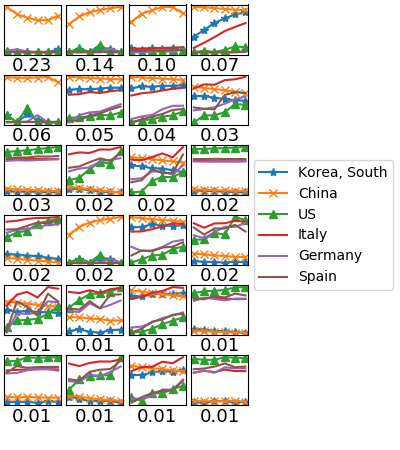
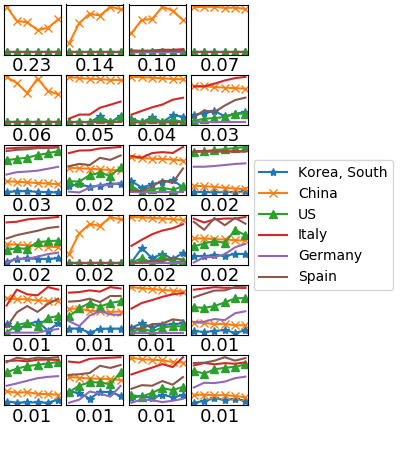
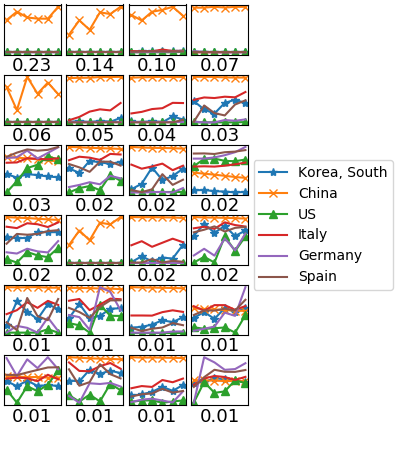
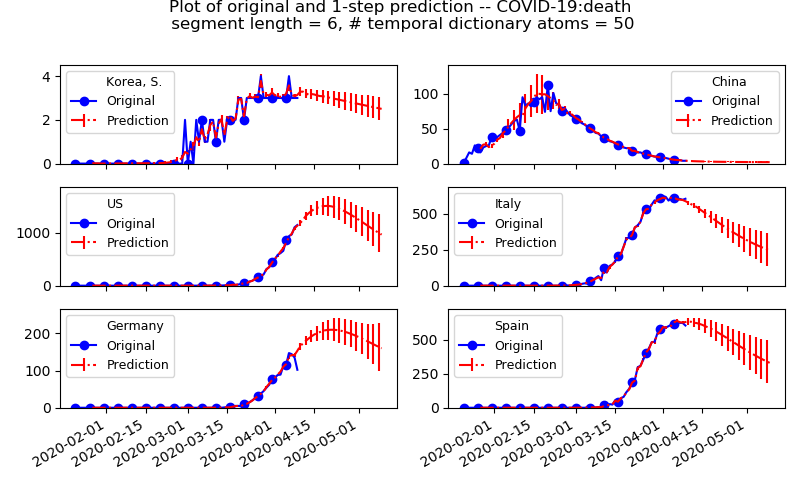
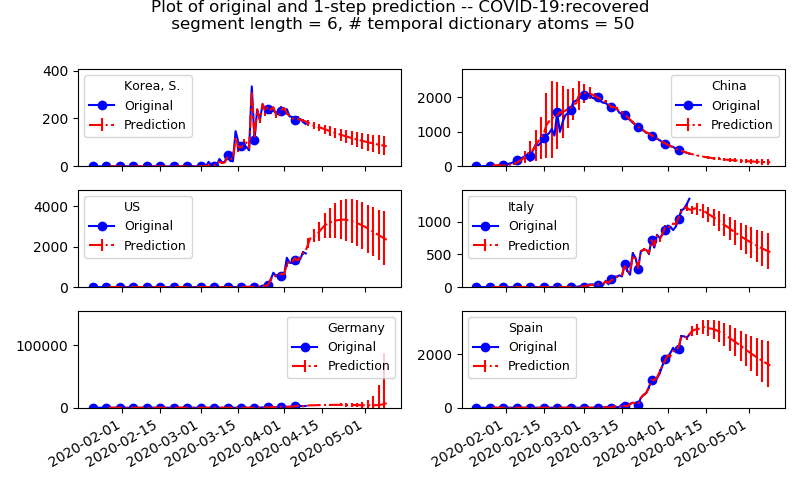


