

Corona Virüs Salgını Teşhis Sayısı Öngörü Analizi v2

Arif Selçuk Öğrenci

v1: 29.3.2020 (TR için 18 gün)

v2: 3.4.2020, 14:00 (TR için 24 gün)

TÜMÜNÜ OKUMAK İSTEMEYENLER İÇİN SONUÇ

GENEL YORUM

Sistem dinamikleri değişkendir, uygulanan yaklaşımlar ve veri yayınlama-analiz politikaları farklıdır, sağlıklı bir öngörünün yapılabilmesi için

- Verinin anlamlı yükselişe geçmesinin üzerinden 35 gün geçmesi gerekli görülmektedir.

Dinamiğin (önlemler, tedavi) değişmesi sebebiyle günlük değişimlerin etkisi yüksektir, yanıltıcı olabilir. TR verisi için öngörünün sağlıklı olabilmesi için 35. Güne kadar (11 gün daha) beklenmelidir.

Temel Varsayım (veriden elde edilen desen)

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 doyuma ulaşma ve nihai maksimum tahminleri yapılmıştır. USA ve TR için sadece bir zıplama gözleendiğinden, USA ve TR için ikinci zıplama tarihi öngörülmüştür. Gözlenen bu tarihler (başlangıçtan itibaren gün sayısı) aşağıda tabloda verilmiştir. ZIPLAMA TARİHİ modelden elde edilecek maksimum değer in zıplayacağı tarihtir, gerçek verinin zıplayacağı bir tarih değildir! Detay için devamını okuyunuz.

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)	

SONUÇ:

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. Bir sonraki analiz 8 Nisan'da yapılacak ve gidişat yorumlanacaktır.

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 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 günde bir güncellenmesi planlanmaktadır.

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

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

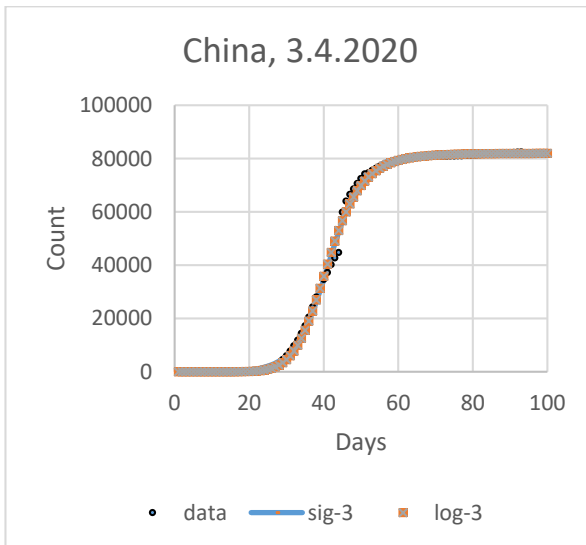
3-parametrelı lojistik (log-3)

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

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

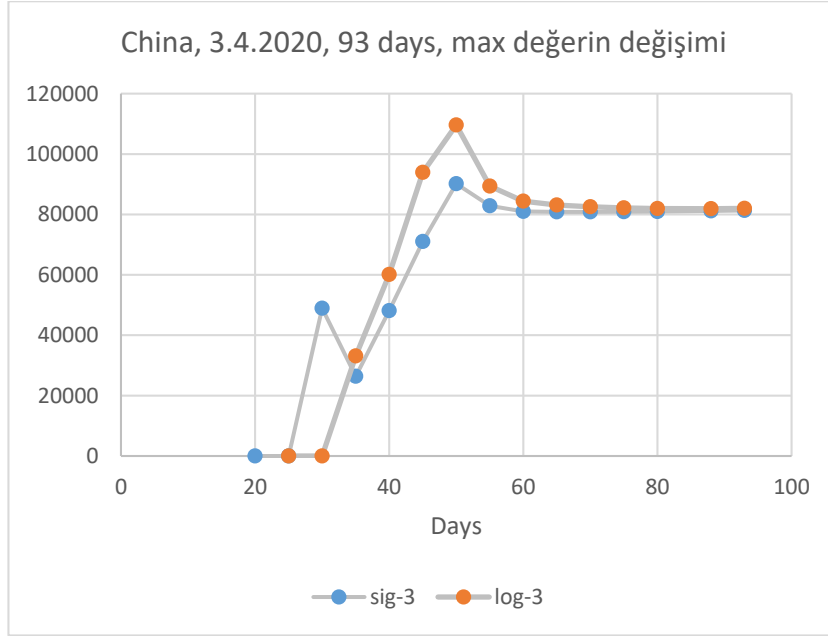


ÇİN (88gün)	sig-3	log-3
a	81138.1	81923.86
b	4.462855	-9.00711
x0	41.31811	41.12196
R ²	0.998362	0.997857

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

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.

