SSM
Scuola Superiore Meridionale

MERC PhD Project Proposal

Title of the research project:

Data Driven Modelling, Numerical Analysis and Control of Epidemic Dynamics using Machine Learning

Keywords (up to five
Epidemiology, Machine Learning, Complex networks, PDEs, SDEs, Numerical analysis, Control

Supervisors (at least two from two different areas):

Supervisor 1 (Constantinos Siettos, MERC PhD Board Member, MERC PhD Board Member, Agent-Based simulations, Numerical Analysis, Machine Learning.)

Supervisor 2 (Mario di Bernardo, MERC PhD Board Member, Control Theory, Dynamical Systems, Complex networks)

Supervisor 3 (Professor Athanasios Yannacopoulos, Department of Statistics, AUEB, Greece*)*

Supervisor 4 (Dr. Lucia Russo, CNR-STEMS, Italy)

Project description (max 5000 characters)

The COVID19 pandemic is impacting millions across the globe, with huge negative consequences in society and economy. A crucial open problem is what type of interventions should be applied to efficiently combat pandemics, avoiding waves of newly emergent and re-emergent infectious diseases while minimizing the negative impacts. Efficient interventions require a profound understanding of the transmission network across different regions of a country. Here, we propose an ambitious interdisciplinary research project that aims at developing an integrated mathematically-based data-driven modeling and multiscale analysis approach based on machine learning to support decision- and policy makers to assess and control pandemic dynamics and more generally dynamics of new-emergent diseases.

State of the art

The analysis of the multiscale interplay between diverse factors including the network of contact transmission as shaped by socio-economical and demographical conditions, the structure, organization and availability of health services across different regions and the efficient distribution of vaccines is essential to efficiently design efficient policies for combating the spread of the pandemics [1]. To address this issue it is fundamental to develop a strongly interdisciplinary approach to derive detailed mathematical models that are able to capture the complex multiscale nature of the problem, calibrate them based on publicly available epidemiological, clinical and mobility data and finally use them to develop an integrated decision making system for the development of optimal local and national intervention strategies as well as to design early warning systems that are crucial in identifying probable rapid outbreaks that require drastic interventions (for a review on mathematical modelling of infectious disease dynamics see e.g. [2]). Such models must be parameterized on real data and defined at the right scales. Current approaches include aggregate compartmental models such as the SEIRD models. Such models treat the population as being homogeneous and therefore may fail to capture important regional differences that are crucial for policy-makers [3]. At a different scale, agent-based models can incorporate details such as features of the transportation network and fluxes among different areas, demographics as well as epidemiological aspects (see e.g. [4,5,6]).

A crucial open problem is to develop models that can describe at the right level the contact transmission network through which the disease spreads and at a higher scale how interconnections and mobility of people between regions can affect the spread across a country (see e.g. [7, 8]). These network models can be used to devise intervention strategies both in terms of vaccination strategies and intermittent social distancing rules. They are constructed on the basis of demographics, mobility data and social mixing patterns between workplaces, schools, leisure, transportation, home and other activities. While, social mixing matrices at a country level are already available from past studies (e.g. the POLYMOD repository), they (a) do not deal with heterogeneities at a regional level and (b) have been constructed using data from past years. Modelling approaches that take into account regional heterogeneities and mobility have been proposed only recently [e.g. 9] but the lack of data and recent developments require an extension of these models and their complete restructuring and re-parameterization of the model from new data.

Objectives

The main objectives of the project are:

1. Use of machine learning to estimate and analyse the contemporary contact transmission causality network for COVID19 and in general for future outbreaks transmitted by close-contacts, at a regional level, and their expansion to the wider country-level network.
2. Use physics informed machine learning to construct and calibrate regional epidemic modes in the form of ODEs, PDEs and SDEs and their closures based on the derived social mixing networks and epidemiological data, using multiscale analysis approaches.
3. Combine the regional models to derive and parameterize an agent-based network model of country-level model.
4. Development of mathematical methodologies to develop reliable early-warning systems that are crucial for the efficient control of the pandemic and based on such signals systematically design intervention policies, that will help public health authorities to intervene in a more efficient and systematic way both at a regional and at a national level avoiding the huge socio-economic costs of a nation-wide lockdown and avoid future recurrent waves of the diseases.

Methodology

We envisage the combined use of physics informed machine learning and numerical analysis methods for modeling and parameterizing large scale complex multiagent systems to construct regional models in the form of ODEs, PDEs and SDEs of the epidemic spread and bridge the micro (agent-based simulations) to the emergent (macro) dynamics. Equation-free framework [10,4,11] will be used to bridge scales between the detailed agent-based simulations and macroscopic dynamics and will be coupled with physics informed machine learning [12, 13] to calibrate and learn the macroscopic models and their closures. Tools from dynamic optimization and feedback control theory will be used to devise strategies to mitigate the epidemic spread and optimize their deployment in order to achieve the highest efficacy while minimizing their overall economic impoact. As part of the possible measures we will consider both NPIs (non pharmaceutical interventions) and vaccination campaign.

Relevance to the MERC PhD Program (max 2000 characters)

Briefly describe how this project fits within the scope of the MERC PhD program describing its interdisciplinary aspects, relevance in application and beneficiaries.