Prediction of COVID-19 daily new confirmed cases

Joint dictionary of 6-day evolution





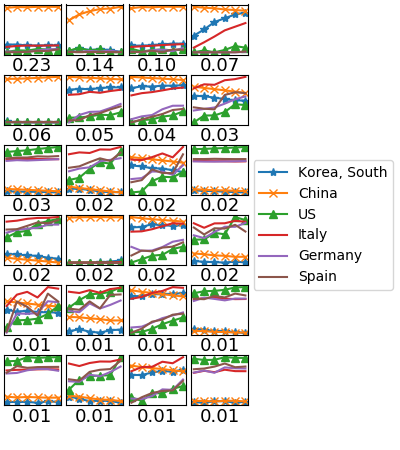
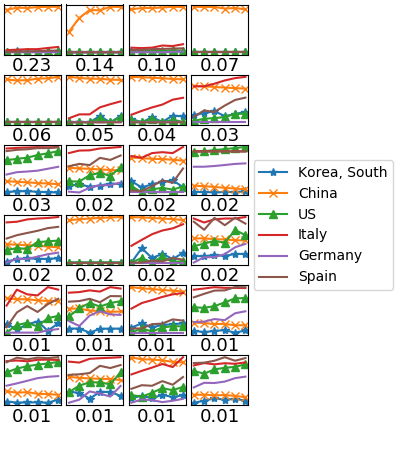
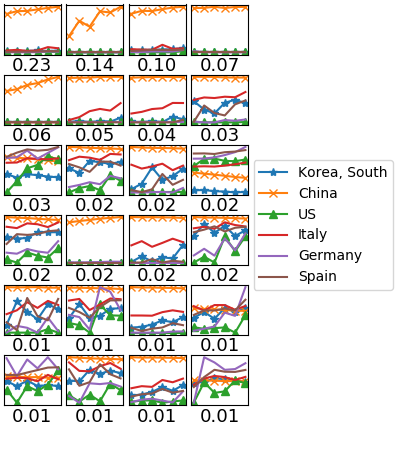
Prediction of COVID-19 daily new recovered cases

Joint dictionary of 6-day evolution

Prediction of COVID-19 daily new deaths

Joint dictionary of 6-day evolution

Joint dictionary of 6-day evolution of three COVID\_19 cases in six countries



Confirmed

Recovered

Death



The goal of our project is to synthesize insights from a diverse collection of COVID-19 related time series data to be able to predict future trends in the number of infections as well as to understand what measures are most effective in fighting the pandemic.

The central concept in our method is dictionary learning, which is a machine learning technique that extracts a reduced number of important patterns (dictionary atoms) in a complex data set. Through our ‘Online Temporal Dictionary Learning’ algorithm based on Online Nonnegative Matrix Factorization, we are able to learn dictionary atoms from an arbitrary set of correlated time-series data (e.g., new daily cases of COVID-19, number of fatal and recovered cases, and degree of observance of social distancing measures). The learned `temporal dictionary atoms’ describe critical short time-evolution patterns in such correlated data set, which gives quantitative measure of correlation and can also be used to predict the future values via our prediction algorithm.

