekliliėi konuŐulurken, gerçekteirilmesindeki organizasyon gclė buna engel olmaktadır. Hem salgının matematiksel bir formlle ifade edilemeyeiŐi, hem veri kalitesi, hem de bu verilerin czmlenmesinin yapılacaėı algoritmanın karar verilmesindeki gclk, bu organizasyonun nndeki en nemli engeller olarak sıralanabilir. lkemizde ise biyosrveyans araŐtırmaları yalnızca tanım aŐamasında kalmıŐ, hastanelerin bildirmek zorunda olduėu aylık salgın raporları dıŐında herhangi bir iŐleyen sistem ortaya konulamamıŐtır. lkemizde bir srveyans sisteminin geliŐtirilmesi için gerekli teknik bilgi yeterli iken, bu konuda yapılan calıŐmalar ve sunulan imknlar yeterli deėildir. Bu tez calıŐması sırasında karŐılaŐılan sorunların da bu cvrede toplandıėı sylenebilir. Gerek kullanılabilir yerli veri kaynaėı bulmak, gerekse bu

konuda alıřma isteęi duyan arařtırmacı bulmak kolay deęildir. Ancak bunlar stesinden gelinmeyecek sorunlar deęildir.

6. Kaynakça

- [1] A. Siu ve Y. C. R. Wong, «Economic Impact of SARS: The Case of Hong Kong,» *Asian Economic Papers*, cilt 3, no. 1, pp. 62-83, 2004.
- [2] T. H. Lotze, *Anomaly Detection in Time Series: Theoretical and Practical Improvements for Disease Outbreak Detection*, Ph.D. Thesis. University of Maryland, 2009.
- [3] W. MM, T. FC, E. JU, D. VM, S. DF, C. RA, M. LF, D. DW, D. MJ ve F. DB., «The emerging science of very early detection of disease outbreaks,» *J Public Health Manag Pract*, cilt 7, no. 6, 2001.
- [4] L. Stern ve D. Lightfoot, «Automated outbreak detection: a quantitative retrospective analysis,» *Epidemiol. Infect.*, pp. 103-110, 1999.
- [5] K. J. Henning, «Overview of Syndromic Surveillance What is Syndromic Surveillance?,» *Morbidity and Mortality Weekly Report*, New York, 2004.
- [6] Hıfzıssıhha Mektebi Müdürlüğü, «Surveyans Sistemi Mevcut Durumu ve Uluslararası Kuruluşların Talep Ettiğ Sağlık Göstergeleri ile Uyumunu,» Refik Saydam Hıfzıssıhha Merkezi Başkanlığı, Ankara, 2007.
- [7] G. Shmueli ve H. S. Burkom, «Statistical Challenges Facing Early Outbreak Detection in Biosurveillance,» *Technometrics (Special Issue on Anomaly Detection)*, no. 1, pp. 39-51, 2010.
- [8] İstanbul Tıp Fakültesi, «Direkt grafi çekimleri Hk,» 09 02 2010. [Çevrimiçi]. Available: [http://www.istanbul.edu.tr/itf/attachments/021_Yaz%C4%B1%20%C4%B0%C5%9Fleri%202010%20Genelgeleri%20\(7\).pdf](http://www.istanbul.edu.tr/itf/attachments/021_Yaz%C4%B1%20%C4%B0%C5%9Fleri%202010%20Genelgeleri%20(7).pdf).
- [9] H. S. Burkom, S. P. Murphy ve G. Shmeuli, «Automated Time Series Forecasting for Biosurveillance,» *Statist. Med.*, 26, no. 22, p. 4202–4218., 2007.
- [10] A. Goldenberg, G. Shmueli, R. A. Caru ve S. E. Fienberg, «Early statistical detection of anthrax outbreaks by tracking over-the-counter medication sales.,» *Proceeding of the National Academy of Sciences*, cilt 99, no. 8, pp. 5237-5240, 2002.
- [11] S. E. Fienberg ve G. Shmueli, «Statistical issues and challenges associated

