

Corona Virüs Salgını Teşhis Sayısı Öngörü Analizi v3

Arif Selçuk Öğrenci

v1: 29.3.2020 (TR için 18 gün)

v2: 3.4.2020, 14:00 (TR için 24 gün)

v2: 11.4.2020, 14:00 (TR için 32 gün)

TÜMÜNÜ OKUMAK İSTEMEYENLER İÇİN SONUÇ

GENEL YORUM

Her bir gün için (ilk n gün kullanılarak) yapılan sig-3 ve log-3 modellerinde log-3 hep sig-3'e göre daha yüksek bir değeri öngörmektedir. Çin ve Kore verisinde zıplama-düşme-zıplama-düşme ve modellerin yaklaşması şeklinde bir desen gözlenmektedir. Bu desen, İtalya ve İspanya'da da gözleendiğinden, İtalya ve İspanya için rapor v2'de doyuma ulaşma ve nihai maksimum tahminleri yapılmıştır. Bu tahminler USA ve TR için sadece bir zıplama gözleendiğinden, USA ve TR için ikinci zıplama tarihi öngörölmüştür. Gözlenen bu tarihler (başlangıçtan itibaren gün sayısı) aşağıda tabloda verilmiştir. ZIPLAMA TARİHİ modelden elde edilecek maksimum değerin zıplayacağı tarihtir, gerçek verinin zıplayacağı bir tarih değildir!

Rapor v2 tablosu

ÜLKE (GÜN, 3.4.2020 itibariyle)	Zıplama-1 (GÜN)	Zıplama-2 (GÜN)	Modellerin birleşmesi (GÜN)	Doyum (max. değeri) (ADET)
ÇİN (93)	25	50	75	82000
KORE (73)	25	41	55	9300
İTALYA (64)	40	53	73 (TAHMİN, 12 NİSAN)	155000-160000 (TAHMİN)
İSPANYA (63)	42	56	76 (TAHMİN, 16 NİSAN)	190000-200000 (TAHMİN)
USA (74)	60	75-80 (TAHMİN)	95 (TAHMİN)	
TR (24)	19	34-39 (TAHMİN)	54 (TAHMİN)	

10 Nisan'da USA zıplama yapmış olursa,
12 Nisan'da İTALYA doyuma ulaşırsa,
16 Nisan'da İspanya doyuma ulaşırsa,
18 Nisan'da TR zıplama yapmış olursa
Modellerin doğruluğu için bir gösterge elde edilmiş olacaktır.

BU RAPORDA GÖZLENEN DURUM:

10 Nisan'da USA zıplama YAPMAMIŞTIR.
12 Nisan'da İTALYA doyuma ULAŞMAMIŞTIR:
16 Nisan'da İspanya'nın doyuma ulaşması olası görölmemektedir.
TR için tahmin bir sonraki raporda doğru/yanlışlanabilecektir.

ÇİN ve KORE için gözlenen dinamikler diğerlerinde gözlenmemektedir, en azından aynı sürelerde gözlenmemektedir. Sig-3 ve log-3 max değerlerin birbirine yakınsaması için gereken süreler uzamıştır. Bunu,

- İspanya, İtalya, USA ve TR'nin uyguladıkları yöntemler ÇİN ve KORE kadar başarılı değildir, diye yorumlamak olasıdır
- İspanya, İtalya, USA ve TR'nin vaka sayılarını belirlemede kullandığı test yaklaşımı ÇİN ve KORE'ye göre daha yaygın bir insan grubunu kapsamaktadır diye yorumlamak olasıdır.

TR için sağlıklı bir doyum noktası tahmini için hala erkendir.

0. Aşağıdaki tüm analizler temelde dinamik sistemin verisinin matematik modellemesine dayanmaktadır. Hastalık yayılımı ile ilgili bir tıbbi modelleme veya bir tıbbi/biyolojik uzmanlık barındırmamaktadır. Amaç, Türkiye'nin ilk 18-24-32 güne ilişkin verisine dayanarak toplam vaka sayısını öngörmenin ne kadar olanaklı olduğunu bulmaktır. Bunun için doyuma ulaşmış ve hala artışta olan bazı ülkelerin verileri incelenecektir. Bu kapsamda bu ülkeler için verinin tamamı yerine ilk-"n" günü kullanılarak farklı sig-3 / log-3 regresyon modelleri elde edilmiş, burada ulaşılması beklenen maksimum hasta sayısının bir desene sahip olup olmadığı, günlük değişimlerin nasıl bir oynaklığa yol açtığı incelenmiştir.

Bu analizin 5-8 günde bir güncellenmesi planlanmaktadır.

1. Kullanılan veri kaynağı: <https://ourworldindata.org/coronavirus-source-data>
2. Modelde kullanılan iki fonksiyon
3-parametrelı sigmoid (sig-3)

$$y = \frac{a}{1 + e^{-\frac{x-x_0}{b}}}$$

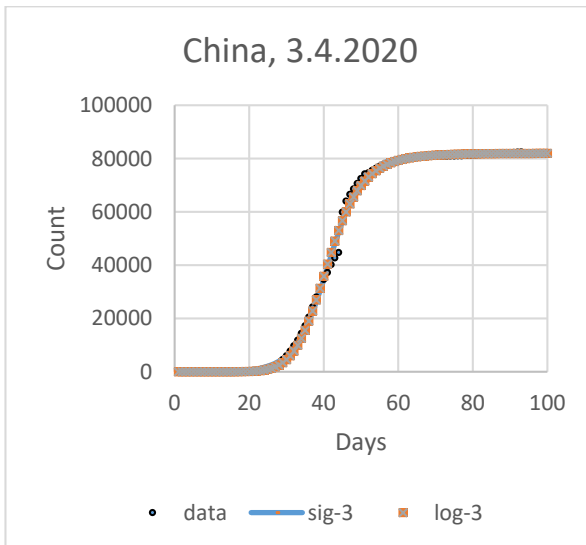
3-parametrelı lojistik (log-3)

$$y = \frac{a}{1 + \left(\frac{x}{x_0}\right)^b}$$

Bu fonksiyonlar salgın modellemesinde SIR, SEIR modellerinde karşımıza sıklıkla çıkmaktadır. Pratik olarak, sig-3 modelinde "a" parametresi en sonunda oluşacak toplam enfekte/hasta sayısını vermektedir. "x0" parametresi oluşan eğrinin büküm noktasını (eğiminin maksimum olduğu nokta), "b" parametresi de eğrinin dikliğini göstermektedir.

