

Study on modeling and monitoring the spread of COVID19 virus

D.Sc. (Tech) Marko Kosunen
markokosunen20@gmail.com
Initiated: March 21, 2020
Current generated version: May 2, 2020



Contents

1 Preface	1
2 Derivation of spreading coefficient limit of declining infection	2
2.1 Parameter estimation with measured data	4
3 Discrete time infection system model	5
3.0.1 Considerations on the feedback model	5
4 Observations and analyses	8
4.1 Action log	8
4.2 Observations	8

1 Preface

This document is created purely from personal interest in mathematics and programming. I have no background in medicine nor in social sciences whatsoever, and everything I present may be commonly known for the people having expertise on those fields. Document does not fulfill the requirements of scientific publication, and it is certainly not peer-reviewed, and is evolving over time. Therefore, the information should be considered more or less unreliable, and you should use your own judgement to evaluate if it is accurate or not.

Regardless of its severity, COVID19 offers a great opportunity to apply mathematics and control theory to societal problems. This work is personal effort to understand ‘the truth’, the laws behind the societal (control) systems, their responses and their stability. It may also reveal discrepancies between the reality and the executed political actions. Ability to identify them I consider very important during the times we are living.

In Puotila, eastern Helsinki, Finland, March 22, 2020

Marko

2 Derivation of spreading coefficient limit of declining infection

In this document I try to develop a model that would help me to understand the essential figures and numbers that we should be monitored and measured in order to distinguish the effective and ineffective measures in limiting the spread of the COVID19 virus.

I start the derivation from the measured reproduction number R_0 . We know that the nature of the growth of the infection is exponential which inherently means the disease is spread by infected persons infecting the uninfected ones over certain period of time. Later on in Chapter 3 I add a more detailed discrete time system model to reveal the effects of the feedback delays. That model is still under construction, and I need to fit the system parameter presented in the following to that model. With COVID19 there is lots of statistics about number of infected persons, daily new cases, deaths and recoveries. Based on these statistics we may derive the estimate of the contamination time during which the infected patient spreads the disease in such a manner that the daily proportional growth equals the monitored reproduction number.

Under assumption that infected persons spread the disease with constant fraction of cases per day, the newly infected persons tomorrow can be expressed as

$$y(n) = Ky(n-1) \quad (2.1)$$

$$K = \frac{y(n)}{y(n-1)} = \frac{y(n) - y(n-1)}{y(n-1)} + 1 \quad (2.2)$$

where $y(n)$ is the number of infected persons on day n . Without limiting measures K is assumed here to be relatively constant.

We may express K as

$$K = 1 + S \quad (2.3)$$

$$S = \frac{y(n) - y(n-1)}{y(n-1)}, \quad (2.4)$$

where S is spreading coefficient, the *ratio of daily increase of cases to active cases*. It should be noted that any estimation error that is in fixed proportion to actually confirmed cases is canceled in calculation of S . Thus we can state that any measure that gives us proportionally relatively stable information will provide us accurate information about S .

As shown before, K and S can be measured. So can be the reproduction number R_0 , which for the COVID19 have been estimated to be $R_0 = 2.06 \dots 2.52$ [1]. These two expansion mechanisms are related, and they should match at least in early phases of epidemic, when deaths, recoveries, countermeasures, and developing immunity do not hinder the spreading. The relation can be expressed

as

$$K^{T_c} = 1 + R_0 \quad (2.5)$$

$$T_c = \frac{\ln(1 + R_0)}{\ln(K)}, \quad (2.6)$$

where T_c is the contamination time, i.e. the effective time during which the patients spread the disease. It can be thought as a time before patient gets so visibly sick that he will be isolated/quarantined from the community and thus does not spread the disease. This time is specific to disease and we may assume here it is relatively constant.

We may now denote the spreading coefficient, measured as the ratio of increase to the active cases, in the beginning of the epidemic as S_0 . Thus

$$T_c = \frac{\ln(1 + R_0)}{\ln(1 + S_0)} \quad (2.7)$$

For present COVID19, S_0 seems to be, with reasonable number of cases, between 0.3 and 0.2, thus we may use $S_0 = 0.25$ in our example calculations. We may now perform an example calculation for contamination period

$$S_0 = 0.25 \quad (2.8)$$

$$R_0 = 2.5 \quad (2.9)$$

$$T_c = \frac{\ln(1 + R_0)}{\ln(1 + S_0)} = 5.6. \quad (2.10)$$

Thus the estimate of the contamination time is 5.6 days, and we assume that to remain constant.

Under ongoing epidemic, public authorities and private persons perform measures to reduce S in order to hinder the spread of the disease. However, as S and R_0 are connected, we may try to calculate *how low we should reduce S in order to stop the epidemic under assumption that T_c does not change*. Stopping the epidemic means $R_0 < 1$. We may now calculate the limit S_l that should result in declining numbers of active patients.

$$(1 + S_l)^{T_c} < 2 \quad (2.11)$$

$$S_l < 2^{\frac{\ln(1+S_0)}{\ln(1+R_0)}} - 1. \quad (2.12)$$

With the given values for current COVID19 outbreak $S_l < 0.13$.

In Table 2.1 the obtained values under various assumptions are presented. These values reveal that with given parameter spread there is no *practical* difference in contamination time, it lies between 3.7 to 6.2 days. This means practically two days difference, easily masked by individual variation and real life uncertainties. Furthermore we may quite certainly say there is no chance of epidemic to stop with spreading factors larger than 0.2. Assuming 0.15 might be wishful thinking but may work, and 0.1 should be quite safe.

As the global data for the cases have been collected and made generally available it is possible to evaluate the theory by comparing it to existing data. That comparison will be presented in Chapter 4.

Table 2.1: Spreading coefficient limit value under various assumptions on reproduction number R_0 and initial spreading coefficient S_0

R_0	S_0	T_c	S_l
2	0.25	4.9	0.15
2	0.3	4.2	0.18
2	0.4	3.7	0.21
2.5	0.25	5.6	0.13
2.5	0.3	4.8	0.16
2.5	0.4	4.2	0.18
3	0.25	6.2	0.11
3	0.3	5.3	0.14
3	0.4	4.6	0.18

2.1 Parameter estimation with measured data

On March 23 friend of mine, Teemu Lehtimäki suggested that the measured relative growth e.g. in Sweden and the trend of measured cases do not support the R_0 parameter value 2.x, but that the correct number could be higher. The arising is, what is the most reliable way to estimate T_t in the beginning of the epidemic, when the behaviour of the individuals is not affected by the counter measures, but when the cases are relatively scarce. I think one possible method is LMS curve fitting based on measured relative growth and difference of cases $y(n) - y(n-x)$.

3 Discrete time infection system model

In order to gain more insight on input-output relation (i.e the transfer function), stability and critical parameters to be identified, I developed a simple discrete-time system model.

Fig 3.1 depicts a discrete time infection system model in Z-domain. $X(Z)$ are the infectious patients entering the community with spreading factor S . $X_q(Z)$ are the infectious patients entering the system through quarantine. They are known to be infectious, but they are isolated.

Fig 3.2 presents the infection feedback eventually causing the exponential growth. Every infectious patient X will initiate an infection chain over T_c days, adding S new infections.

Due to the feedback through $W(Z)$, the system will eventually have positive feedback causing exponential type of growth. The parameters characterizing the system behaviour are listed in Table 3.1.

3.0.1 Considerations on the feedback model

In the beginning of the April I had already several bothersome days of thinking the accuracy of the feedback model. There seemed to be a discrepancy between the exponential growth model I describe in the beginning of this chapter, and the R_0 -based exponential model. As per now, (April 2), I think I have developed an explanation that helps to get the modeling forward.

The parameters s_k define the feedback system $W(Z)$. In simplest case, $s_0 = R_0$, and the other feedback coefficients are zero. This will provide the growth due to feedback on a single day with the delay of T_i . To use relative growth as a measure in this kind of a system, we should estimate the T_i accurately, and use the past values $y(n - T_i)$ in calculation of relative growth. Currently I use $y(n - 1)$ in calculation of the relative growth, and it does not take into account the incubation period at all. Adding the delay would increase the relative growth values on by the average gain of over incubation period, (which should be R).