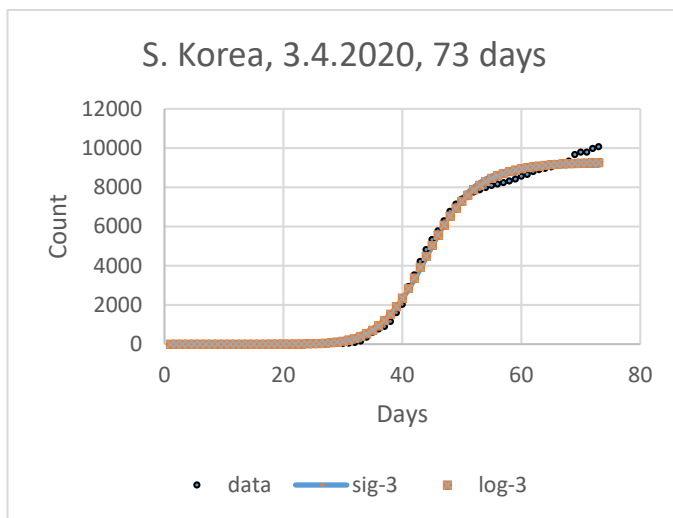
ÇİN verisinin tamamı yerine ilk-n günü kullanılarak sig-3 ve log-3 regresyon modelleri elde edilmiş ve bunların doyum değerleri ("a" parametresi) aşağıda grafikte verilmiştir.



(*) 0 görülen 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 (25 Gün) sonra düşme, tekrar zıplama (50 Gün) ve sonra da her iki modelin birbirine yaklaşması (75 gün) şeklinde gözlenmektedir.

6. G. KORE 73 günlük veri

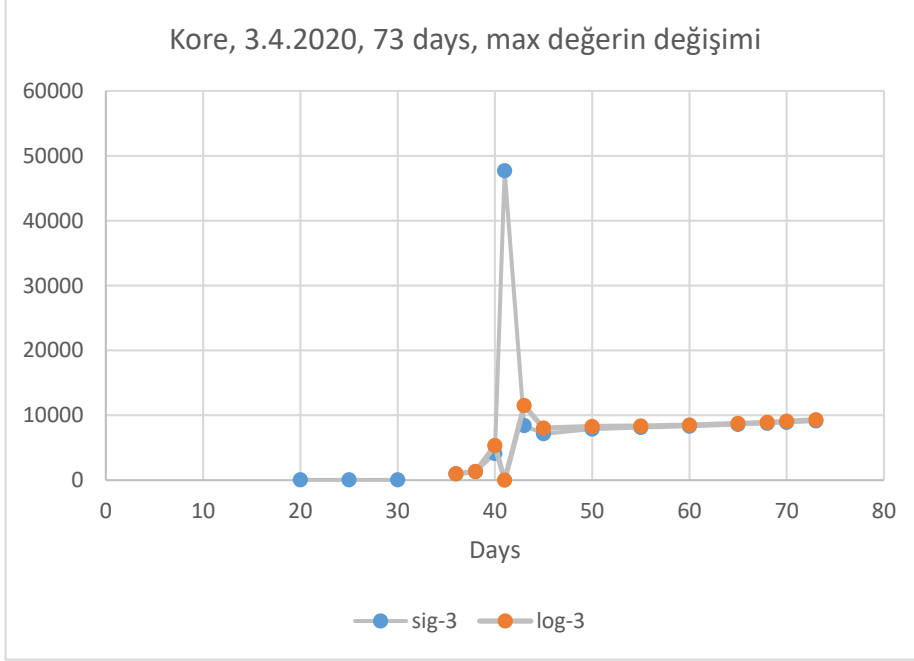


G. KORE	sig-3	log-3
a	8756.041	8897.114
b	3.524001	-11.8825
x0	43.76948	43.76377
R^2	0.997411	0.998371

G. KORE	sig-3	log-3
a	9128.179	9301.795
b	3.929003	-10.6228
x0	44.2795	44.3027
R^2	0.994231	0.995916

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 44 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.