R² değeri en basit anlatımla yapılan regresyonun veri noktalarından "ne kadar iyi" geçtiğinin göstergesidir, bu değer 0-1 aralığında olur ve 1 olması, eğrinin veri noktalarının tam üstünden geçtiğini, mükemmel bir "fit" olduğunu gösterir. Sonuçlarda görülen 0.99xxx gibi bir değer oldukça başarılı bir fit yapıldığını gösterir.

3. Aşağıda farklı ülkelerin enfekte hasta sayısı verisine bu iki fonksiyonla SIGMAPLOT yazılımında regresyon uygulanarak model parametreleri elde edilmiştir. Yatay eksen ilk hastanın ilan edildiği andan itibaren gün sayısını (Days), dikey eksen de hasta sayısını (Count) vermektedir.
4. Bu haftaki raporda, ilgili ülkeler için sig-3 ve log-3 modellerinin verdiği maksimum sayının kullanılan gün sayısına göre değişim grafikleri de verilmiştir. Burada bir desen varlığı gözlenmektedir.
5. ÇİN
93 günlük veri



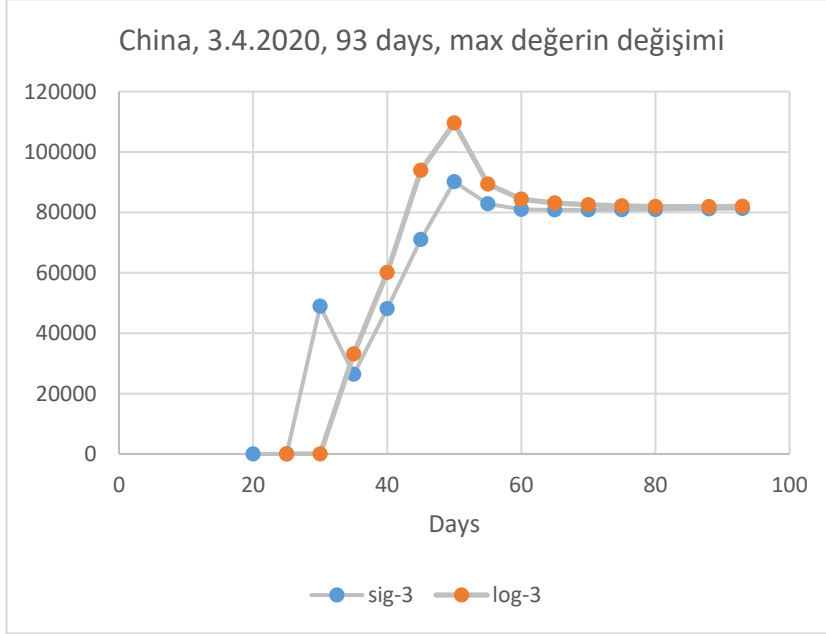
ÇİN (88gün)	sig-3	log-3
a	81138.1	81923.86
b	4.462855	-9.00711
x0	41.31811	41.12196
R ²	0.998362	0.997857

ÇİN (93gün)	sig-3	log-3
a	81300.63	81995.88
b	4.482799	-8.98657
x0	41.34455	41.13311
R ²	0.998409	0.997975

Çin için 3.4.2020 ile 11.4.2020 arasında vaka sayısı 632 artmıştır (<%1) modelde bir değişim yoktur.

Yorum: Her iki fonksiyon da oldukça yakın bir şekilde veriyi modellemektedir. 80 gün civarında doyuma ulaşmış görünen eğri için regresyon modelleri 41 günde bir büküm noktası öngörmektedir. Ulaşılan maksimum hasta değeri 82465 olup, regresyon modelleri bu değere %1.5 yakınlıktadır.

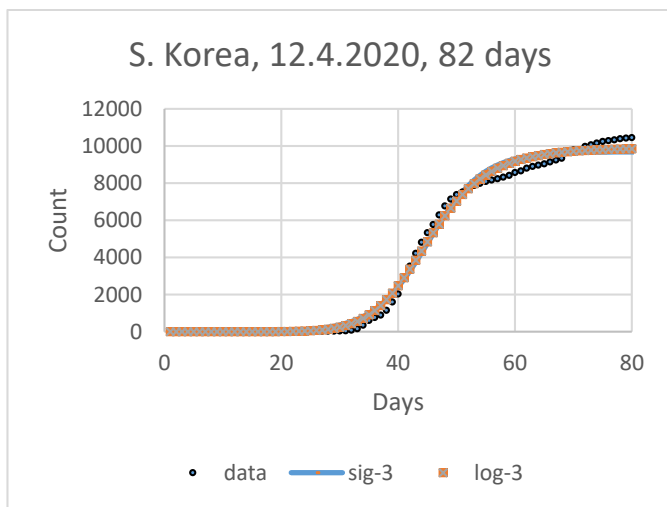
ÇİN verisinin tamamı yerine ilk-n günü kullanılarak sig-3 ve log-3 regresyon modelleri elde edilmiş ve bunların doyum değerleri ("a" parametresi) aşağıda grafikte verilmiştir.



