Episimmer Validation on IITJ Data

Experiment

Aim : To train the simulator to predict the number of disease positive agents in a specific environment.

Process : On building a campus with available data, episimmer will fit the disease spread based on the testing and result timeline provided. Follow up by comparing with other simulators.

Note that Episimme is highly flexible and can accommodate any amount of data. The validation conducted does not use Episimmer's complete capability due to the lack of data.

Modelling

Agents
Episimmer works on creating virtual agents and modelling them. Based on the hostel occupancy data we have created a set of agents to adhere to the given conditions.

Duration
Simulation training in the months of March and April 2021, Testing during the month of May 2021.

Environment
The data is highly incomplete. Thus some basic assumptions have been made. From the hostel data, hostels and corresponding populations have been reconstructed. Due to a lack of non-student data, it has been done in a two pronged approach. First we model purely the interactions that occur in the hostel (Based on IITJ data) followed by modelling the entire campus with some additional assumptions.
If additional data on room data, mess data, faculty residence data, extra curricular activities can be provided, a more rigorous analysis can be conducted. Despite receiving some faculty data, it is purely the size of the faulty rooms and nothing about the faculty has been provided. Additionally due to a lockdown on the campus in the given duration, the class data has proven to be redundant.

Testing Policy

Given the testing timeline and resultant positives, Episimmer models this strategy and tries to fit a disease model. Since no specifications were given, the testing was assumed to be random testing. On testing positive, the agent is quarantined for a duration of 10 days (changeable). While testing, we test only those who are present on campus and not quarantined. This testing can be further modified given additional details (group testing, pool testing, testing contacts, testing symptomatic).

Disease Model

Given the available data, there are only a few models that can be considered. Despite Covid-19 having a large dichotomy between Symptomatic and Asymptomatic cases, due to the lack of distinguishable data it is not considered relevant. But with any specific or meta data it is possible. We consider two basic Models

- Stochastic model
  This is a standard disease model mimicking ODE's by having a probability of transition between states.

- Scheduled model
  This is a more stable model that is more realistic. It accounts for the incubation period, recovery time and associated distributions.

  The data for this model has been fit using data from multiple verified sources.
  References : Nature paper, India states covid rates data, Jodhpur district data

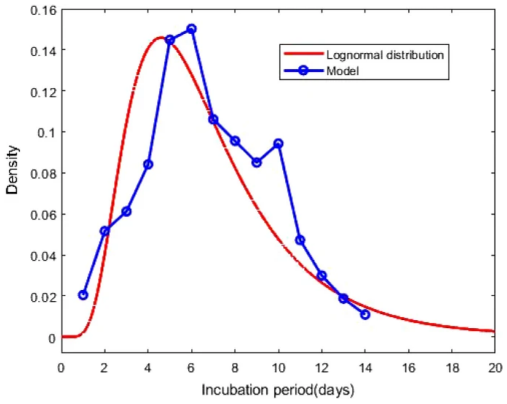
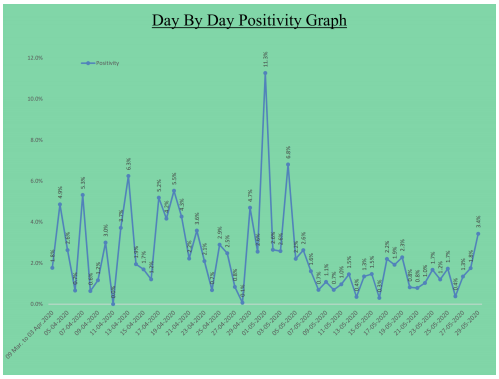
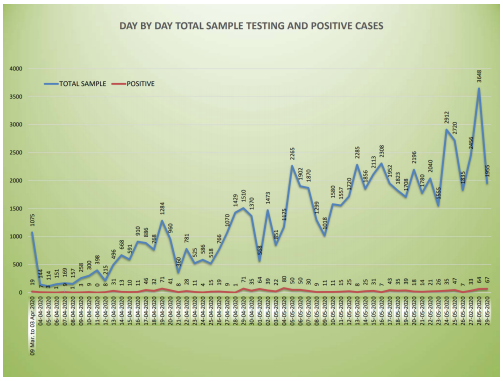


Table 1. Average recovery duration in days of COVID-19 patients in Indian states.