So, the question remains why my model seems to work, even though it is timing wise incorrect. My current thoughts are the following and subject to change. The way I originally thought the exponential growth process is, that a patient infects persons as geometric series so that total number of persons over the contamination person is increased by reproduction number R . This is technically

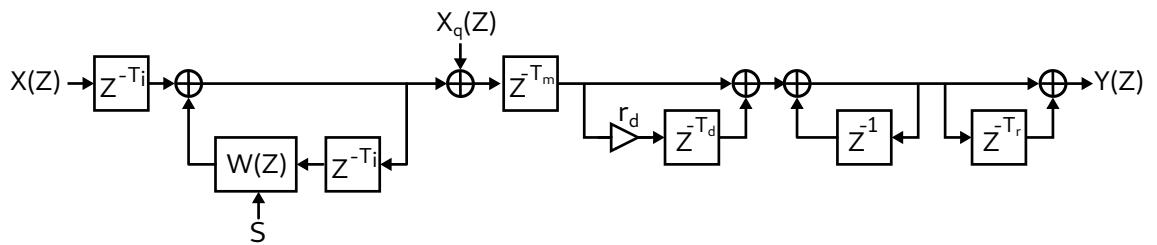


Figure 3.1: Discrete time Infection system in Z-domain.



Figure 3.2: Infection feedback system $W(Z)$

Table 3.1: Infection system parameters

Parameter	Description
S	Spreading factor.
T_i	Incubation period. It is assumed that the patient is not infectious during this time.
T_c	Contamination period. Patient spreads the disease during this time until is isolated from the population because of sickness or treatment.
T_m	Monitoring delay. The time from the patient becoming infectious to registration of the case.
r_d	Death rate. The share of the registered patients dying.
T_d	Delay of death. Time from registration to death.
T_r	Recovery time.

incorrect. However, it can be modeled as s_k values in Fig. 3.2

$$s_k = s(1+s)^k, \quad k = [0, T_c - 1]. \quad (3.1)$$

This is the sum of the growth after T_c , which compensates the lack of T_i . Actually, as

$$s_{T_c} = (1+s)^{T_c} - 1 = R, \quad (3.2)$$

the model works exactly as the incubation period were T_c . So the timing error in the relative growth estimate is the difference of the contamination period T_c , and incubation period T_i . Most likely this is not zero, but the question is, is that difference typically smaller than the estimation error of the incubation period? If it is, it is, using this model provides better estimate than the discrete time model with estimated incubation period. One can also think, that the incubation period is the period over which the patient is infectious but not sick enough to isolate himself. At the end of the incubation period, the patient has infected R patients. This effectively means $T_c = T_i$, which would mean that the model is accurate, and measuring the relative growth would reflect quite accurately the characteristics of the system. It also distributes the exponential growth over the T_c with exponential emphasis on the last days. More considerations are required to determine if this is accurate enough.

The remaining delay is the measurement delay T_m , which is not related to the disease, but the system measuring it. This does not affect the relative growth

number, only delays the result by T_m . It should also be noted, that we may measure the contamination period by monitoring the relative growth at the point where the increase of the number of active cases stops.

Therefore following observations about the nature of the system can be made.

- System is not linear time invariant, but can be considered piecewise linear over the coherence time on S , i.e. during the periods S is not altered, and linear time invariant, if S is considered constant.
- As the fraction of recovered and dead patiences increases, the S is decreased. This phenomenon is not included in this model, indicating infinite population.
- Timing relation of the signal and S in the feedback path indicates that the effect of changes in policies affecting S are immediate to the spreading of the disease, and visible in the statistics after the measurement delay T_m .

4 Observations and analyses

To evaluate the validity of the above theory, I have developed and used analysis software available https://github.com/mkosunen/Covid19_SydeKick. Software fetches the data automatically from Johns Hopkins database and plots the relative growth and the number of current cases for chosen countries under interest. Observations of the growth and number of cases will be compared to the presented theory so the theory can be adjusted accordingly, or the root causes /explanations for the anomalies will be developed.

4.1 Action log

To this section I try to log the implemented governmental actions that supposedly decrease S . These actions can be compared to data to evaluate the effectiveness of the actions. A quite comprehensive list of actions in Finland is presented in <https://timoheinonen.fi/koronapaivakirja-nain-covid-19-mullisti-maailman>.

April 24

Recently I added a trend (LMS linear fit) line to daily growth monitoring. Today I noticed that the ratio of the first and the last value of 17-day trend is a measure of efficiency of restrictions applied in different countries. Added that efficiency number to the curves. Later on, plotting the efficiency curve over the course of the epidemic can be used to present the 'efficiency history'.

4.2 Observations

Covid cases in Finland

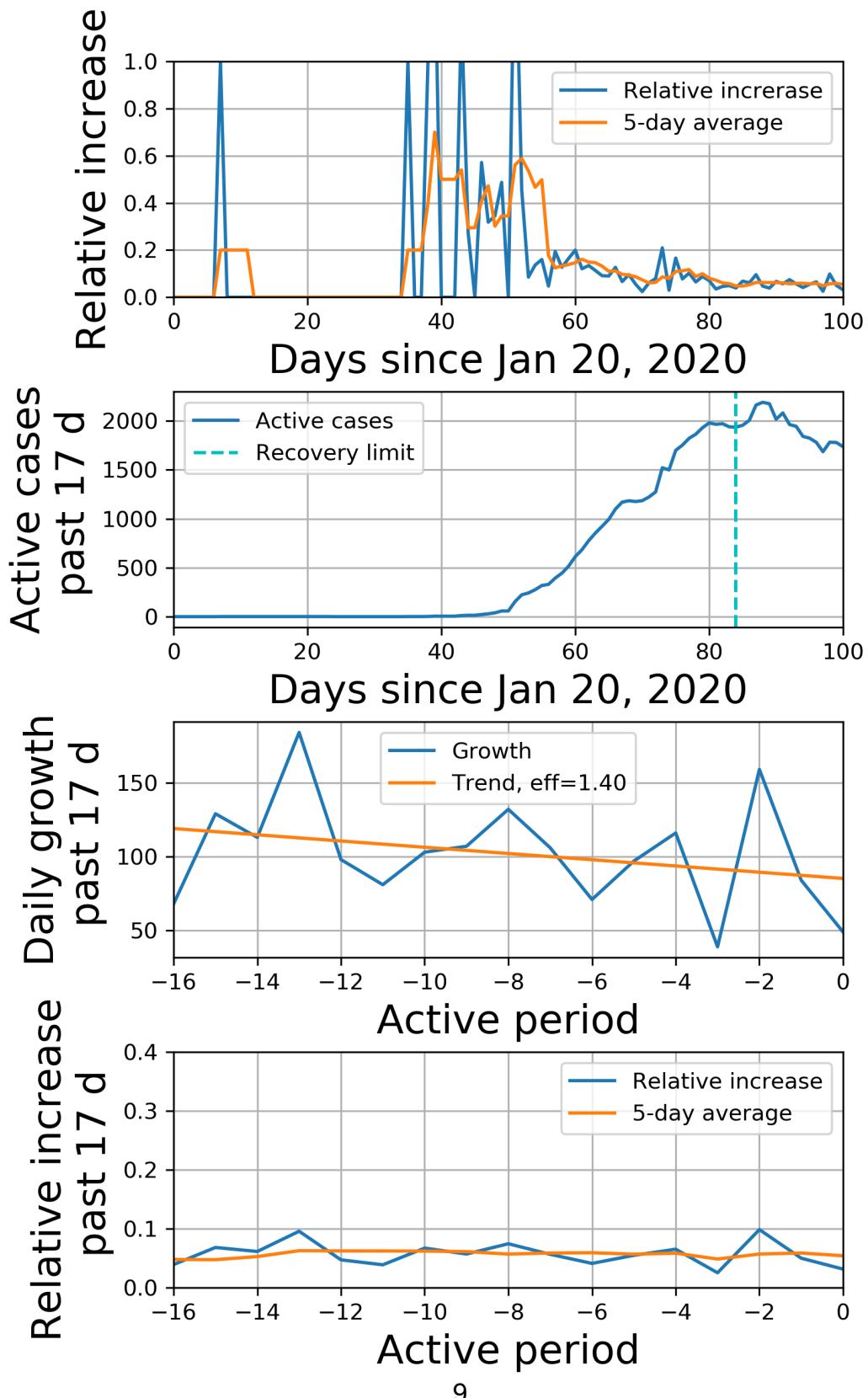


Figure 4.1: Relative growth and number of cases in Finland.