(*) 0 görülen deęerler için yapılan regresyonda algoritma yakınsamayı bařaramamıřtır. Aslında bu noktalarda grafik "uuk" bir tepe yapmaktadır.

Yorum: Sistemin dinamięi hesaplanan a (max sayı) deęerlerinde önce bir zıplama (25 Gn) sonra dřme, tekrar zıplama (50 Gn) ve sonra da her iki modelin birbirine yaklařması (75 gn) řeklinde gzlenmektedir.

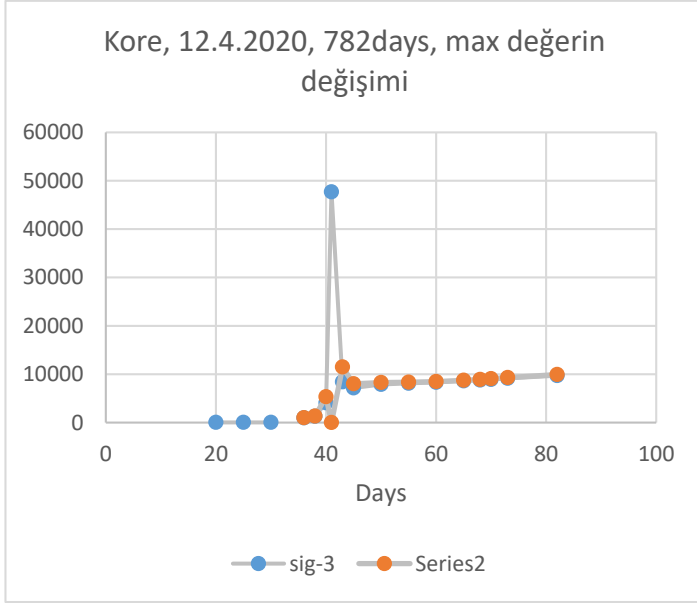
6. G. KORE 82 gnlk veri



G. KORE(82)	sig-3	log-3
a	9689.138	9911.097
b	4.675034	-8.94792
x0	45.17243	45.23732
R^2	0.990434	0.993337

Yorum: Veri doyuma ulaşmamıştır, 50 günden sonra doğrusal bir değişim gözlenmiştir. Ancak sigmoidal bölgede her iki fonksiyon da oldukça yakın bir şekilde veriyi modellemektedir. 80 gün civarında doyuma ulaşacak regresyon modelleri 45 günde bir büküm noktası öngörmektedir.

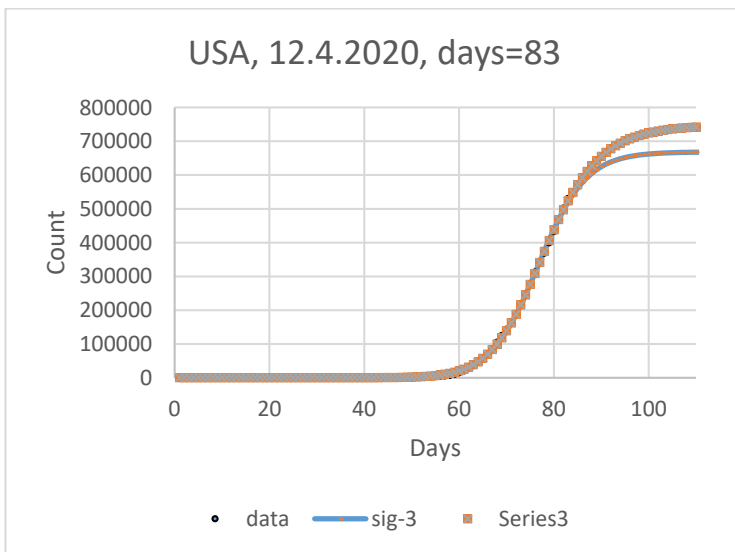
G. Kore verisinin tamamı yerine ilk-n günü kullanılarak sig-3 ve log-3 regresyon modelleri elde edilmiş ve bunların doyum değerleri ("a" parametresi) aşağıda grafikte verilmiştir.



(*) 0 görülen değerler için yapılan regresyonda algoritma yakınsamayı başaramamıştır. Aslında bu noktalarda grafik "uçuk" bir tepe yapmaktadır.

Yorum: Sistemin dinamiği hesaplanan a (max sayı) değerlerinde önce bir zıplama (20-30 Gün, ortalama 25 gün) sonra düşme, tekrar zıplama (41. Gün) ve sonra da her iki modelin birbirine yaklaşması (55 gün) şeklinde gözlenmektedir.

