Addressing Global Challenges with HPC and Big Data Technologies

Presenters
Dennis Hoppe – **Introduction** (HLRS)
Derek Groen – **Pandemic Research** (Brunel University London)

Global Challenges

- Disaster Resilience
- Energy
- Environment
- Food
- Global Health
- Governance
- Learning
- Prosperity
- Shelter
- Security
- Space
- Water

Can HPC support Global Challenges?

- Support in decision making for global challenges
  - Challenges of this magnitude cannot any longer rely on gut feeling
- High-performance computing as a crucial tool to solve these complex problems affecting our society
  - An accurate and precise description of these grand challenges becomes only feasible by means of HPC with support of AI
- Inevitable to bring together an interdisciplinary team of actors through workshops, symposiums, and projects
  - Experts: HPC + GSC (Global System Challenges) + AI

HiDALGO: HPC and Big Data for Global Challenges

- European Project. 13 Partners. 7 Countries.
  - Coordination by Atos (Spain)
  - Technical coordination by HLRS
  - Runtime: 2018 until 11/2021
  - https://hidalgo-project.eu
- Key Objectives
  - Foster dialogue with a multitude of communities
  - Throw a bridge between infrastructures (HPC, HPDA, AI)
  - Integration of AI with simulations on HPC

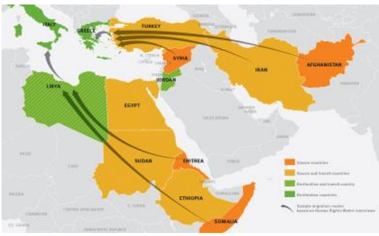
How does HiDALGO address Global Challenges?

- HiDALGO aims to bring together
  - HPC, HPDA, and Global Systems Science (GSS) communities in order to address global challenges and bridge the gap between traditional HPC and data-centric computation
- HiDALGO enables highly-accurate
  - - simulations
    - data analytics, artificial intelligence, and
    - data visualization
- HiDALGO provides mechanisms on how to integrate
  - - various, hybrid workflows
    - data coming from heterogeneous sources

HiDALGO's Ambition

- High-Level Ambition
  - Provide a **single entry-point** for decision makers, technical experts & other relevant entities in the GC ecosystem
  - Connect & **train** the different **communities**
- Project Targets
  - Establish the **baseline for simulations** (ABMS, CFD, …), HPDA and AI oriented computing in the domain of Global Challenges
  - Advance state-of-the-art mechanisms for **data analytics** and develop an **AI-based integrated simulation workflow**
  - Focus on highly accurate models & **significantly improved simulation results**
  - Build-up **coupled simulations** for highly complex phenomena
    Integrate real-world data in static simulations
    Enhance and advertise a **multi-domain portal** for the GC community (https://hidalgo-project.eu)

HiDALGO’s Pilots

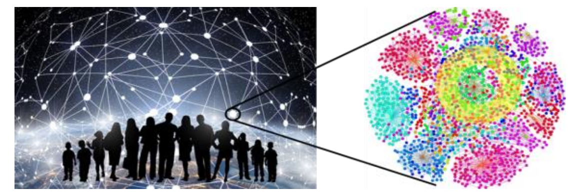


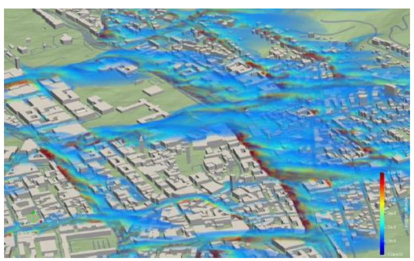
Migrants simulation (Mig)

- Develop realistic models for simulating refugee streams
- Investigate the consequences of certain policies (i.e. a nation closing its borders)

Twitter message simulation (SNA)

- Analyzing the structure of social networks
- Simulating the spread of messages (i.e. Fake news) among users and look for countermeasures



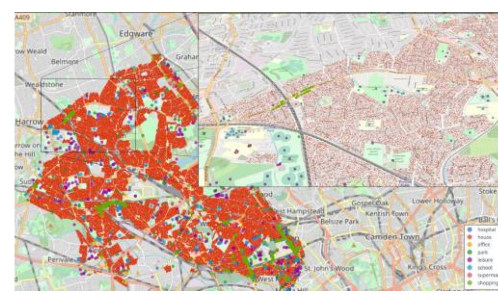


Urban air pollution simulation (UAP)