Covid cases in Sweden

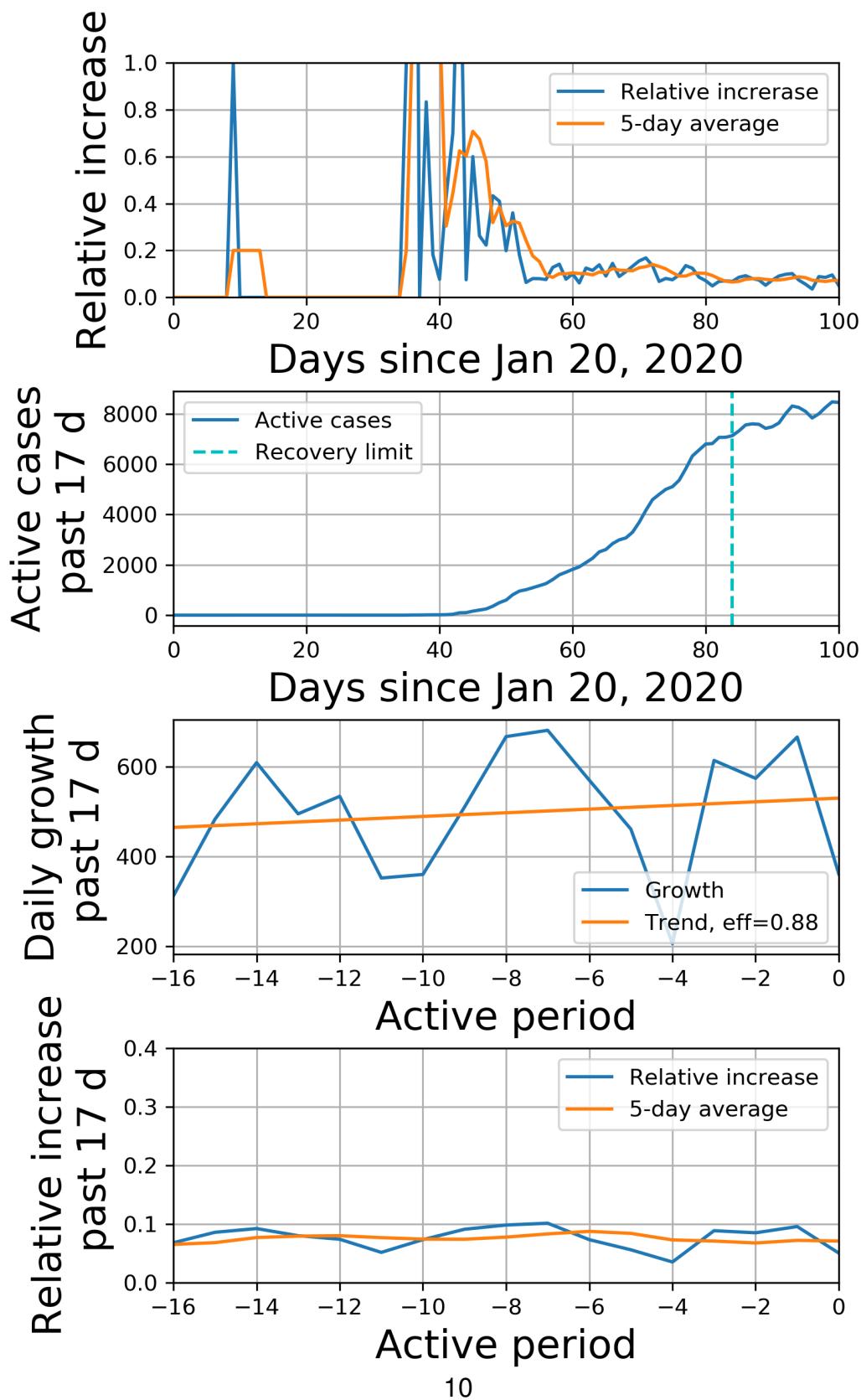


Figure 4.2: Relative growth and number of cases in Sweden.

Covid cases in Austria

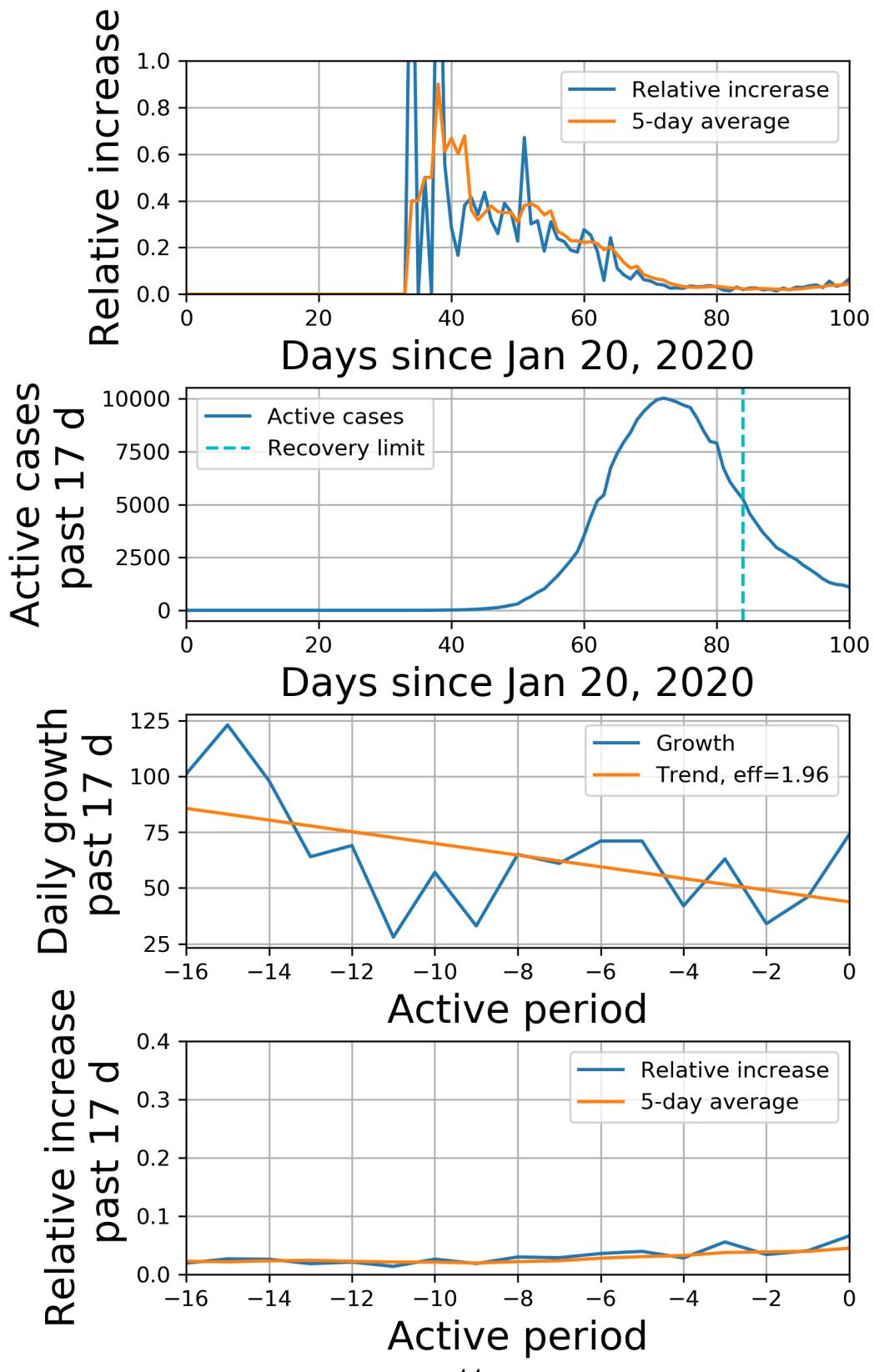


Figure 4.3: Relative growth and number of cases in Austria.