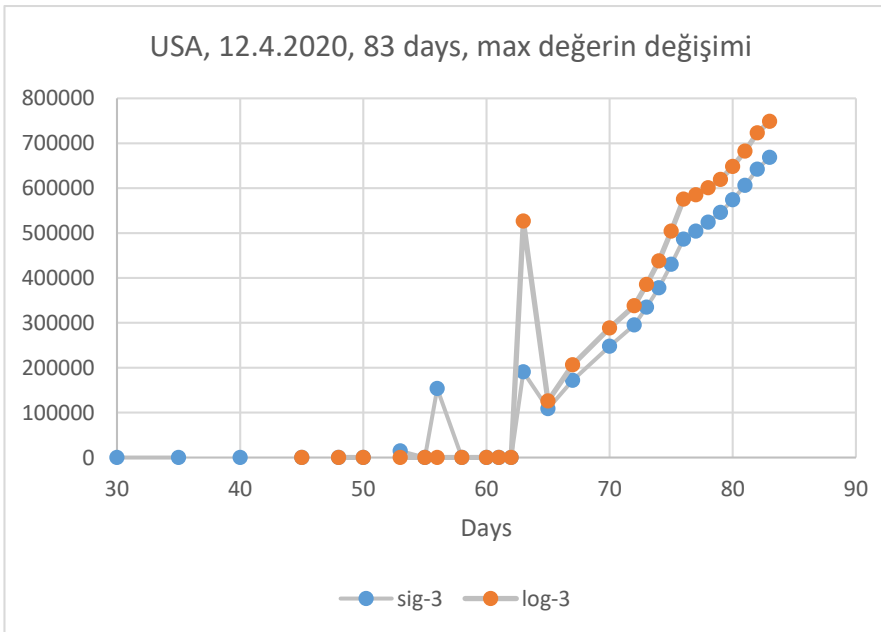
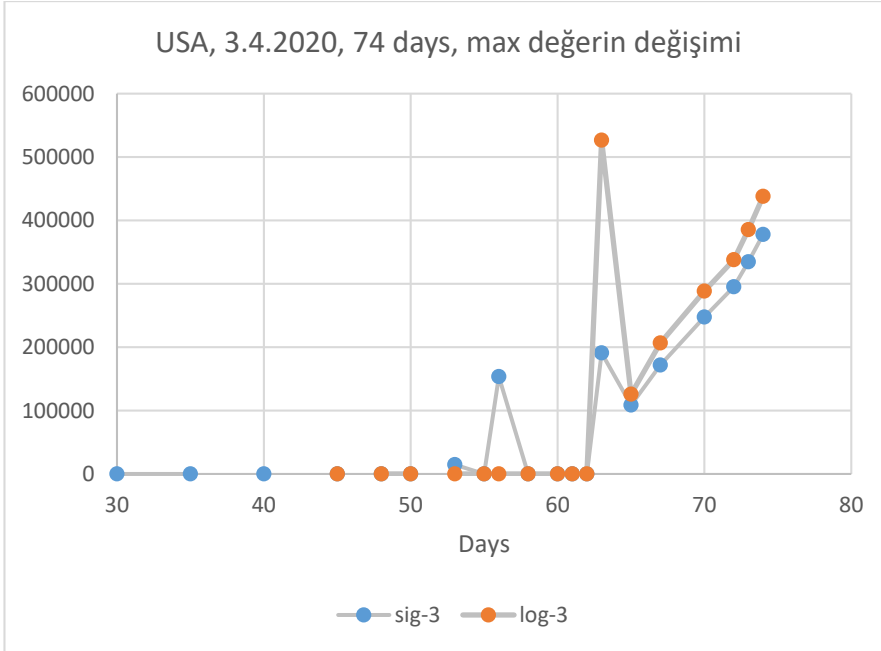
7. USA 83 günlük veri



USA(83gün)	sig-3	log-3
a	668447	748871.8
b	4.980113	-13.6283
x0	76.7518	78.02046
R^2	0.999209	0.999587

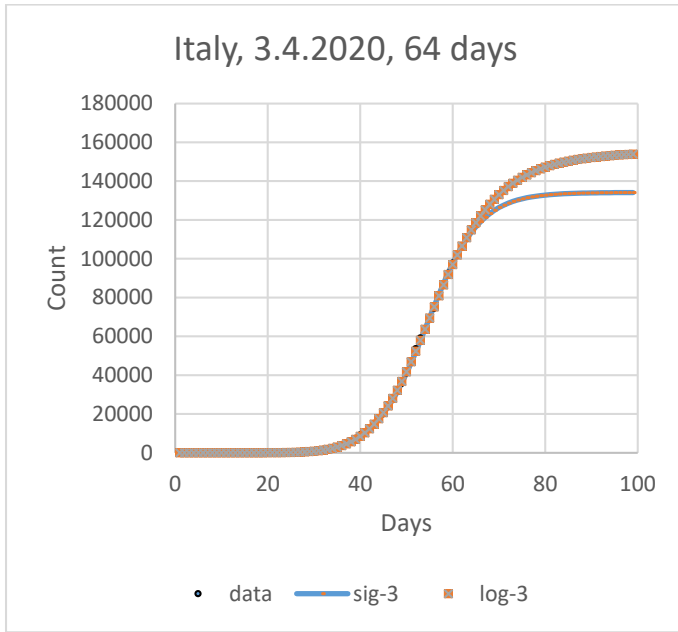
Yorum: Veri doyuma ulaşmamıştır, HALA sigmoidal bölgede olup her iki fonksiyon da oldukça yakın bir şekilde veriyi modellemektedir. 110 gün civarında doyuma ulaşacak regresyon modelleri 76-78 günde bir büküm noktası öngörmektedir. Log-3 modeli daha yüksek bir maksimum öngörmektedir. Önceki rapora göre doyuma ulaşma ve büküm süreleri artmıştır.

USA verisinin tamamı yerine ilk-n günü kullanılarak sig-3 ve log-3 regresyon modelleri elde edilmiş ve bunların doyum değerleri ("a" parametresi) aşağıda grafikte verilmiştir.