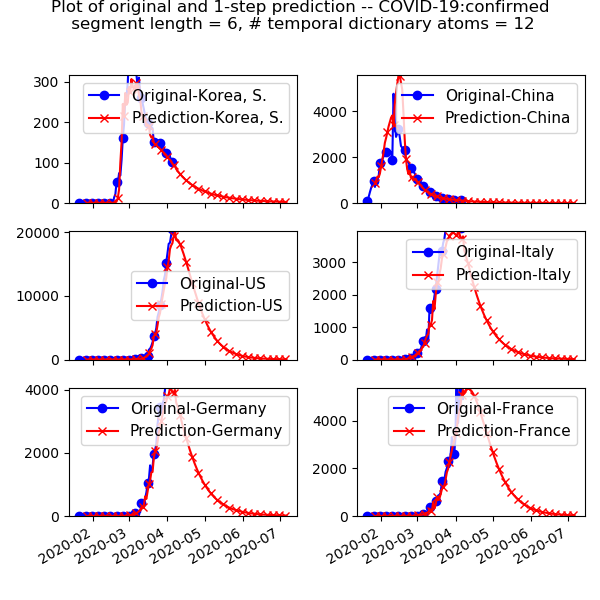
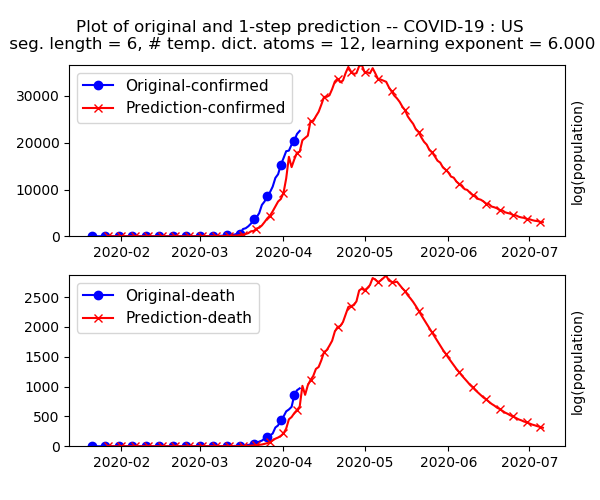
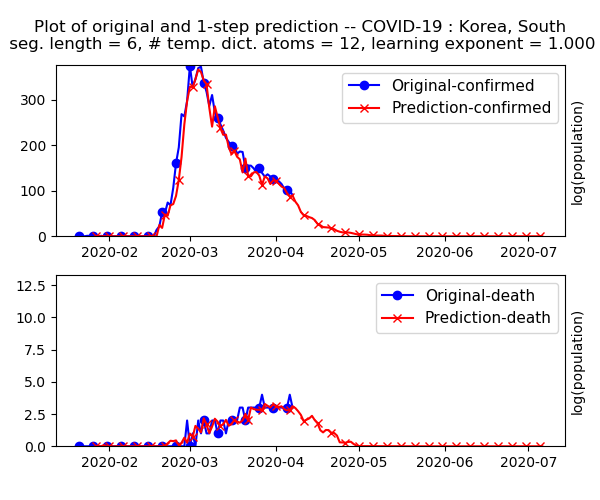
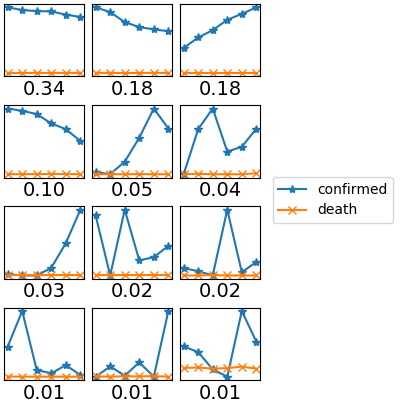
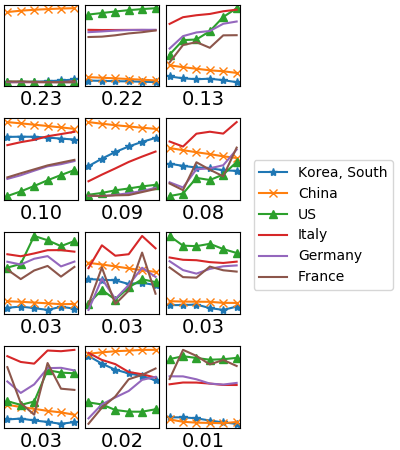
There are some notable advantages in our proposed method. First, our method learns dictionary from a bundle of any correlated time-series data set, so one may augment a given time-series (e.g., new daily cases) with any other ones that may yield some extra predictive power through non-trivial correlation (e.g., measures of social distancing and enforced quarantine, time-series generated by compartmental models, related information spread on social networks). Second, due to the aforementioned flexibility, our method can be used to predict not only the spread of the virus itself, but also many other crucial relevant behaviors such as supply shortages, critical patient numbers, and media coverage and information. Many epidemiological models may not necessarily apply to other relevant COVID-19 related data.