Covid cases in Norway

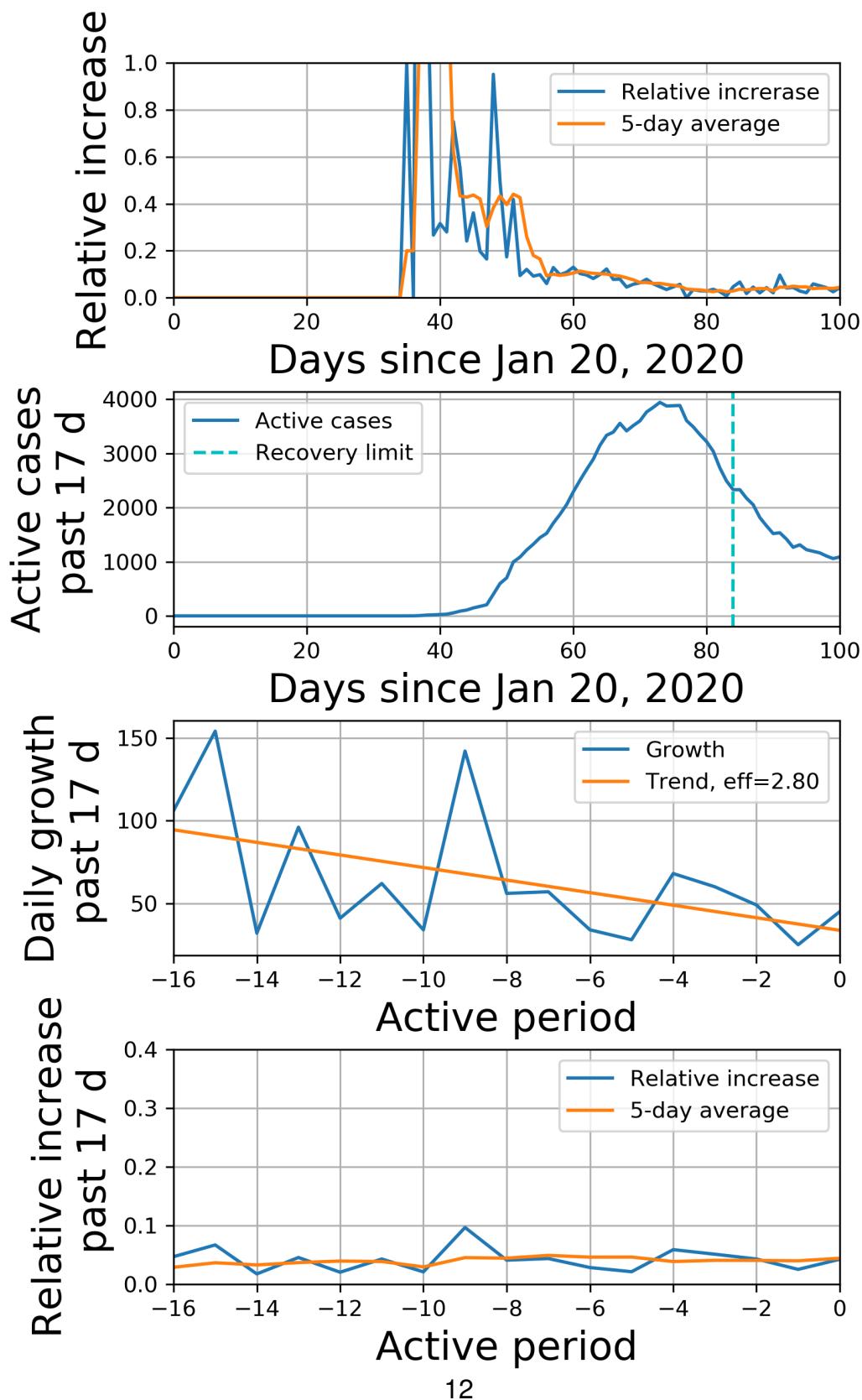


Figure 4.4: Relative growth and number of cases in Norway.