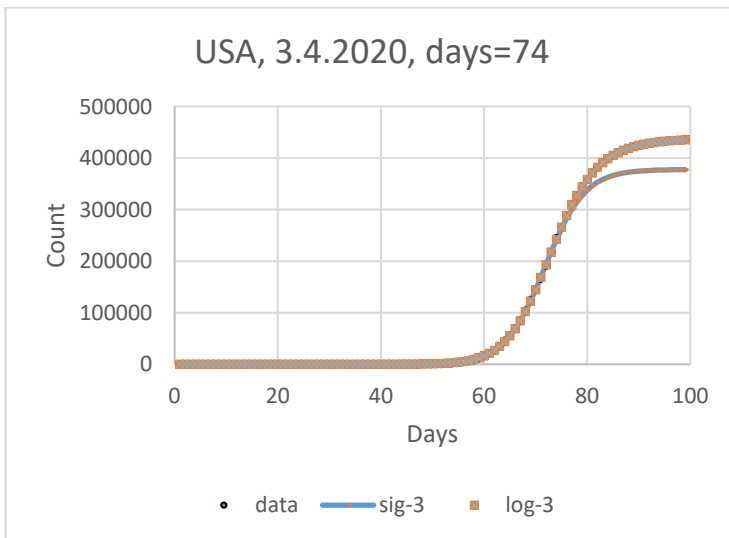


(*) 0 görülen değerler için yapılan regresyonda algoritma yakınsamayı başaramamıştır. Aslında bu noktalarda grafik "uçuk" bir tepe yapmaktadır.

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

74 günlük veri

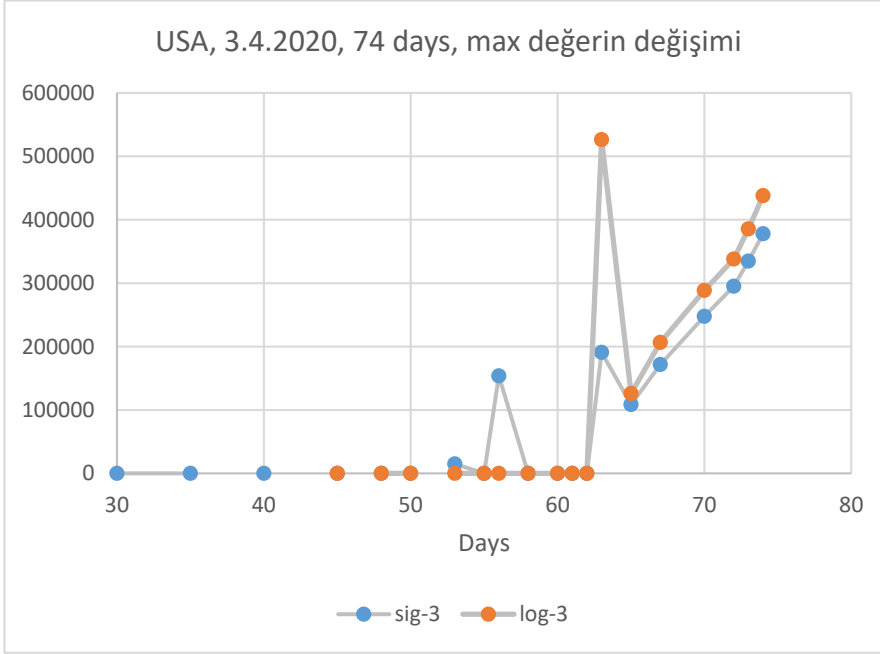


USA(67 gün)	sig-3	log-3
a	171721.5	206513.3
b	2.939834	-20.2975
x0	67.06216	68.18214
R ²	0.999285	0.999337

USA(74gün)	sig-3	log-3
a	377853.3	437860.8
b	3.829695	-16.5932
x0	71.82137	73.05493
R ²	0.999316	0.999564