New rigorous mathematical numerical-assisted methodologies and machine learning algorithms that address the bridging between the scales across which the spread of pandemics emerges have the potential to facilitate novel modeling as well as to enhance our understanding, forecasting capability and design of better public health interventions. Such an attempt is extremely challenging and novel and it is at the forefront of current research interest in the area. Thus, the level of innovation of the proposed research is high, both in terms of the anticipated creation of new knowledge as well as in terms of the expected results and their impact on the society. The scientific keystones of the proposed study are strongly interdisciplinary integrating recent advances from different fields of research, namely: (i) mathematical modeling (ii) data analysis and machine learning, (iii) control engineering of complex systems and networks, and (iv) epidemiology. All these necessary competences are fully covered by the three participating institutions and the participants.
We wish to emphasize that our approach can be exploited in the future to study the pandemic in other countries, especially those with a federal structure or to estimate the spread of the diseases across neighboring countries in Europe or elsewhere.

Key references

[1] Anastassopoulou C., Russo, L., Tsakris A., Siettos, C., 2020, Data-Based Analysis, Modelling and Forecasting of the COVID-19 outbreak, PLoS ONE 15(3): e0230405. https://doi.org/10.1371/journal.pone.0230405.

[2] Siettos, C. I., Russo, L., 2013, Mathematical Modeling of Infectious Disease Dynamics, *(Invited Review Article),* *Virulence*, **4** (4), 295-306

[3] Russo, L., Anastassopoulou, C., Tsakris, A., Bifulco, G.N., Campana E.F., Toraldo, G., Siettos, C.I., 2020, Tracing Day-Zero and Forecasting the COVID-19 Outbreak in Lombardy, Italy: A Compartmental Modelling and Numerical Optimization Approach, PLoS ONE, 15(10): e0240649

[4] Siettos, C. I., 2011, Equation-Free Multiscale Computational Analysis of Individual-Based Epidemic Dynamics on Networks, Applied Mathematics and Computation, 218, 324-336.

[5] Siettos, C., Anastassopoulou, C., Russo, L., Grigoras, C., Mylonakis, E., 2015, Modeling the 2014 Ebola Virus Epidemic – Agent-Based Simulations, Temporal Analysis and Future Predictions for Liberia and Sierra Leone. PLoS Currents. 2015 Mar 9. Edition 1. Doi:10.1371/currents.outbreaks.8d5984114855fc425e699e1a18cdc6c9.

[6] Siettos, C., Anastassopoulou, C., Russo, L., Grigoras, C., Mylonakis, E., 2016, Modeling,

Forecasting and Control Policy Assessment for the Ebola Virus Disease (EVD) Epidemic in Sierra Leone Using Small-World Networked Model Simulations, BMJ Open, 6, e008649. Doi:10.1136/bmjopen-2015-008649.

[7] Coraggio, M., Xie, S., Russo, G., di Bernardo, M., 2021, Intermittent non-pharmaceutical strategies to mitigate the COVID-19 epidemic in a network model of Italy via constrained optimization", submitted to *IEEE Conference on Decision and Control* (CDC), (available at https://arxiv.org/abs/2103.16502)

[8] Della Rossa, F., Salzano, D., Di Meglio, A., De Lellis, F., Coraggio, M., Calabrese, C., Guarino, A., Cardona-Rivera, R., De Lellis, P., Liuzza, D., Lo Iudice, F., Russo, R., di Bernardo, M., 2020, A network model of Italy shows that intermittent regional strategies can alleviate the COVID-19 epidemic, *Nature Communications,* 11, 5106.

[9] M. Gatto et al., 2020, Spread and dynamics of the COVID-19 epidemic in Italy: Effects of emergency containment measures. Proc. Natl. Acad. Sci. U.S.A. 117, 10484–10491.

[10] Kevrekidis IG, et al., 2003, Equation-free coarse-grained multiscale computation: enabling microscopic simulators to perform system-level analysis. Comm. Math. Sciences, 1(4):715-762.

[11] Siettos, C. I., Gear, C. W., Kevrekidis, I. G., 2012, An Equation-free Approach to Agent-Based Computation: Bifurcation Analysis and Control of Stationary States, EPL (Europhysics Letters), 99, 48007.

[12] G. E. Karniadakis, I. G. Kevrekidis, L. Lu, P. Perdikaris, S. Wang, and L. Yang, 2021, Physics informed machine learning, Nature Reviews Physics, 3 (2021), pp. 422–440

[13] Lee, S., Kooshkbaghi, M., Spiliotis, K., Siettos, C.I., Kevrekidis, I.G., 2020, Coarse-scale PDEs from Fine-scale Observations via Machine Learning, Chaos, 30, 013141.

Joint supervision arrangements

Describe joint supervision arrangements, e.g. weekly/monthly meetings with one or both supervisors, how will the joint supervision be split etc.

The supervisors will work closely with the PhD student for the successful completion of the PhD thesis. The student will meet with at least one of the supervisors on a weekly basis while there will be other regular meetings with all supervisors, including during the period abroad when such meetings will be held remotely. There will be also collaboration with Prof. Yannis Kevrekidis, Johns Hopkins Universtiy, USA

Location and length of the study period abroad (min 12 months)

The student is planned to spend 12 months at the Laboratory of Prof. Athanasios Yannakopoulos at the Department of Statistics at AUEB, Athens including a short visit at the Laboratory of Prof. Yannis Kevrekidis, Johns Hopkins University, USA.