Lastly, our method requires relatively few parameters and learns the model directly from the data. When it comes to predicting the spread of COVID-19, one of the main difficulty for both the mathematical modeling and deep-learning based approaches is that only a very limited amount of daily COVID-19 cases are available (less than hundred days of data), and we are still in the early stage of the pandemic so that most countries are experiencing exponential growth. This critically restricts the amount of information we can get out of the data for fitting the model parameters. In our approach, the dynamical information on time evolution is encoded in dictionary atoms, which we directly learn from a given data set. This allows one to transfer the model learned from more informative data set (e.g., South Korea) for predicting other data sets of more unknowns (e.g., US), which is crucial in our particular application to COVID-19 data set.

The idea of our technique is to consider simultaneously time series quantifying e.g. public awareness of COVID-19, degree of observance of social distancing measures, extent to which masks are being used, and much more.

In contrast to mainstream mathematical modeling approaches, our method learns directly from the data as opposed to making modeling assumptions and trying to tune parameters to fit the data. We comment that the competing approach to time series prediction using deep neural networks is not ideal for the current problem, as such techniques generally require vast amounts of data in order to be effective.

While basic relationships between the spread of the virus in relation to these and other data are common sense (for instance, an increase in social distancing measures should lead to a decrease in number of new infections), the precise relationship is difficult to quantify. Even more difficult, and important, is to understand the complex way in which these different time series interact. It is for such problems that our algorithm shows its strength, as it systematically learns “latent” relationships between different time series. These relationships can yield non-obvious insights about the original time series data which may be useful for informing policy decisions. Furthermore, understanding the latent relationships between our time series data can enhance the predictive power of the model.