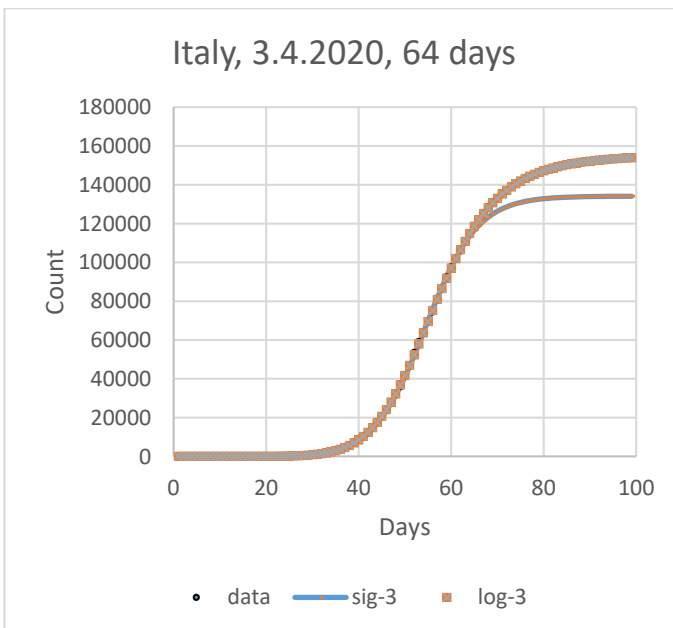
Yorum: Veri doyuma ulaşmamıştır, HALA sigmoidal bölgede olup her iki fonksiyon da oldukça yakın bir şekilde veriyi modellemektedir. 90 gün civarında doyuma ulaşacak regresyon modelleri 67-72 günde bir büküm noktası öngörmektedir. Log-3 modeli daha yüksek bir maksimum öngörmektedir.

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 dinamięi ortaya çıkmamıştır ancak ilk 30+45 =75 günden sonra tam olarak anlaşılması beklenebilir. Birkaç günlük eklemeler bile tahminlerde ciddi farklara yol açmaktadır. ÇİN ve KORE örneklerine bakılırsa, 60+20=80 günde bir tepe (uçuk deęer) yapıp sistemin 100 günde doyuma ulaşması beklenebilir.

8. İTALYA 64 GÜNLÜK VERİ

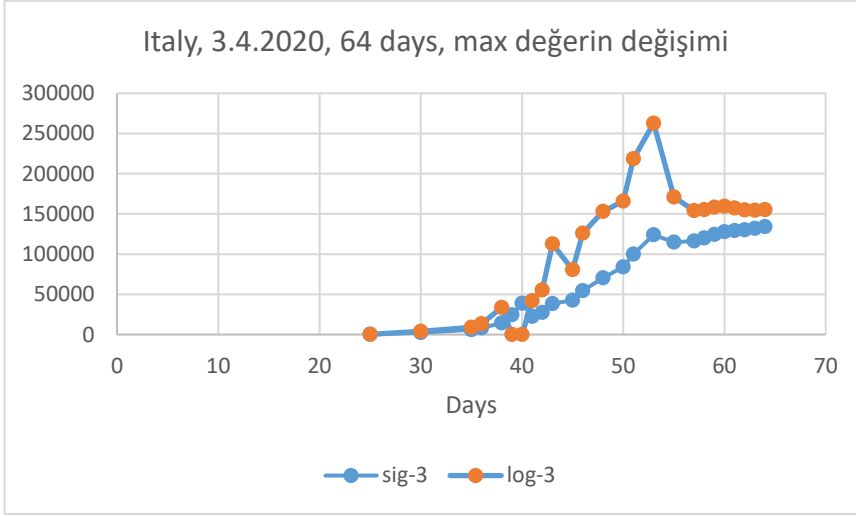


Italy (57)	sig-3	log-3
a	116205.9	154095.5
b	5.103287	-8.29438
x0	52.95111	56.34824
R^2	0.999727	0.999719