Region	State	R Square	F Statistic	B	b_1	Average Duration (Days)	Lower Limit	Upper limit
Central	Chhattisgarh	0.941	4184.161	2.511	0.969	12.319	10.406	14.583
	Uttar Pradesh	0.883	2101.031	3.222	0.896	25.073	19.353	32.484
	Madhya Pradesh	0.761	835.595	4.225	0.726	68.360	50.348	92.814
East	Odisha	0.956	5823.254	2.562	0.940	12.958	11.168	15.035
	West Bengal	0.948	4789.756	2.441	0.980	11.482	9.497	13.881
	Bihar	0.923	3143.014	2.808	0.914	16.569	13.637	20.132
	Jharkhand	0.871	1695.179	3.204	0.867	24.631	19.806	30.632
North	Rajasthan	0.937	4179.219	2.650	0.952	14.157	11.801	16.984

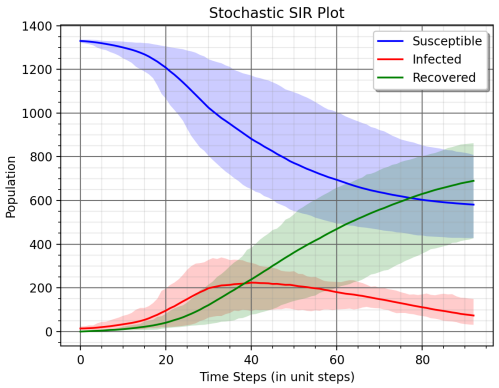




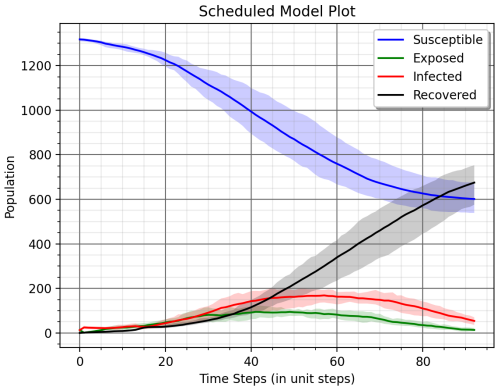
This data allows us to tune incubation period, recovery rate, external prevalence and associated values for the Jodhpur district in which IITJ is situated.

Results

Predicted Disease Spread using different models



Intensity = Hostel:Campus :: 50:1

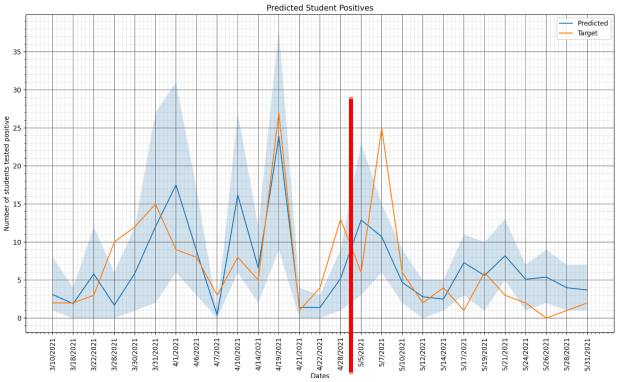


Intensity = Hostel:Campus :: 30:1

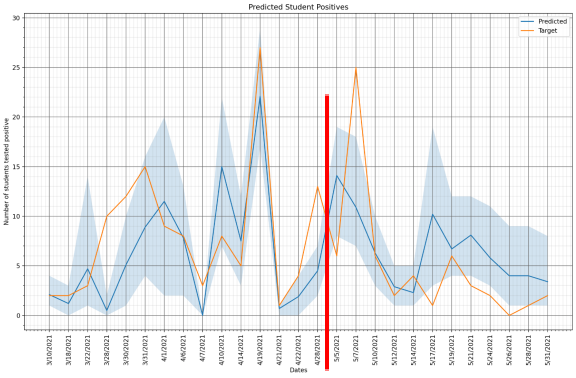
The plots above show the averaged time series of the predicted disease spread. As shown, we consider both Stochastic (SIR) and Scheduled (SEIR) Models for our analysis. The bands show the region of variance for possible disease spread.

Fitting the model

A low intensive grid search was carried out till April 30th. The plots below show the predicted number of positive agents using the testing schema provided by IITJ. We see the predicted values for May after the red line (Validation period). We have not overfitted the data due to the non-deterministic nature of disease spread in general.