- Simulate pollution in cities based on real-world sensor data and coupling agent-based simulations with CFD and weather
- Provide stakeholders of politics and industry decision models to leverage green growth

COVID-19 spread simulation

- Determine people infected, ICU occupation, et cetera, through people movement
- Effects of applying certain policies (i.e. curfews)



Pilot Requirements: Coupling and AI

- Coupling mechanisms towards better accuracy
  - Weather and climate data coupling (provided by ECMWF)
    - Implemented REST API for async coupling
    - Completed for Mig (South Sudan) and UAP (City of Gyor) pilots
  - Implemented Kafka-based solution for management of streaming data
  - Devised an approach for the integration of real-world sensor data into the simulation execution
  - Introduced simulations involving coupling with numerous static data sources
  - Coupled simulations from different scientific domains (e.g., CFD+ABMS)
- Introduce AI to GSS workflows
  - Designed 3 AI algorithms to support pilots
  - Introduced generalized GSS workflow

Benchmarking & Scalability towards ExaScale

Pilot	End 2018	Initial Status	Mid 2020	Current Status
Migration	Flee v1.0	Limited parallel efficiency, communication-bound	Flee v2.0	Scalable up to 16K cores for 1B agents
Urban Air Pollution	Fenics-HPC version	Installation/ Procurement issues	OpenFOAM version	Scalable up to 4K cores for mesh of 10M points
	ANSYS Fluent	Not scalable		
Social Networks	Exact version (MPI)	Memory-hungry, not scalable	Approximate version (Python)	Scalability up to 32K cores for the Pokec graph

HiDALGO Case Study
Modelling the Spread of COVID-19

Publication

Imran **Mahmood**, Hamid **Arabnejad**, Diana **Suleimenova**, Isabel **Sassoon**, Alaa **Marshan**, Alan **Serrano-Rico**, Panos **Louvieris**, Anastasia **Anagnostou**, Simon J E **Taylor**, David **Bell** & Derek **Groen** (2020) FACS: a geospatial agent-based simulator for analysing COVID-19 spread and public health measures on local regions, Journal of Simulation, DOI: 10.1080/17477778.2020.1800422

Motivation

- The recent Covid-19 outbreak has had a tremendous impact on the world.