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

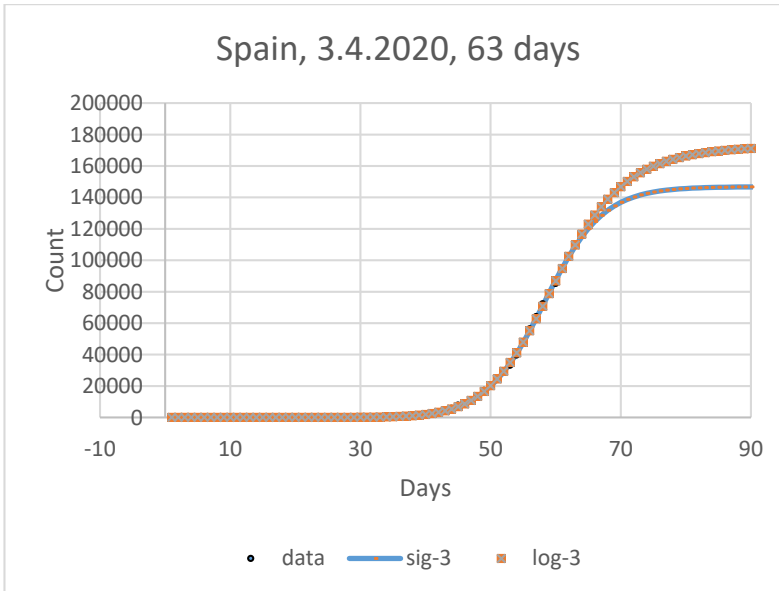
Yorum: Veri doyuma ulaşmamıştır, HALA sigmoidal bölgede olup her iki fonksiyon da oldukça yakın bir şekilde veriyi modellemektedir. 90 gün civarında doyuma ulaşacak regresyon modelleri 55 günde bir büküm noktası öngörmektedir. Log-3 modeli daha yüksek bir maksimum öngörmektedir.

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 zıplam yapıp (sig-3'de de bir zıplama vardır), maksimum değerde ikinci zıplamayı 53 günde yapmış ve sonra iki model birbirine yaklaşımaya başlamıştır. Bu iki modelin $53+20=73$ günde birbirine ulaşması, 155000-160000 arasında doyuma ulaşması beklenmektedir. Verinin 65 (20. Güne kadar düşük olan sayı + 45 günlük büküm, $20+45=65$) günden itibaren "doęru" maksimumu vermesi beklenebilir.

9. İSPANYA 63 GÜNLÜK VERİ

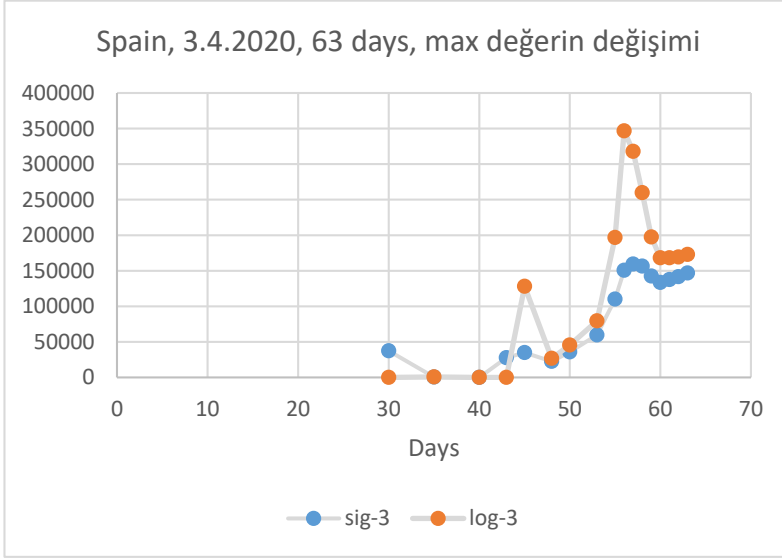


SPAIN	sig-3	log-3
a	150675.1	346608.4
b	4.45008	-10.0974
x0	58.42115	65.97722
R^2	0.998683	0.999182