- with rapid detection of bio-terrorist attack,» *Statistics in Medicine*, cilt 24, no. 4, pp. 513-529, 2005.
- [12] T. Lotze, G. Shmueli ve I. Yahav, «Simulating Multivariate Syndromic Time Series and Outbreak Signatures,» *Robert H. Smith School Research Paper No. RHS-06-054*, 2007.
- [13] B. DL, B. H, C. M, H. WR ve M. AW., «Algorithms for rapid outbreak detection: a research synthesis.,» *Journal of Biomedical Informatics*, cilt 38, no. 2, pp. 99-113, 2005.
- [14] M. A. Stoto, R. D. Fricker, A. J. Jain, A. Diamonds, J. O. Davies-Cole, C. Glymph, K. Gebreyesus, G. Lum, L. Jones, K. Dehan ve C. Yuan, «Evaluating Statistical Methods for Syndromic Surveillance,» %1 içinde *Statistical Methods in Counterterrorism; Game Theory, Modeling, Syndromic Surveillance, and Biometric Authentication*, Springer, 2006, pp. 141-172.
- [15] T. Burr, T. Graves, R. Klamann, S. Michalak, R. Picard ve N. Hengartner, «Accounting for seasonal patterns in syndromic surveillance data for outbreak detection,» *BMC Medical Informatics and Decision Making*, cilt 6, no. 40, 2006.
- [16] M. KD, R. B ve C. C, «Measuring outbreak-detection performance by using controlled feature set simulations.,» *Morbidity and Mortality Weekly Report*, cilt 53, pp. 130-136, 2004.
- [17] S. N. Mukhi, *Integrated Approach to Real-Time Biosurveillance in a Federated Data Source Environment*, Manitoba, 2007.
- [18] J. S. Lombardo, H. Burkom ve . J. Pavlin, «ESSENCE II and the Framework for Evaluating Syndromic Surveillance Systems,» *Morbidity and Mortality Weekly Report (MMWR)*, cilt 53, p. 159–165, 2004.
- [19] M. W. Wagner, J. M. Robinson, F.-C. Tsui ve J. U. Espino, «Design of a National Retail Data Monitor for Public Health,» *Journal of the American Medical Informatics Association*, cilt 10, no. 5, p. 409–418, 2003.
- [20] CDC, «Early Aberration Reporting System,» 9 2012. [Çevrimiçi]. Available: <http://www.bt.cdc.gov/surveillance/ears/>. [10 2012 tarihinde erişilmiştir].
- [21] W. A. Shewhart, «Quality Control Charts,» *Bell System Technical Journal*, cilt 5, no. 4, pp. 593-603, 1926.

- [22] J. S. Oakland, *Statistical Process Control*, 6 dü., A Butterworth-Heinemann Title, 2007.
- [23] D. M. Hawkins ve D. H. Olwell, *Cumulative Sum Charts and Charting for Quality Improvement*, New York: Springer, 1998.
- [24] D. F. Stroup, G. D. Williamson ve J. L. Her, «Detection of aberrations in the occurrence of notifiable diseases surveillance data,» *Statistics in Medicine*, cilt 8, no. 3, pp. 323-329, 1989.
- [25] W. H. Woodall, «The use of control charts in health-care and public-health surveillance,» *Journal of Quality Technology*, cilt 38, no. 2, pp. 89-104, 2006.
- [26] E. S. Page, «Continuous Inspection Schemes,» *Biometrika Trust*, cilt 41, no. 1/2, pp. 100-115, 1954.
- [27] F. F. Gan, «An optimal design of CUSUM control charts for binomial counts,» *Journal of Applied Statistics*, cilt 20, no. 4, pp. 445-460, 1993.
- [28] M. Kartal, *İstatistiksel Kalite Kontrolü*, Erzurum: Şafak Yayınevi, 1999.
- [29] M. M. Wagner, A. W. Moore ve R. M. Aryel, *Handbook of biosurveillance*, Amsterdam ; Boston: Academic Press Inc, 2006.
- [30] D. A. Dunfee ve B. L. Hegler, *Biological Terrorism Preparedness: Evaluating the Performance of the Early Aberration Reporting System (EARS) Syndromic Surveillance Algorithms[Tez]*, 2007.
- [31] S. W. Roberts, «Control Chart Tests Based on Geometric Moving Averages,» *Technometrics*, cilt I, no. 3, pp. 239-250, 1959.
- [32] J. F. Muth, «Optimal Properties of Exponentially Weighted Forecasts,» *Journal of the American Statistical Association*, cilt 55, no. 390, pp. 299-306, 1960.
- [33] G. E. P. Box, G. M. Jenkins ve J. F. MacGregor, «Some Recent Advances in Forecasting and Control,» *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, cilt 23, no. 2, pp. 158-179, 1974.
- [34] J. S. Hunter, «The Exponentially Weighted Moving Average,» *Journal of Quality Technology*, cilt 18, no. 4, pp. 203-210, 1986.
- [35] S. V. Crowder, «A Simple Method for Studying Run Length Distributions of Exponentially Weighted Moving Average Control Charts,» *Technometrics*, cilt

29, no. 4, pp. 401-407, 1987.