-

Prediction of COVID-19 daily new cases

Joint dictionary leaned from six countries

Prediction of COVID-19 daily new cases – S. Korea

using dict. learned from S. Korea

Dict

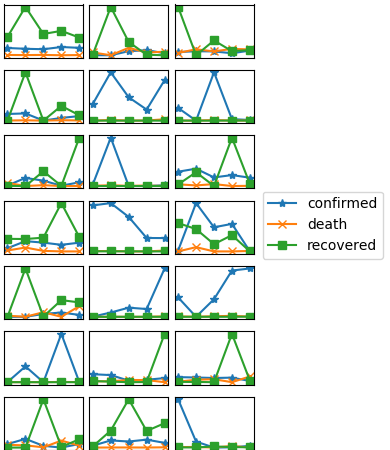
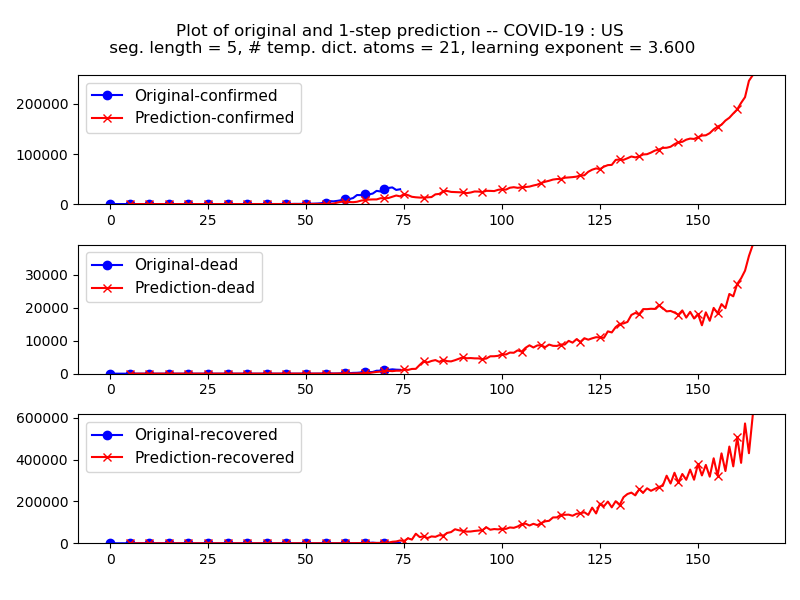
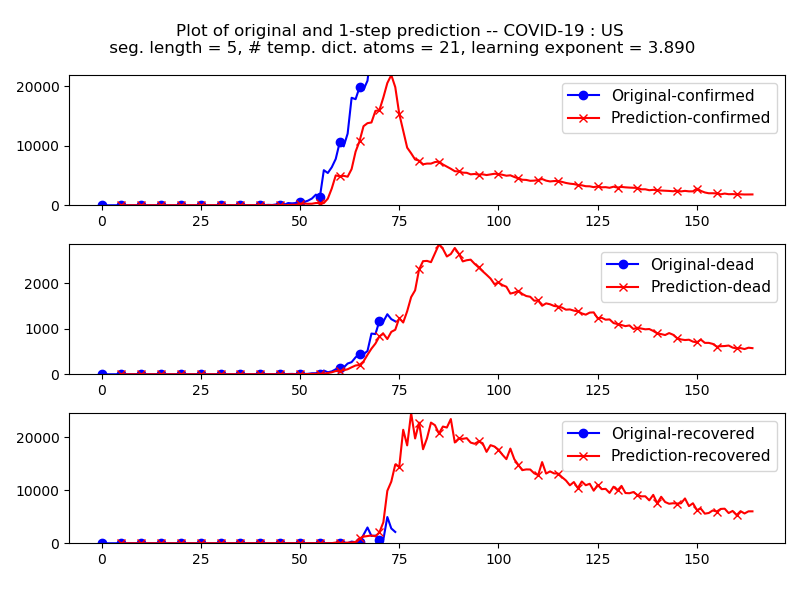
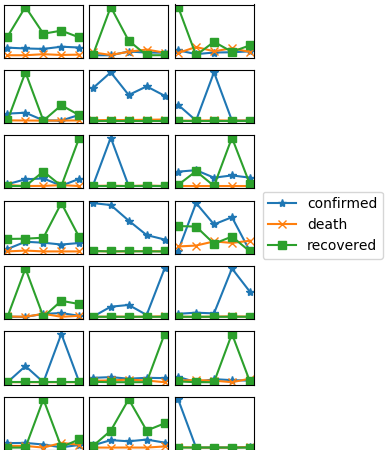
Prediction of COVID-19 daily new cases – US