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

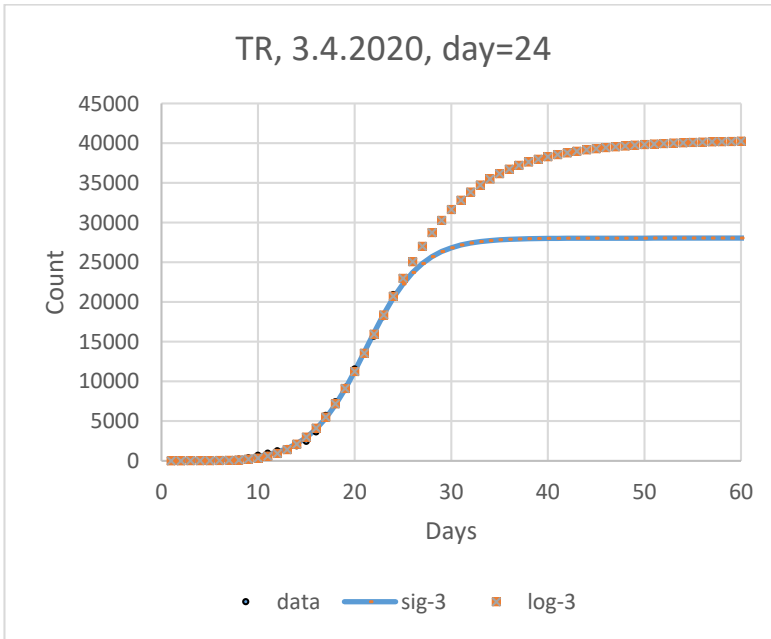
Yorum: Veri doyuma ulaşmamıştır, HALA sigmoidal bölgede olup her iki fonksiyon da oldukça yakın bir şekilde veriyi modellemektedir. 90 gün civarında doyuma ulaşacak regresyon modelleri 60 günde bir büküm noktası öngörmektedir. Log-3 modeli daha yüksek bir maksimum öngörmektedir.

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 diğer 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 24 GÜNLÜK VERİ

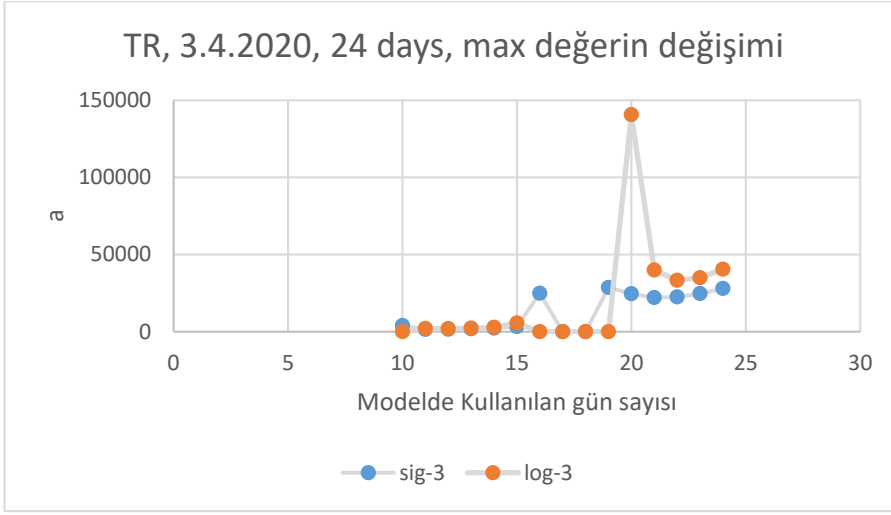


TR(18)	sig-3	log-3
a	2164010	10893223
b	2.994145	-5.13025
x0	34.96183	74.79085
R ²	0.994184	0.989646

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

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 22 günde bir büküm noktası öngörmektedir. Log-3 modeli daha yüksek bir maksimum öngörmektedir.

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 gün arası, ortalama 19 gün. Diğer ülkelerdeki desene uygun olarak $19+15=34$ günde ikinci bir zıplama ve ardından doyuma ulaşmanın $34+20=54$ günde oluşması beklenebilir.

GENEL YORUM

Sistem dinamikleri değışkendir, uygulanan yaklaşımlar ve veri yayınlama-analiz politikaları farklıdır, sağlıklı bir öngörünün yapılabilmesi için

- Verinin anlamlı yükselişe geçmesinin üzerinden 35 gün geçmesi gerekli görülmektedir.

Dinamiğin (önlemler, tedavi) değışmesi sebebiyle günlük değışimlerin etkisi yüksektir, yanıltıcı olabilir. TR verisi için öngörünün sağlıklı olabilmesi için 35. Güne kadar (11 gün daha) beklenmelidir.

Temel Varsayım (veriden elde edilen desen)

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 doyuma ulaşma ve nihai maksimum tahminleri yapılmıştır. USA ve TR için sadece bir zıplama gözleendiğinden, USA ve TR için ikinci zıplama tarihi öngörülmüştür. 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!

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)	

SONUÇ:

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. Bir sonraki analiz 8 Nisan'da yapılacak ve gidışat yorumlanacaktır.