Covid cases in Denmark

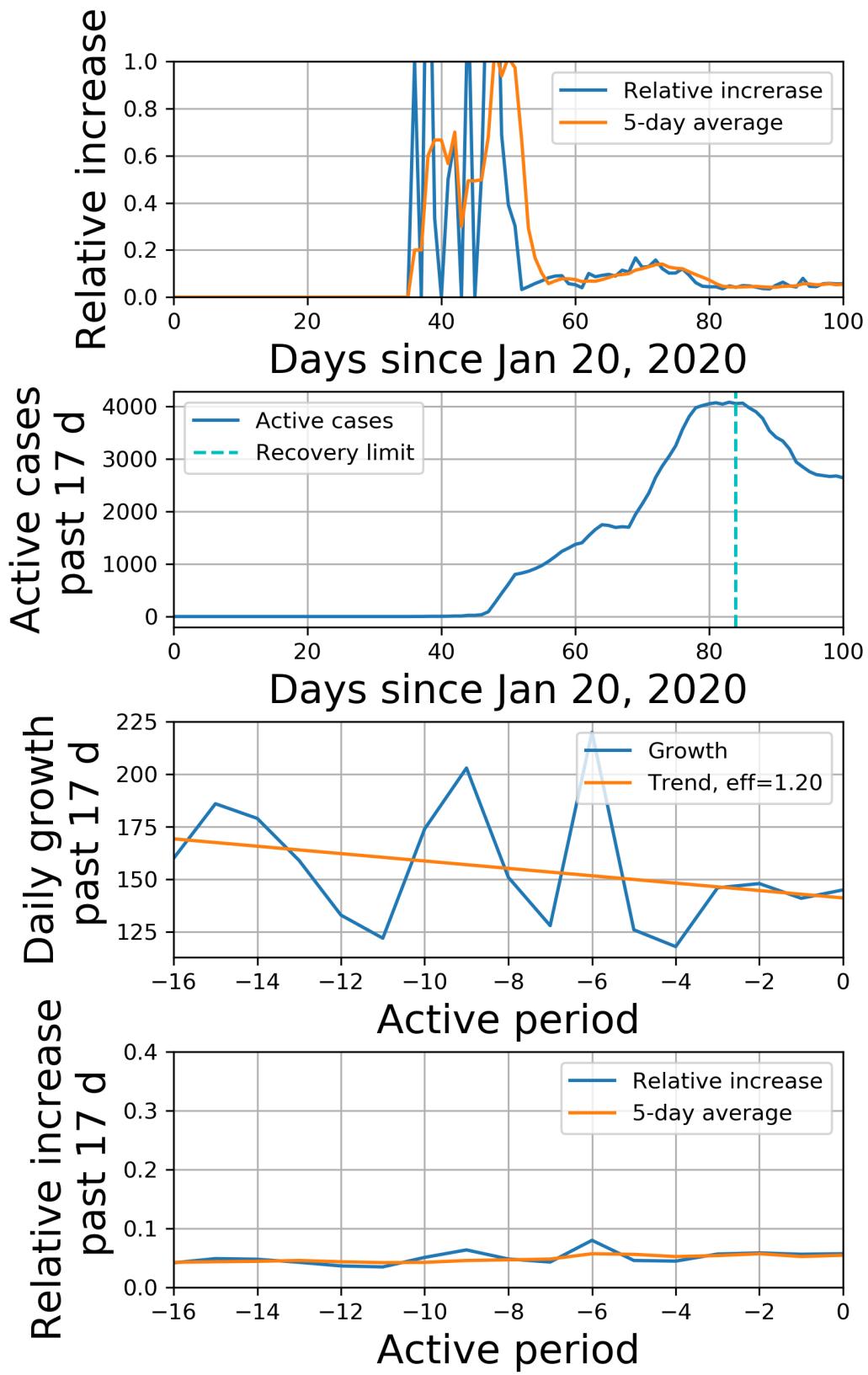
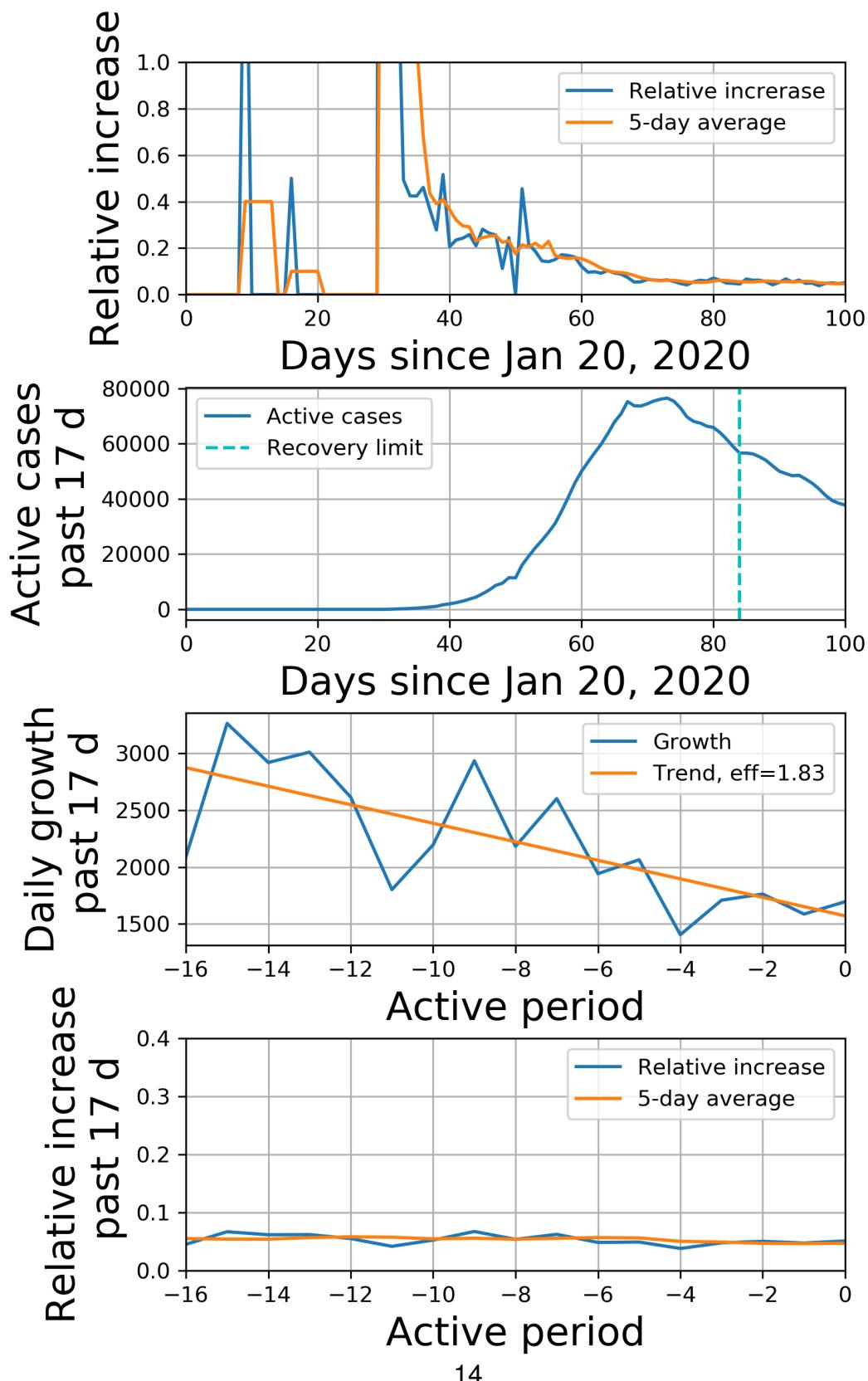


Figure 4.5: Relative growth and number of cases in Denmark.

Covid cases in Italy



14

Figure 4.6: Relative growth and number of cases in Italy.

Covid cases in Germany

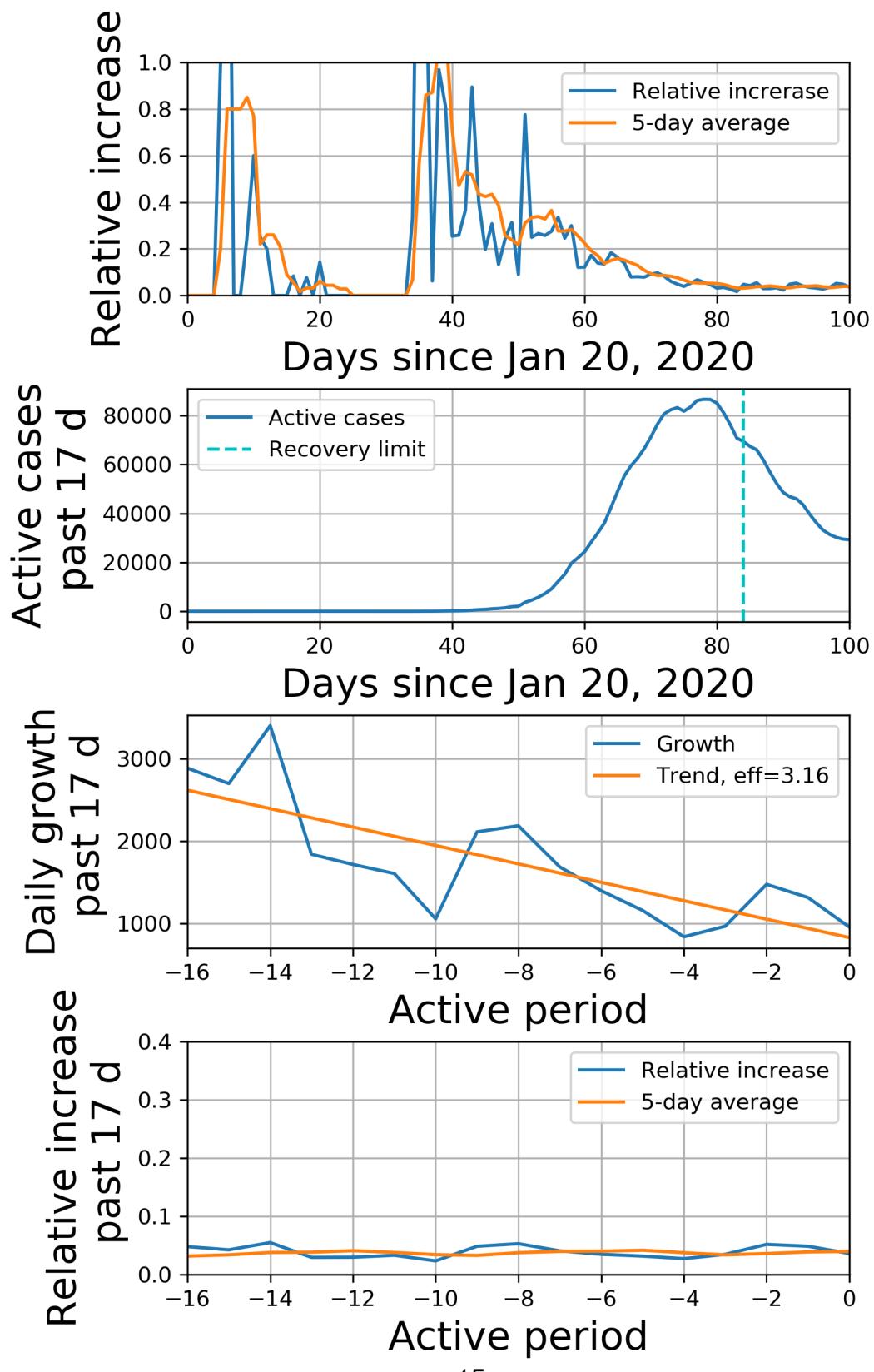


Figure 4.7: Relative growth and number of cases in Germany.