Dictionary learned from S. Korea

Learned temporal dictionary atoms

Predicting COVID-19 daily new cases in US

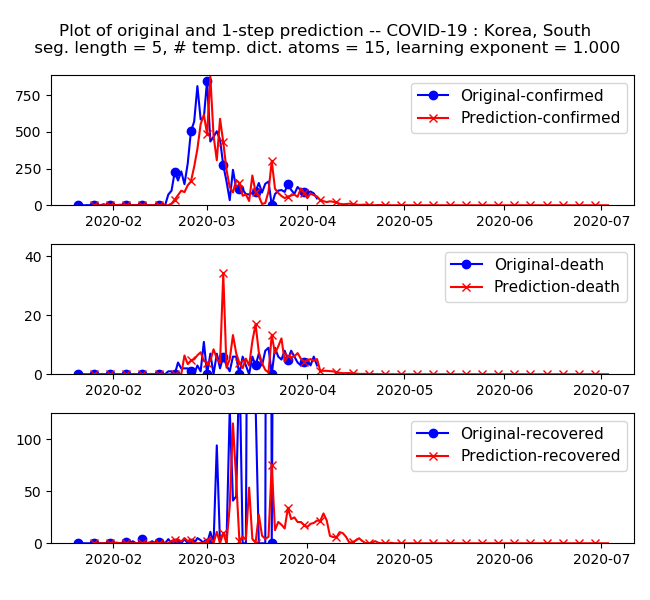
using dict. from S. Korea with



Learned temporal dictionary atoms

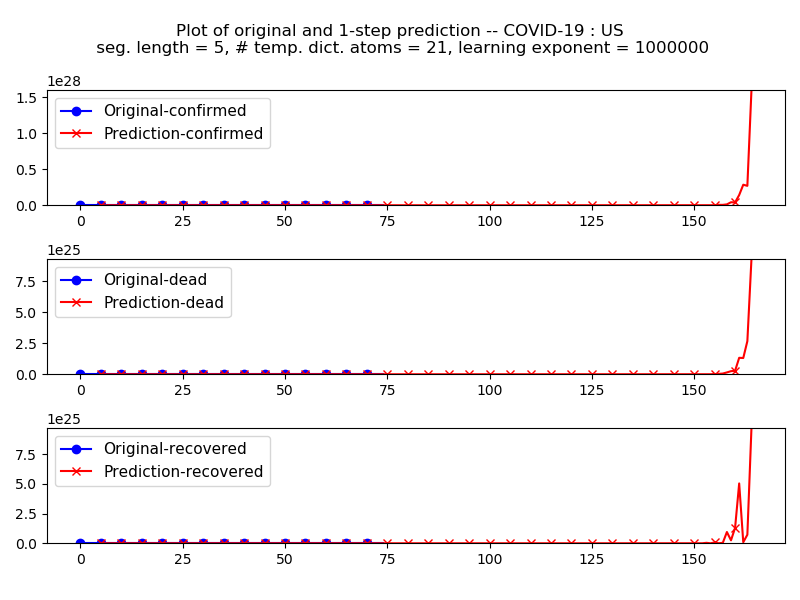
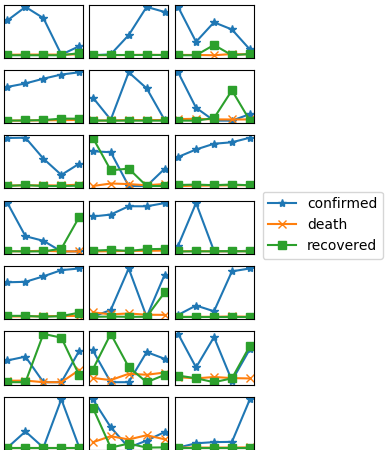
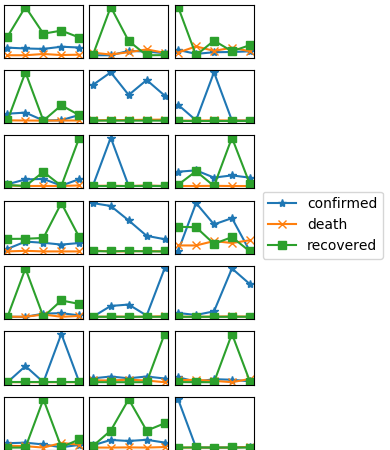
Predicting COVID-19 daily new cases in US

using dict. from S. Korea with



Learned temporal dictionary atoms

Predicting COVID-19 daily new cases in S. Korea,



Learned temporal dictionary atoms

Predicting COVID-19 daily new cases in US,

using dict. from S. Korea with

Learned temporal dictionary atoms

Predicting COVID-19 daily new cases in US