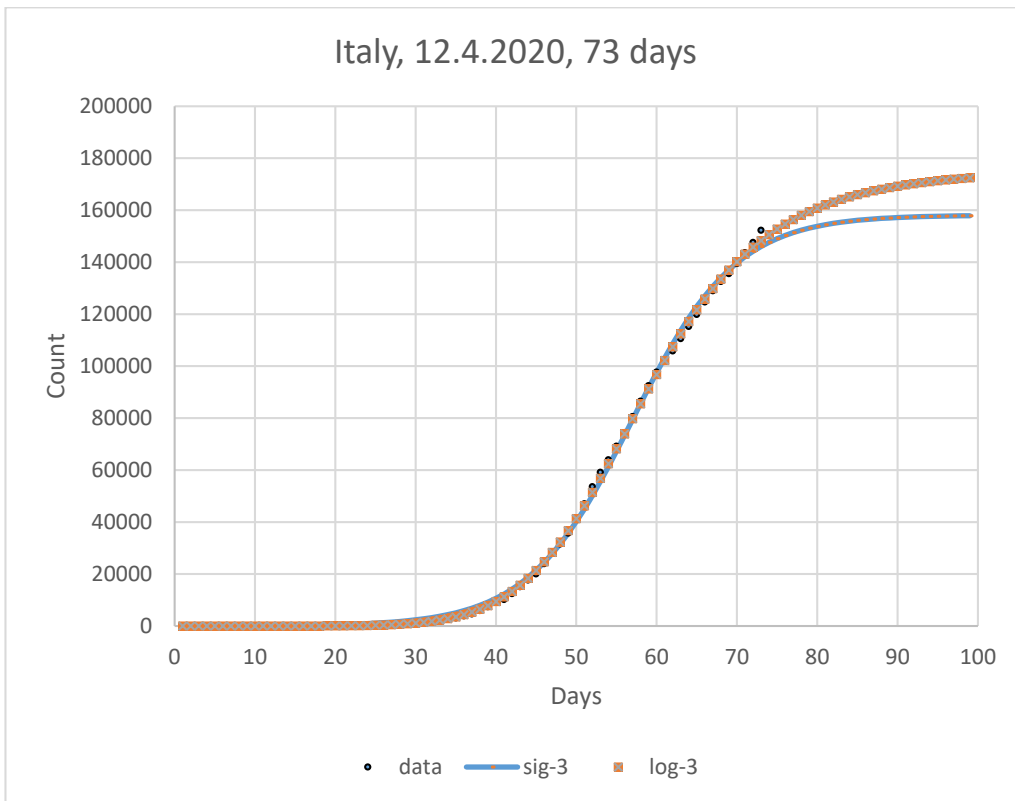
Yorum: İlk 30 günde hasta sayısı 15'dir. Sistemin gerçek dinamiği ortaya çıkmamıştır ancak ilk 30+45 =75 günden sonra tam olarak anlaşılması beklenebilir. Birkaç günlük eklemeler bile tahminlerde ciddi farklılara yol açmaktadır. ÇİN ve KORE örneklerine bakılırsa, 60+20=80 günde bir tepe (uçuk değer) yapıp sistemin 100 günde doyuma ulaşması beklenebilirdi, ancak 76 günde bir davranış değişikliği olmuştur.

8. İTALYA
73 GÜNLÜK VERİ



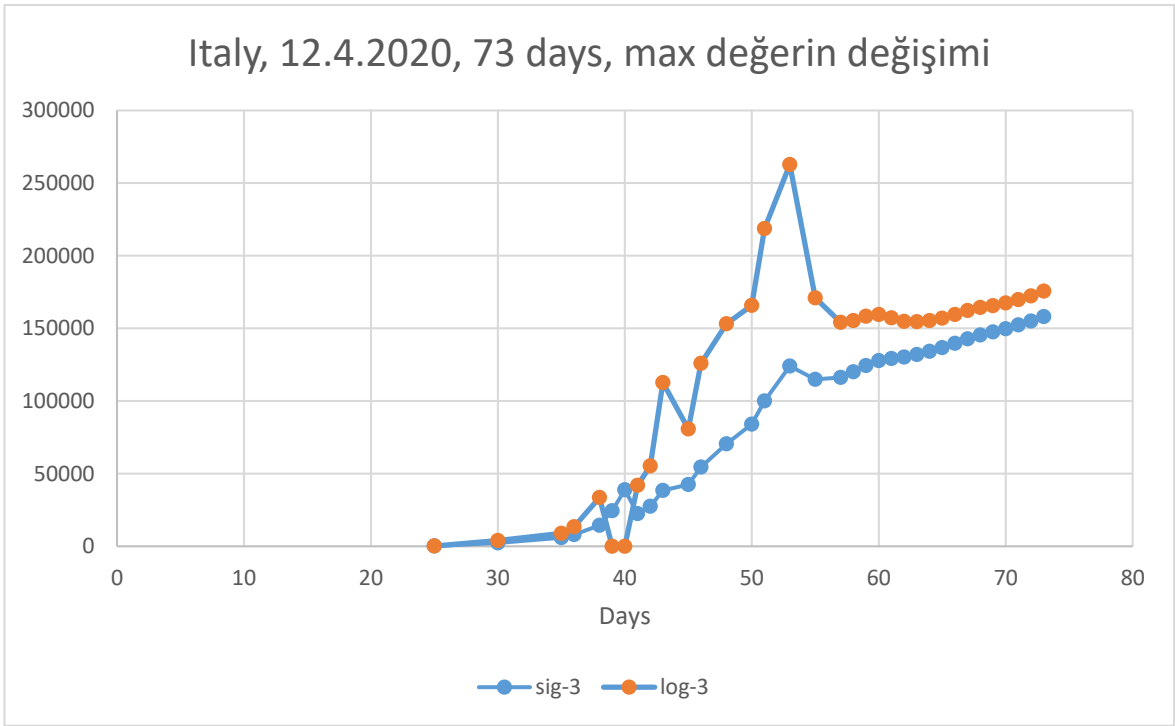
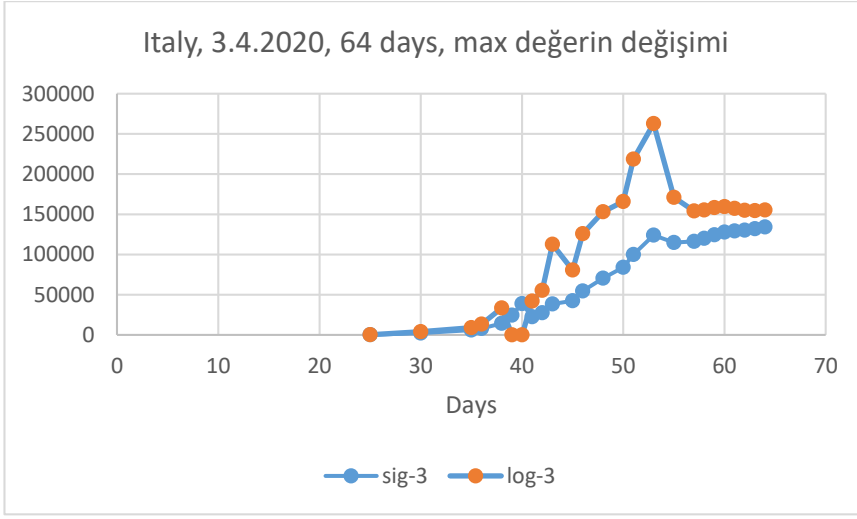
Italy (64 gün)	sig-3	log-3
a	134152.7	155335.4
b	5.535365	-8.28829
x0	54.55298	56.42819