Covid cases in France

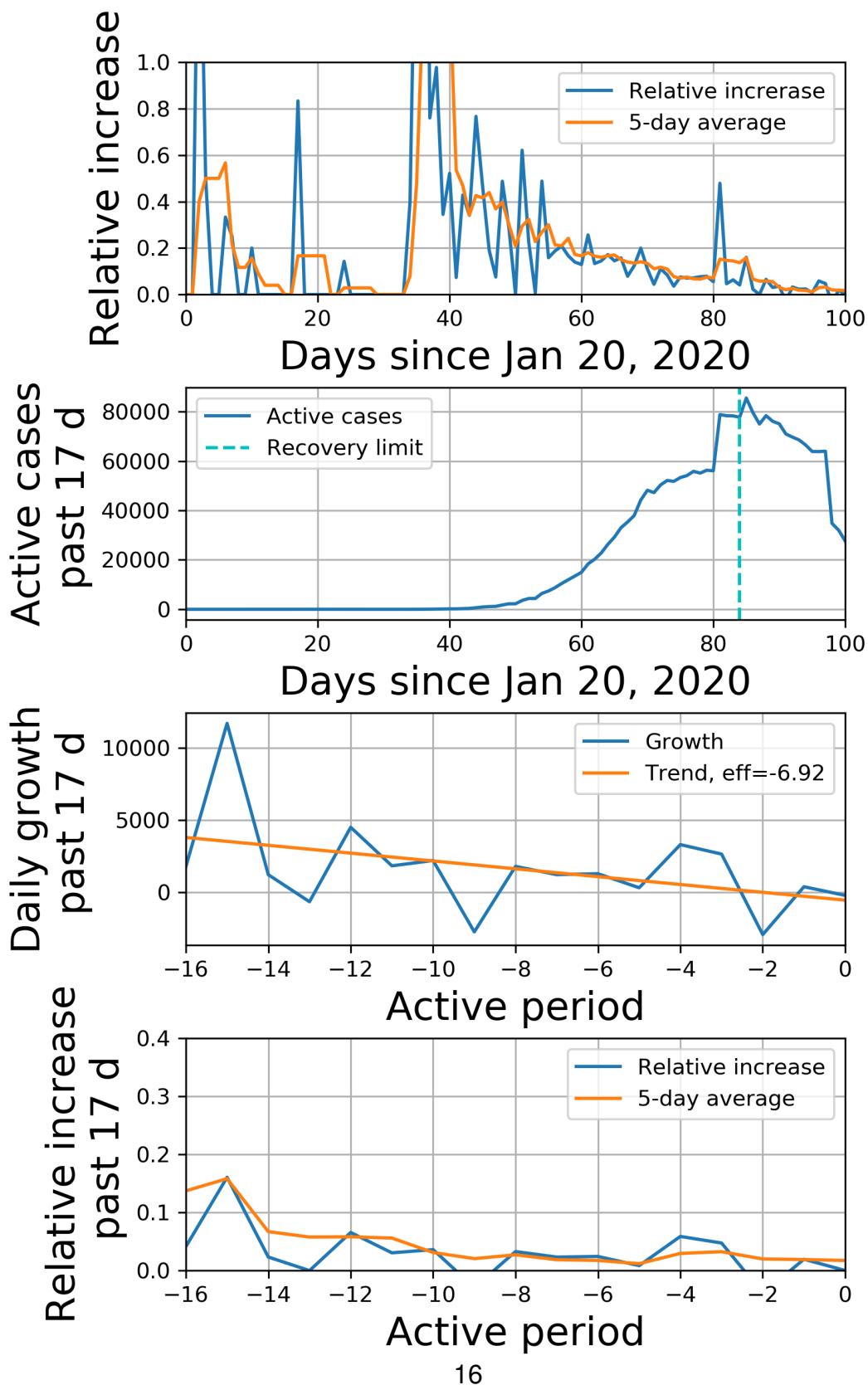


Figure 4.8: Relative growth and number of cases in France.