Stochastic SIR model



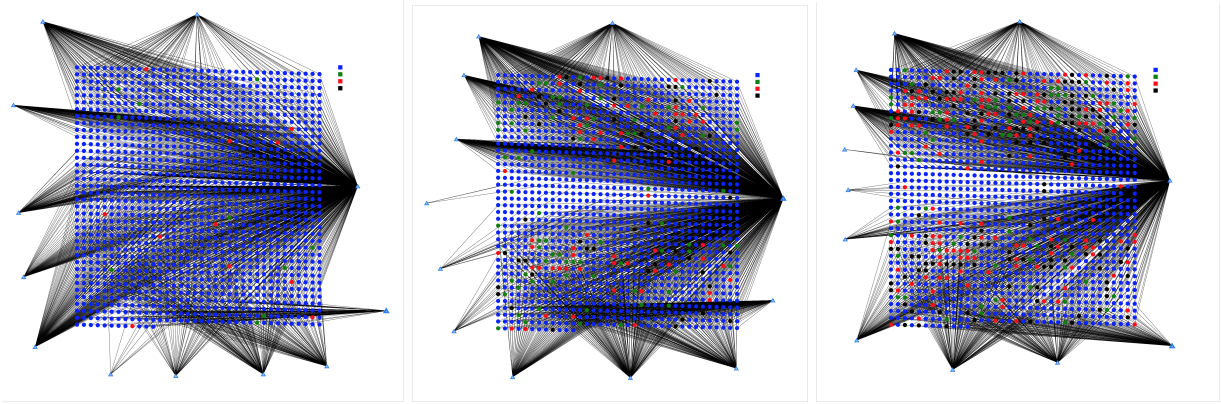
Scheduled SEIR model

The same predicted outcomes have been shown in the table below. The mean square errors have also been shown in the following table.

Data			Predicted Student Positives	
Date of Testing	**Number of Students Tested**	**No of Student Positive**	**SIR Stochastic**	**SEIR Scheduled**
3/10/2021	170	2	**3.1**	**2.1**
3/18/2021	50	2	**1.9**	**1.2**
3/22/2021	111	3	**5.8**	**4.7**
3/28/2021	20	10	**1.7**	**0.5**
3/30/2021	68	12	**5.9**	**5.1**
3/31/2021	143	15	**12.0**	**8.9**
4/1/2021	148	9	**17.5**	**11.5**
4/6/2021	84	8	**8.8**	**7.8**
4/7/2021	2	3	**0.4**	**0.0**
4/10/2021	123	8	**16.2**	**15.0**
4/14/2021	54	5	**6.6**	**7.5**
4/19/2021	146	27	**23.9**	**22.1**
4/21/2021	6	1	**1.4**	**0.7**
4/22/2021	12	4	**1.4**	**1.9**
4/28/2021	26	13	**5.2**	**4.5**
5/5/2021	83	6	**12.9**	**14.1**
5/7/2021	63	25	**10.7**	**10.9**
5/10/2021	40	6	**4.7**	**6.3**
5/12/2021	17	2	**2.8**	**2.9**
5/14/2021	22	4	**2.5**	**2.3**
5/17/2021	64	1	**7.3**	**10.2**
5/19/2021	49	6	**5.6**	**6.7**
5/21/2021	68	3	**8.2**	**8.1**
5/24/2021	53	2	**5.1**	**5.8**
5/26/2021	45	-	**5.4**	**4.0**
5/28/2021	38	1	**4.0**	**4.0**
5/31/2021	45	2	**3.7**	**3.4**
Total	1750	180	**184.7**	**172.2**

	Mean Square Error	
Model	**Train (Till April 31st)**	**Validation (From May 1st)**
Stochastic SIR	23.398	31.185
Scheduled SEIR	23.327	35.062

Visual snapshots of the Scheduled SEIR dynamic graph of student-hostel interaction
[Blue - Susceptible, Green - Exposed, Red - Infected, Black - Recovered]



3/1/2021 **4/1/2021** **4/16/2021**

Conclusion

Despite a lack of complete data we see that Episimmer can fit the data with significant precision (Given by the MSE table). Thus with more data Epismmer can provide for a more rigorous and complete analysis. Since it is not in the scope of the validation experiment and due to ambiguity of some data the following tasks have not been carried out.

- Vulnerability Detection
  - Detect agents most at risk : requires anonymised agents data
  - Detect agents who pose the largest risk of spreading disease: requires anonymised agents data
- Optimization/Recommendation
  - Recommendation on optimal testing procedure : Requires anonymised agents data
  - Optimal schedule of classes to reduce risk : requires student-class relationship
  - When to lockdown and what to lockdown
  - Compare different policies