Italy (73 gün)	sig-3	log-3
a	158108.1	175600.9
b	6.442535	-7.58724
x0	56.9438	58.40055
R ²	0.998833	0.999641



Yorum: Veri doyuma ulaşmamıştır, HALA sigmoidal bölgede olup her iki fonksiyon da oldukça yakın bir şekilde veriyi modellemektedir. 90-100 gün civarında doyuma ulaşacak regresyon modelleri 57-58 günde bir büküm noktası öngörmektedir. Log-3 modeli daha yüksek bir maksimum öngörmektedir.

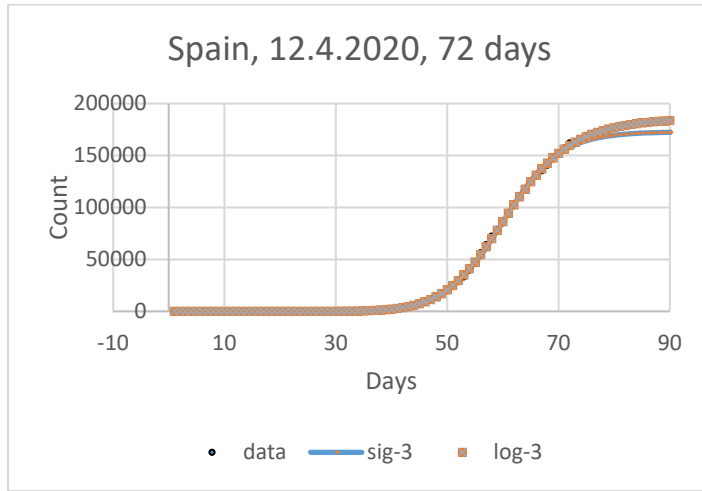
İtalya verisinin tamamı yerine ilk-n günü kullanılarak sig-3 ve log-3 regresyon modelleri elde edilmiş ve bunların doyum değerleri ("a" parametresi) aşağıda grafikte verilmiştir.



YORUM İtalya verisi 40 günde log-3 için anormal zıplama yapıp (sig-3'de de bir zıplama vardır), maksimum deęerde ikinci zıplamayı 53 günde yapmış ve sonra iki model birbirine yaklařmaya bařlamıřtır. Bu iki modelin $53+20=73$ günde birbirine ulařması, 155000-160000 arasında doyuma ulařması beklenmektedir. Ancak bu olmamıřtır, iki model neredeyse paralel hareket etmektedir.

9. İSPANYA

72 GÜNLÜK VERİ

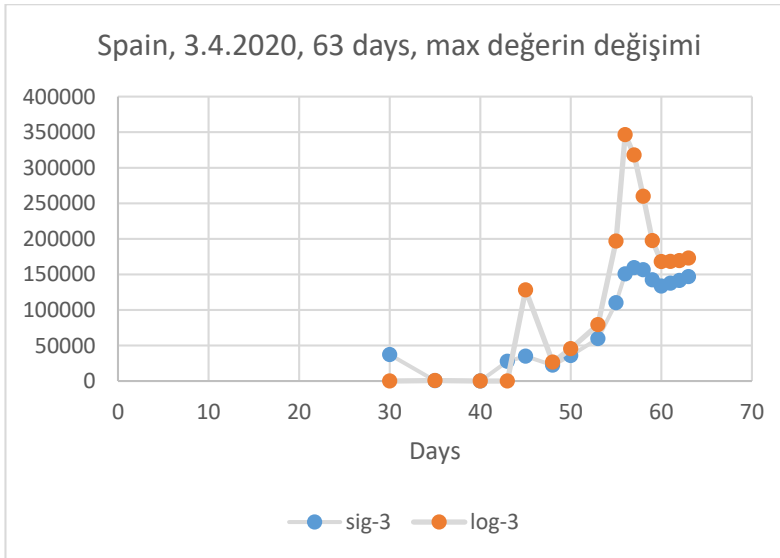


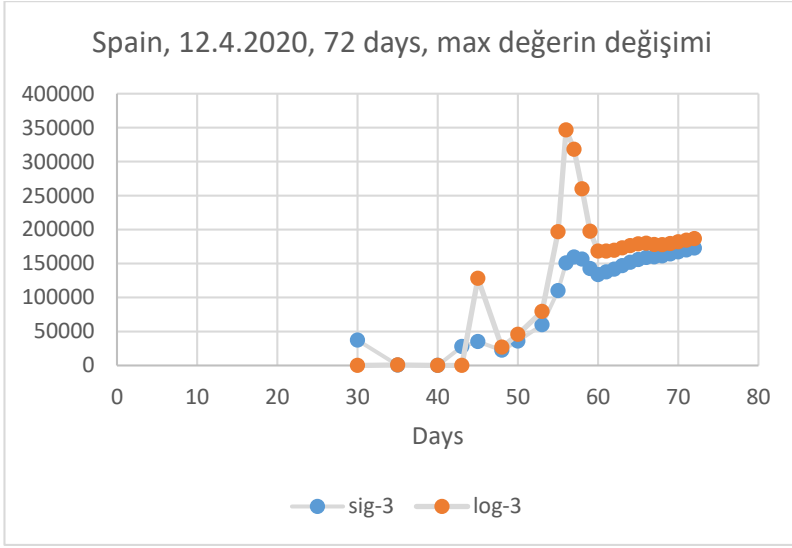
SPAIN(63)	sig-3	log-3
a	146797.3	172988.1
b	4.466438	-11.1365
x0	58.28243	59.96144
R ²	0.999579	0.999652