Covid cases in Spain

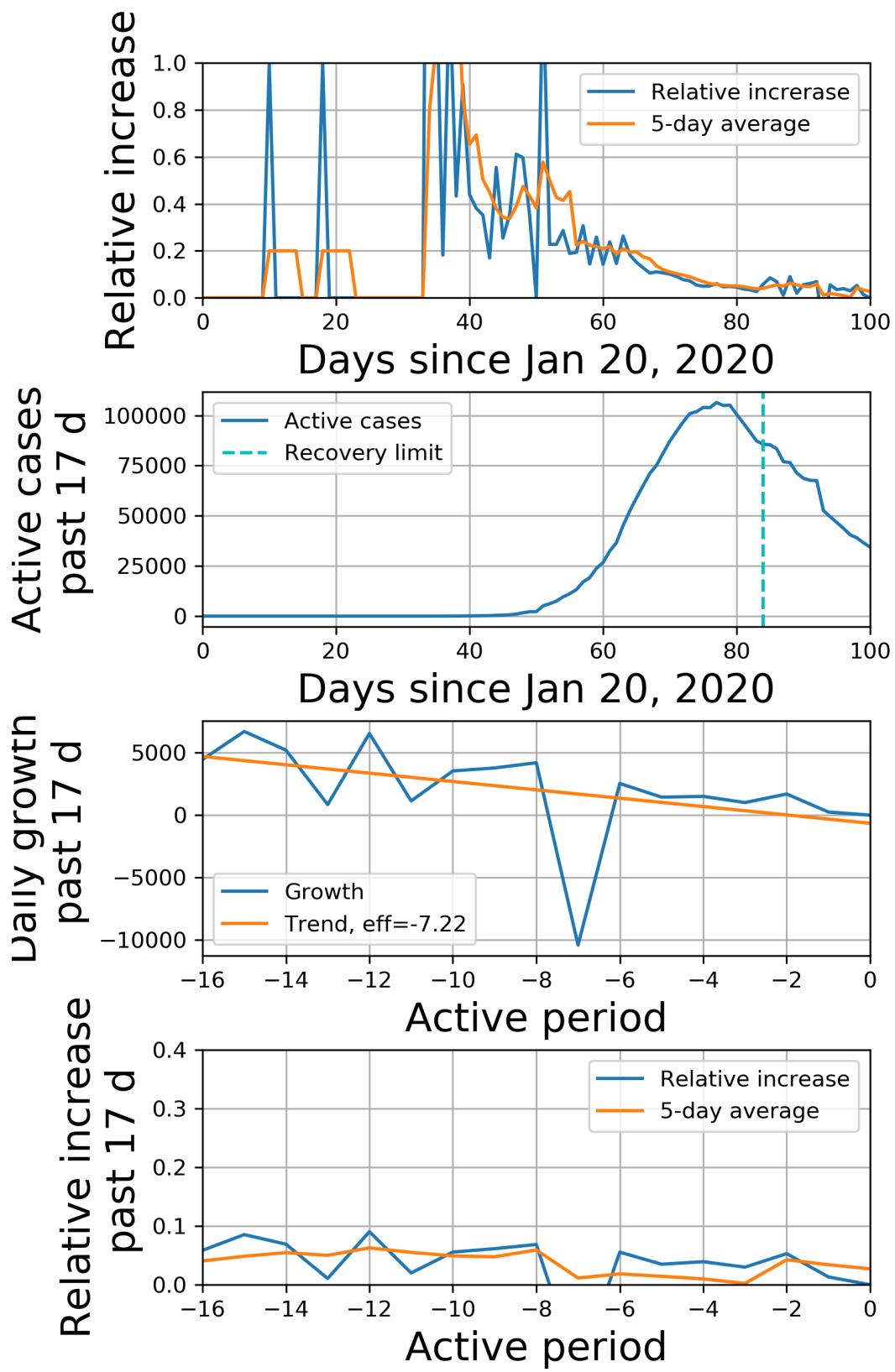
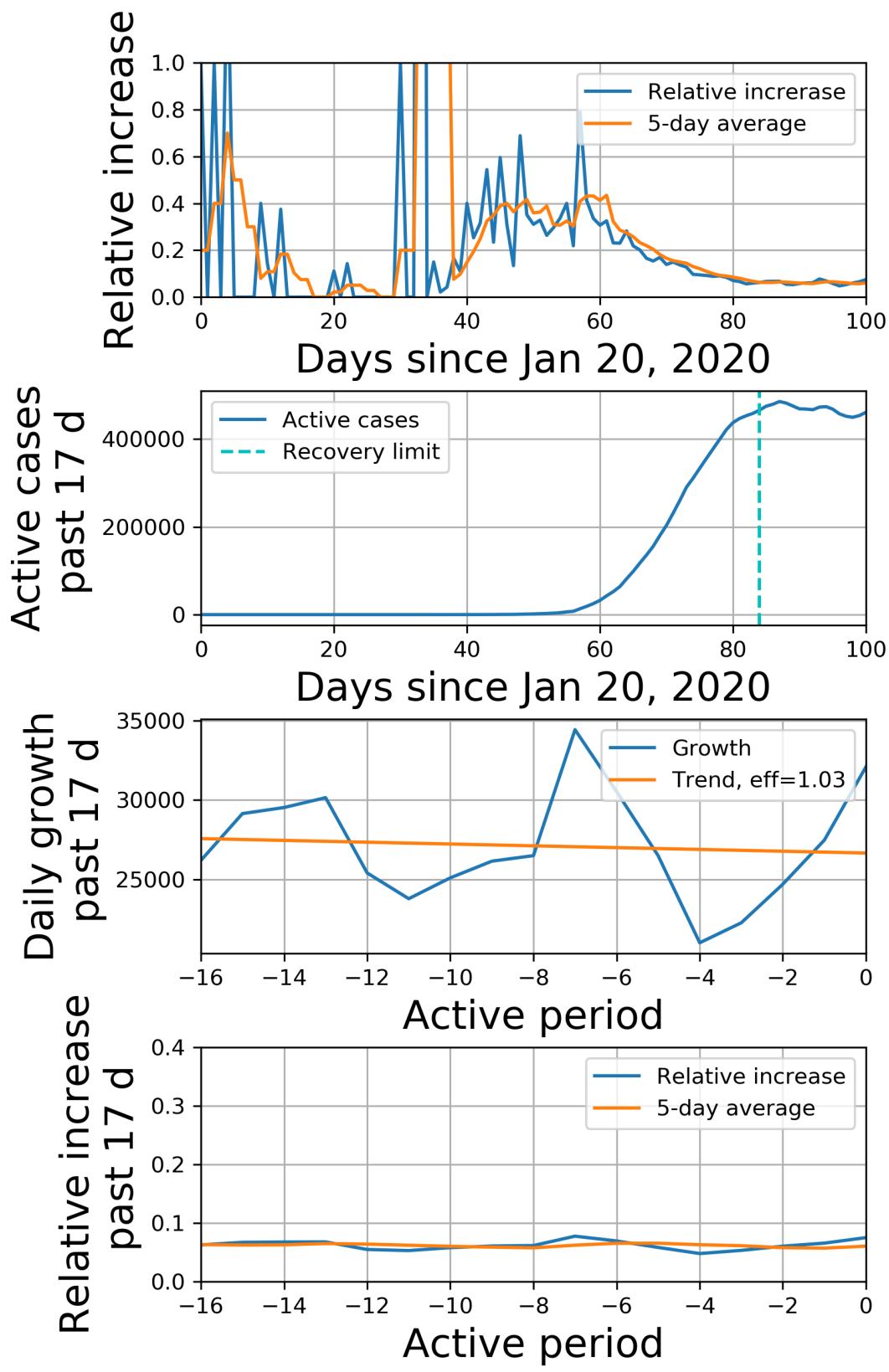


Figure 4.9: Relative growth and number of cases in Spain.

Covid cases in US



18

Figure 4.10: Relative growth and number of cases in United States.

South.png

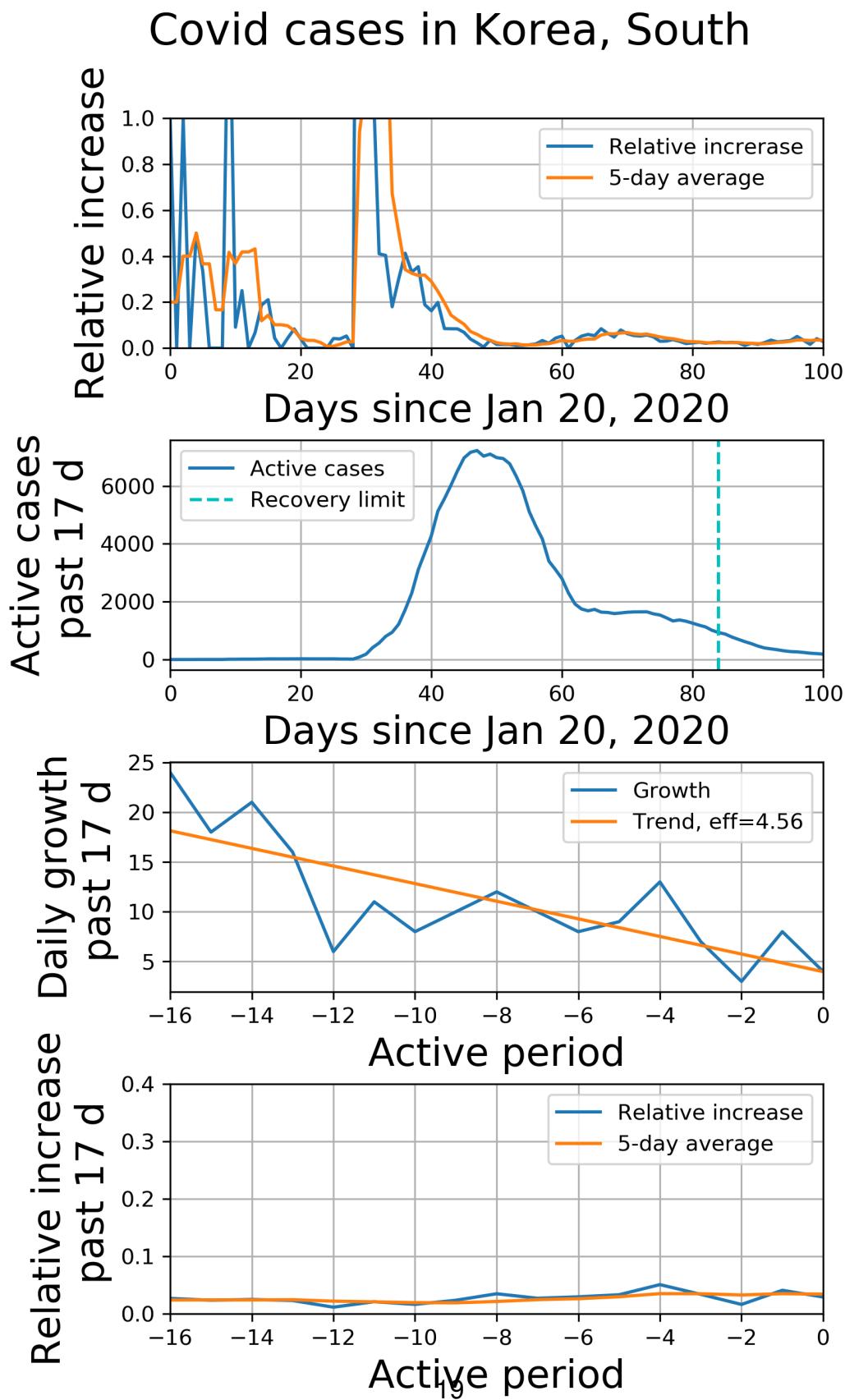


Figure 4.11: Relative growth and number of cases in South Korea.