- [36] D. C. Montgomery, J. S. Gardiner ve B. A. Pizzano, «Statistical Process Control Methods for Detecting Small Process Shifts,» *Frontiers in Statistical Quality Control*, cilt 3, pp. 161-178, 1987.
- [37] J. M. Lucas ve M. S. Sanucci, «Exponentially Weighted Moving Average Control Schemes: Properties and Enhancement,» *Technometrics*, cilt 32, no. 1, 1990.
- [38] R. D. Fricker, *Introduction to Statistical Methods for Biosurveillance*, Cambridge University Press[Basımda], 2013.
- [39] D. C. Montgomery, *Introduction to Statistical Quality Control*, 5 dü., John Wiley & Sons., 2004.
- [40] R. E. Watkins, S. Eagleson, B. Veenendaal, G. Wright ve A. J. Plan, «Applying cusum-based methods for the detection of outbreaks of Ross River virus disease in Western Australia,» *BMC Medical Informatics and Decision Making*, cilt 8, no. 37, 2008.
- [41] L. Hutwagner, W. Thompson ve G. M. Seema, «The bioterrorism preparedness and response Early Aberration Reporting System (EARS),» *Journal of Urban Health: Bulletin of the New York Academy of Medicine*, cilt 80, no. 2, pp. 89i-96i, 2003a.
- [42] R. D. Fricker Jr, B. L. Hegler ve D. A. Dunfee, «Comparing syndromic surveillance detection methods: EARS' versus a CUSUM-based methodology,» *Statistics in Medicine*, no. 17, p. 3407–3429, 2008.
- [43] I. Szarka JL, G. L ve W. WH., «Comparison of the early aberration reporting system (EARS) W2 methods to an adaptive threshold method.,» *Statistics in Medicine*, cilt 30, no. 5, pp. 489-504, 2011.
- [44] R. E. Watkins, S. Eagleson ve B. Veen, «Disease surveillance using a hidden Markov model,» *BMC Medical Informatics and Decision Making*, cilt 9, no. 39, 2009.
- [45] A. M. Pelecanos, P. A. Ryan ve M. L. Gatt, «Outbreak detection algorithms for seasonal disease data: a case study using ross river virus disease,» *BMC Medical Informatics and Decision Making*, cilt 10, no. 74, 2010.
- [46] G. Shmueli ve S. E. Fienberg, «Current and Potential Statistical Methods for

Monitoring Multiple Data Streams for Biosurveillance,» *Statistical Methods in Counterterrorism*, Springer New York, 2006, pp. 109-140.

- [47] «Wikipedia,» [Çevrimiçi]. Available: http://en.wikipedia.org/wiki/Regression_analysis.
- [48] K. Kleinman, R. Lazarus ve R. Platt, «A Generalized Linear Mixed Models Approach for Detecting Incident Clusters of Disease in Small Areas, with an Application to Biological Terrorism,» *Am. J. Epidemiol*, cilt 159, no. 3, pp. 217-224, 2004.
- [49] D. Muscatello, «An adjusted cumulative sum for count data with day-of-week effects: application to influenza-like illness,» *Syn. Surv. Conf.*, Boston, 2004.
- [50] Y. Zhuangt, L. Chen, X. Wang ve J. Lian, «A Weighted Moving Average-based Approach for Cleaning Sensor Data,» *Distributed Computing Systems, 2007. ICDCS '07. 27th International Conference on*, Toronto, ON, 2007.
- [51] C. S. Holt, «Forecasting trends and seasonal by exponentially weighted averages,» *ONR Memorandum*, cilt 52, 1957.
- [52] C. C. Holt, «Forecasting seasonals and trends by exponentially weighted moving averages,» *International Journal of Forecasting*, cilt 20, no. 1, pp. 5-10, 2004.
- [53] P. R. Winters, «Forecasting Sales by Exponentially Weighted Moving Averages,» *Management Science*, cilt 6, no. 3, pp. 324-342, 1960.
- [54] S. G. Makridakis, S. C. Wheelwright ve R. J. Hyndman, *Forecasting: Methods and Applications*, New York: John Wiley and Sons, 1998.
- [55] K. Singh, A. Kumar ve H. Chandra, «Forecasting Techniques in Agriculture,» 2003. [Çevrimiçi]. Available: <http://www.iasri.res.in/ebook/FET/>. [Ekim 2012 tarihinde erişilmiştir].
- [56] R. J. Hyndman ve G. Athanasopoulos, «Forecasting: principles and practice,» Mart 2012. [Çevrimiçi]. Available: <http://otexts.com/fpp/>. [Ekim 2012 tarihinde erişilmiştir].
- [57] R. G. Brown, *Smoothing, forecasting and prediction of discrete time series*, Mineola, NY: Dover Publications, 2004.