SPAIN(72)	sig-3	log-3
a	172757.3	186449.1
b	5.052229	-10.6127
x0	60.04394	60.85307
R ²	0.999448	0.999788

Yorum: Veri doyuma ulaşmamıştır, HALA sigmoidal bölgede olup her iki fonksiyon da oldukça yakın bir şekilde veriyi modellemektedir. 90 gün civarında doyuma ulaşacak regresyon modelleri 60 günde bir büküm noktası öngörmektedir. Log-3 modeli daha yüksek bir maksimum öngörmektedir.

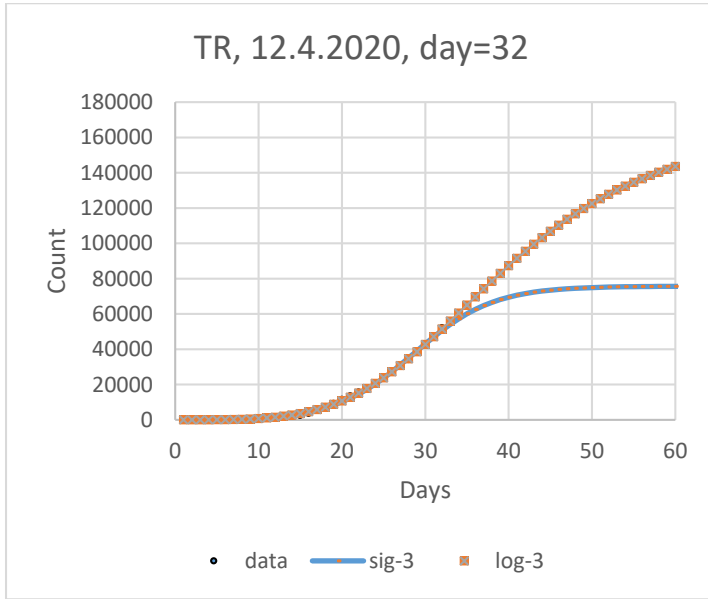
İspanya verisinin tamamı yerine ilk-n günü kullanılarak sig-3 ve log-3 regresyon modelleri elde edilmiş ve bunların doyum değerleri ("a" parametresi) aşağıda grafikte verilmiştir.





YORUM İspanya verisi 40 günde sig-3 ve log-3 için anormal zıplama yapıp, maksimum değerde diğerk zıplamaları 45 ve 56 günde yapmış ve sonra iki model birbirine yaklaşılmaya başlamıştır. Bu iki modelin 56+20=76 günde birbirine ulaşması, 190000-200000 arasında doyuma ulaşması beklenmektedir.

10. TÜRKİYE 32 GÜNLÜK VERİ

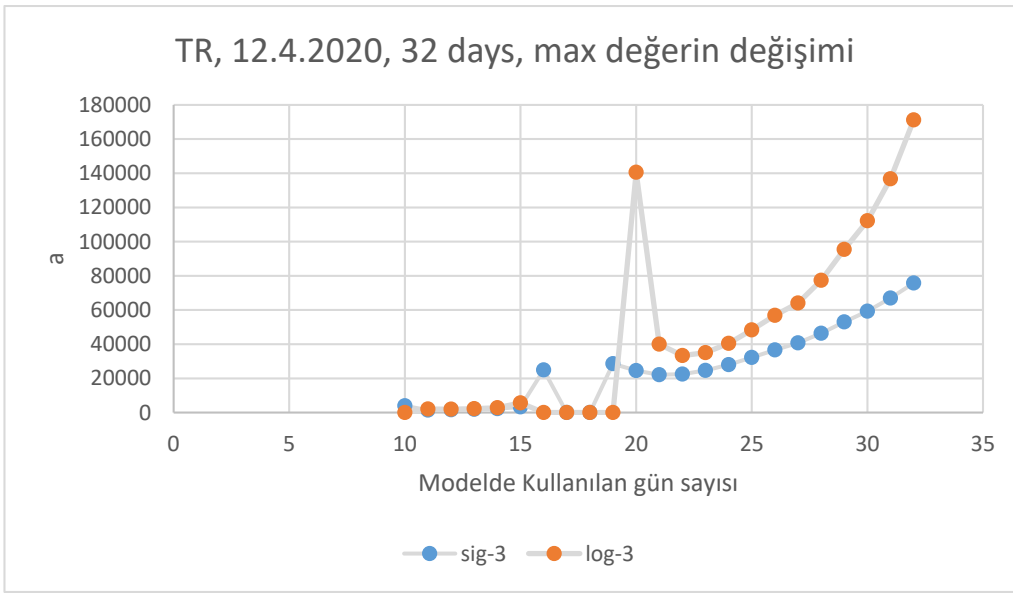
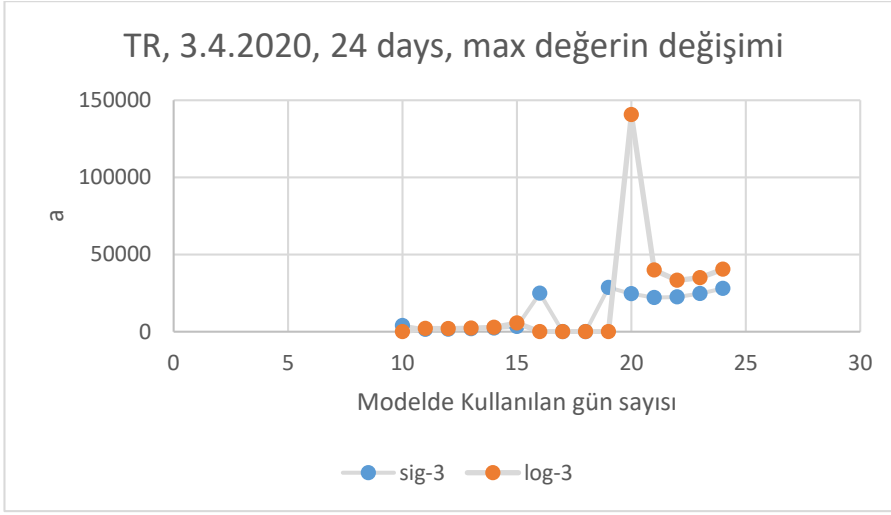