Covid cases in China

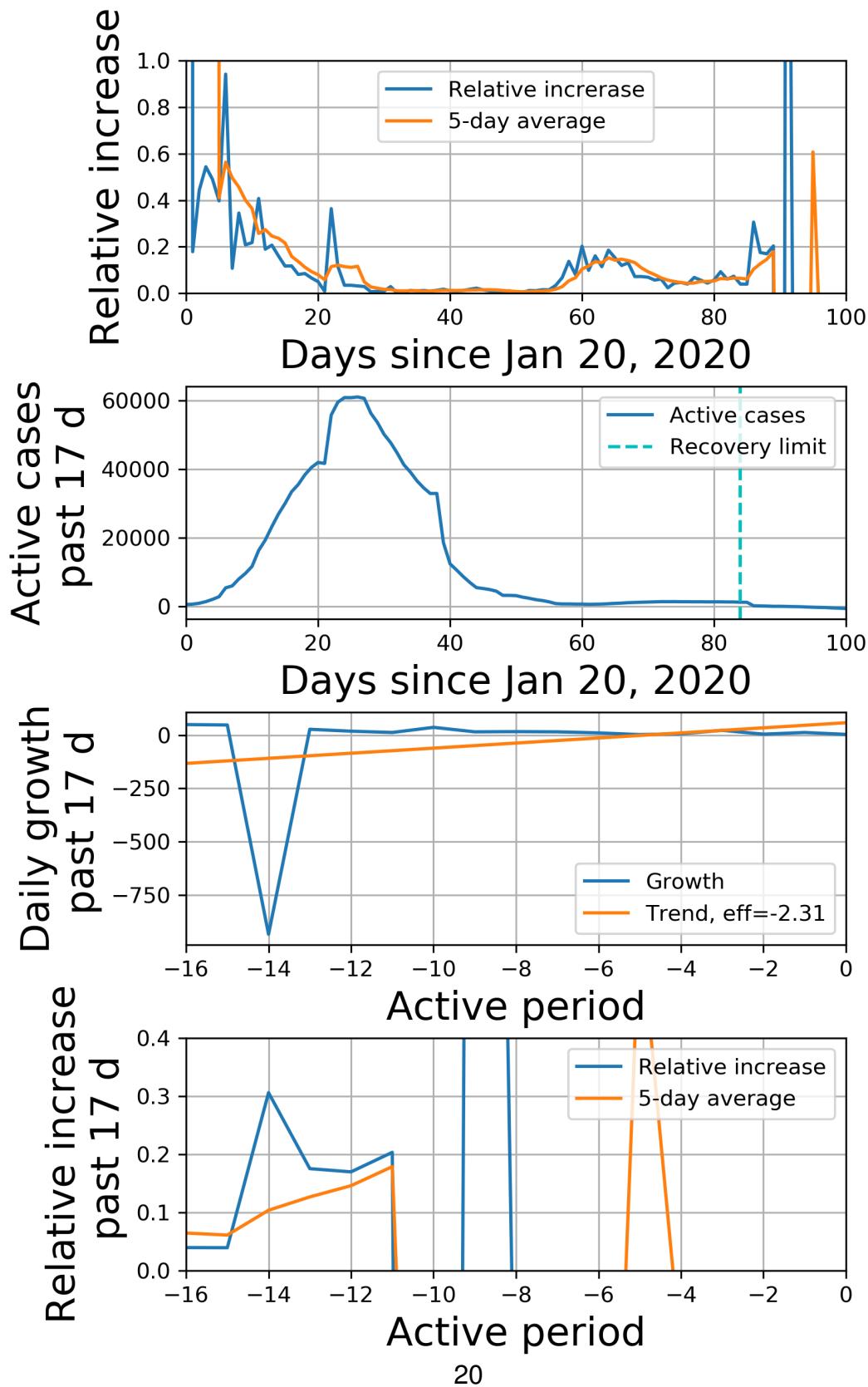


Figure 4.12: Relative growth and number of cases in China.

Covid cases in selected countries

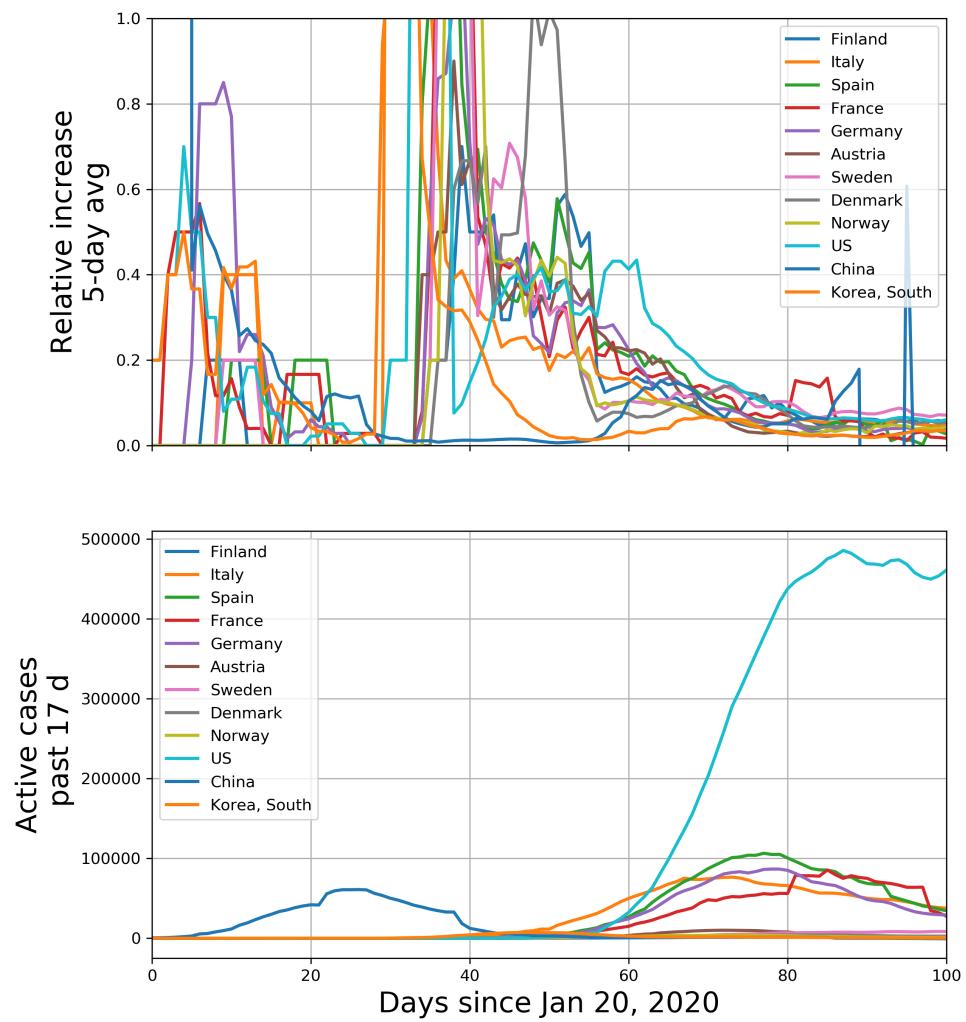


Figure 4.13: Relative growth and number of cases in Selected countries.

Bibliography

- [1] S. Zhang, M. Diao, W. Yu, L. Pei, Z. Lin, and D. Chen, “Estimation of the reproductive number of novel coronavirus (COVID-19) and the probable outbreak size on the Diamond Princess cruise ship: A data-driven analysis,” *International Journal of Infectious Diseases*, vol. 93, pp. 201–204, 2020.