- [58] B. Mandel, «The Regression Control Chart,» *Journal of Quality Technology*, cilt 1, no. 1, pp. 1-9, 1969.
- [59] F. Güneysu ve N. Yamak, «İMKB'de Haftanın Günü Etkisinin Kriz Dönemleri için Araştırılması,» *Finans Politik & Ekonomik Yorumlar*, cilt 48, no. 560, pp. 33-44, 2011.
- [60] A. K. Tunçel, «Haftanın Günü Etkisi'ne yeni bir yaklaşım:İMKB örneği,» *Trakya Üniversitesi Sosyal Bilimler Dergisi*, cilt 10, no. 2, pp. 246-271, 2008.
- [61] N. Ergül, S. Dumanoğlu ve V. Akel, «İMKB'de Günlük Anomaliler,» *Marmara Üniversitesi İİBF Dergisi*, cilt 25, no. 2, 2008.
- [62] R. M. C. Apolinario, O. M. Santana ve L. J. Sales, «Day of the Week Effect on European Stock Markets,» *International Research Journal of Finance and Economics*, no. 2, pp. 53-70, 2006.
- [63] J. C. Brillman, T. Burr, D. Forslund, E. Joyce, R. Picard ve E. Umland, «Modeling emergency department visit patterns for infectious disease complaints: results and application to disease surveillance,» cilt 5, no. 4, 2005.
- [64] T. Lotze, G. Shmueli ve I. Yahav, «Simulating and Evaluating Biosurveillance Datasets,» *Biosurveillance: A Health Protection Priority*, Chapman and Hall, BASIMDA.
- [65] J. I. Tokars, H. Burkom, J. Xing, R. English, S. Bloom, K. Cox ve J. A. Pavlin, «Enhancing time series detection algorithms for automated biosurveillance,» *Emerg Infect Dis.*, cilt 15, no. 4, pp. 533-539, 2009.
- [66] M. R. Busse, C. R. Knittel ve F. Zettelmeyer, «Pain at the Pump: The Differential Effect of Gasoline Prices on New and Used Automobile Markets,» *NBER WORKING PAPER SERIES*, 2009.
- [67] A. Aslan, «Türkiye'de Ekonomik Büyüme ve Turizm İlişkisi Üzerine Ekonometrik Analiz,» *Erciyes Üniversitesi SBE Dergisi*, 2008.
- [68] T. C. M. Bankası, «Enflasyon Raporu,» TCMM, 2011-IV.
- [69] S. Çalık, «Ekonomik Zaman Serilerinde Mevsimsellik Analizi,» TÜİK Uzmanlık Tezi, 2009.
- [70] R. E. Serfling, «Methods for current statistical analysis of excess pneumonia-

influenza,» *Public Health Rep*, cilt 78, no. 6, pp. 494-506, 1963.

- [71] T. Lotze, S. Murphy ve G. Shmueli, «Implementation and Comparison of Preprocessing Methods for Biosurveillance Data,» *Advances in Disease Surveillance*, cilt 6, no. 1, 2008.
- [72] C. Chatfield ve M. Yar, «Holt-Winters Forecasting: Some Practical Issues,» *Journal of the Royal Statistical Society. Series D (The Statistician)*, cilt 2, no. 37, pp. 129-140, 1988.
- [73] S. P. Murphy ve H. Burkom, «Recombinant Temporal Aberration Detection Algorithms for Enhanced Biosurveillance,» *Journal of the American Medical Informatics Association*, cilt 15, no. 1, pp. 77-87, 2008.
- [74] U.S. Department of Health & Human Services, "Pandemic Flu History," [Online]. Available: <http://www.flu.gov/pandemic/history/index.html#>. [Accessed 30 12 2012].

ÖZGEÇMİŞ

Adı Soyadı : Yasin ŞAHİN

Doğum Yeri : Amasya

Doğum Yılı : 1986

Medeni Hali : Bekâr

Eğitim ve Akademik Durumu:

Lise : 2001-2004 Meram Anadolu Anadolu Lisesi

Lisans : 2004-2009 Hacettepe Üniversitesi Mühendislik Fakültesi
Bilgisayar Mühendisliği Bölümü

Yabancı Dil : İngilizce

İş Tecrübesi:

2009-... Araştırma Görevlisi, Hacettepe Üniversitesi Bilgisayar
Mühendisliği Bölümü

2012-... Hacettepe Üniversitesi Hastane Otomasyonu Analiz ve
Optimizasyonu