TR(24)	sig-3	log-3
a	28043.25	40490.99
b	2.861378	-5.49789
x0	21.1376	23.79911
R^2	0.99855	0.99867

TR(32)	sig-3	log-3
a	75753.42	171181.5
b	4.688543	-3.96448
x0	28.69752	39.59832
R^2	0.997255	0.999163

Yorum: Veri doyuma ulaşmamıştır, HALA sigmoidal bölgede olup her iki fonksiyon da oldukça yakın bir şekilde veriyi modellemektedir. 50 gün civarında doyuma ulaşacak regresyon modelleri 29-40 günde bir büküm noktası öngörmektedir. Log-3 modeli daha yüksek bir maksimum öngörmektedir. Bu değerler bir önceki rapora göre artmıştır.

TR verisinin tamamı yerine ilk-n günü kullanılarak sig-3 ve log-3 regresyon modelleri elde edilmiş ve bunların doyum değerleri ("a" parametresi) aşağıda grafikte verilmiştir.



YORUM TR verisi anormal zıplamalar yapmaktadır: 18 ve 20 gn arası, ortalama 19 gn. Dięer lkelerdeki desene uygun olarak $19+15=34$ gnde ikinci bir zıplama ve ardından doyuma ulařmanın $34+20=54$ gnde oluřması beklenebilir.

GENEL YORUM

Sistem dinamikleri deęiřkendir, uygulanan yaklařımlar ve veri yayınlama-analiz politikaları farklıdır, saęlıklı bir ngrnn yapılabilmesi iin

- Verinin anlamlı ykseliře gemesinin zerinden 35 gn gemesi gerekli grlmektedir.

Temel Varsayım (veriden elde edilen desen)

Her bir gn iin (ilk n gn kullanılarak) yapılan sig-3 ve log-3 modellerinde log-3 hep sig-3'e gre daha yksek bir deęeri ngrmektedir. in ve Kore verisinde zıplama-dřme-zıplama-dřme ve modellerin yaklařması řeklinde bir desen gzlenmektedir. Bu desen, İtalya ve İřpanya'da da gzlendięinden, İtalya ve İřpanya iin rapor v2'de doyuma ulařma ve nihai maksimum tahminleri yapılmıřtır. Bu tahminler USA ve TR iin sadece bir zıplama gzlendięinden, USA ve TR iin ikinci zıplama tarihi ngrlmřtr. Gzlenen bu tarihler (bařlangıtan itibaren gn sayısı) ařaęıda tabloda verilmiřtir. ZIPLAMA TARİHİ modelden elde edilecek maksimum deęerin zıplayacaęı tarihtir, gerek verinin zıplayacaęı bir tarih deęildir!

Rapor v2 tablosu

ÜLKE (GÜN, 3.4.2020 itibariyle)	Zıplama-1 (GÜN)	Zıplama-2 (GÜN)	Modellerin birleşmesi (GÜN)	Doyum (maksimum değeri) (ADET)
ÇİN (93)	25	50	75	82000
KORE (73)	25	41	55	9300
İTALYA (64)	40	53	73 (TAHMİN, 12 NİSAN)	155000-160000 (TAHMİN)
İSPANYA (63)	42	56	76 (TAHMİN, 16 NİSAN)	190000-200000 (TAHMİN)
USA (74)	60	75-80 (TAHMİN)	95 (TAHMİN)	
TR (24)	19	34-39 (TAHMİN)	54 (TAHMİN)	

10 Nisan'da USA zıplama yapmış olursa,

12 Nisan'da İTALYA doyuma ulaşırsa,

16 Nisan'da İspanya doyuma ulaşırsa,

18 Nisan'da TR zıplama yapmış olursa

Modellerin doğruluğu için bir gösterge elde edilmiş olacaktır.

BU RAPORDA GÖZLENEN DURUM:

10 Nisan'da USA zıplama YAPMAMIŞTIR.

12 Nisan'da İTALYA doyuma ULAŞMAMIŞTIR:

16 Nisan'da İspanya'nın doyuma ulaşması olası görülmemektedir.

TR için tahmin bir sonraki raporda doğru/yanlışlanabilecektir.

ÇİN ve KORE için gözlenen dinamikler diğerlerinde gözlenmemektedir, en azından aynı sürelerde gözlenmemektedir. Sig-3 ve log-3 max değerlerin birbirine yakınsaması için gereken süreler uzamıştır.

Bunu,

- c) İspanya, İtalya, USA ve TR'nin uyguladıkları yöntemler ÇİN ve KORE kadar başarılı değildir, diye yorumlamak olasıdır
- d) İspanya, İtalya, USA ve TR'nin vaka sayılarını belirlemede kullandığı test yaklaşımı ÇİN ve KORE'ye göre daha yaygın bir insan grubunu kapsamaktadır diye yorumlamak olasıdır.

TR için sağlıklı bir doyum noktası tahmini için hala erkendir.