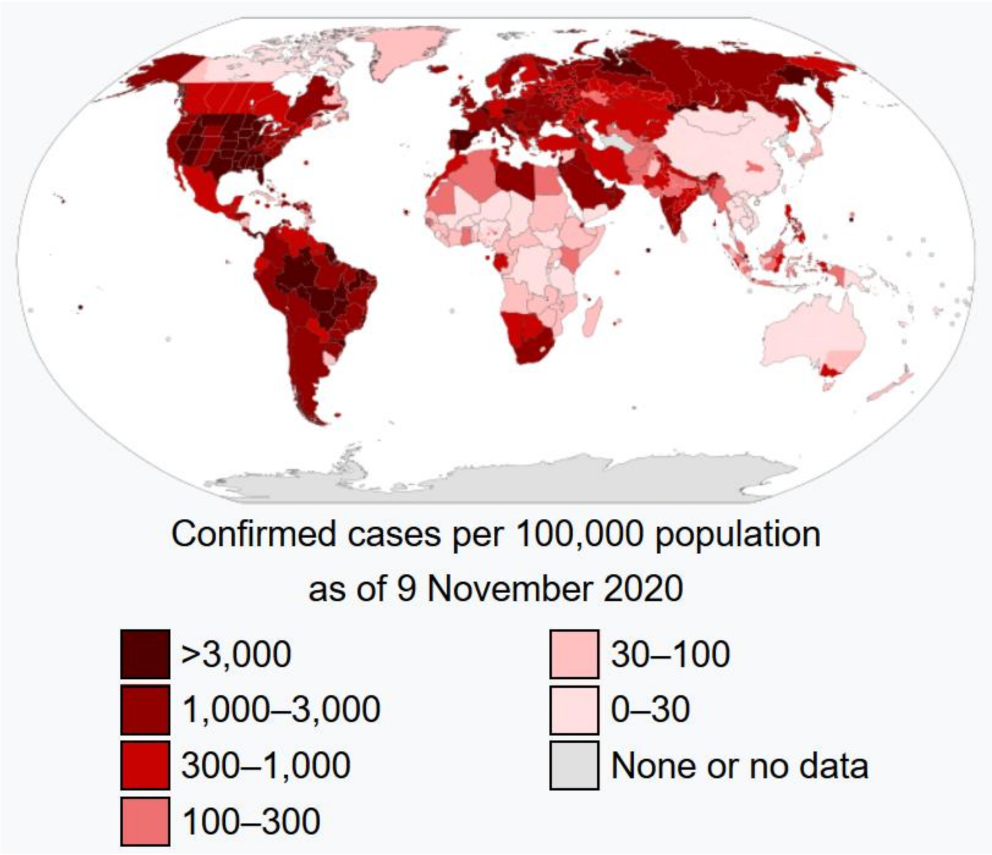
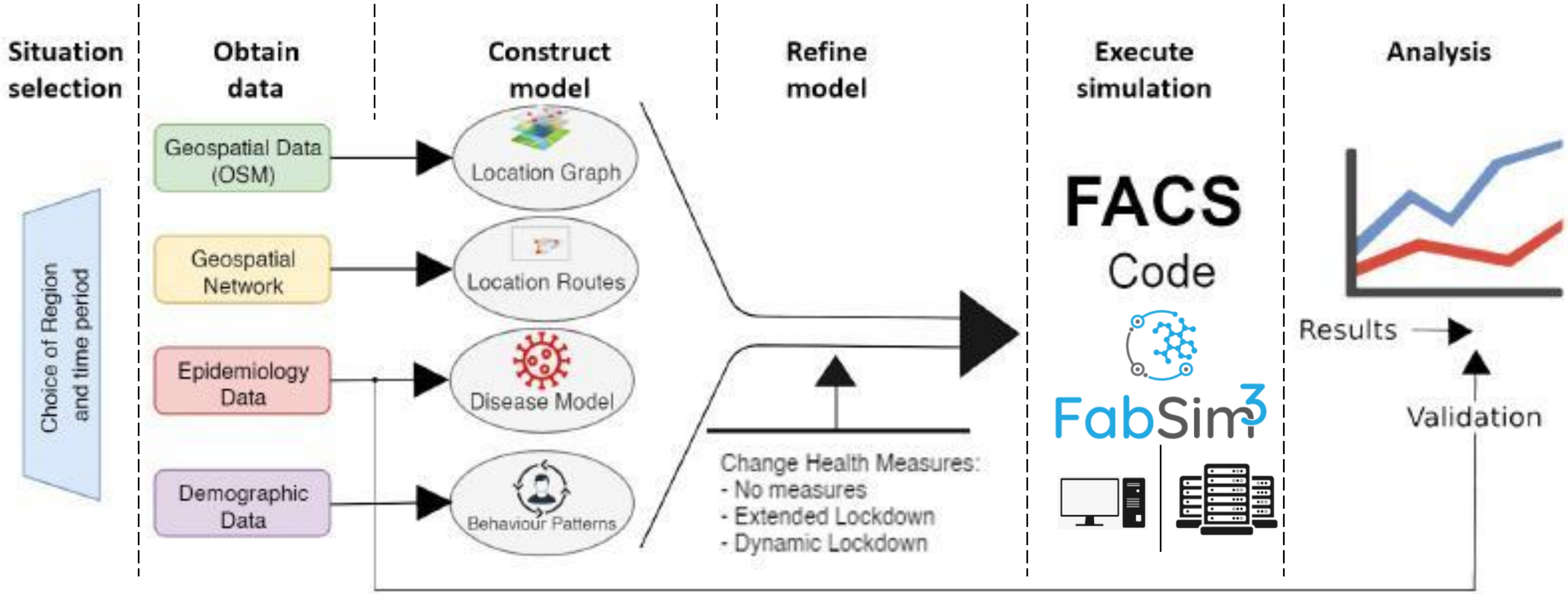


Image source: „Map of the COVID-19 verified number of infected per capita as of November 2020“ by Raphael Dunant is licensed by CC By 4.0.

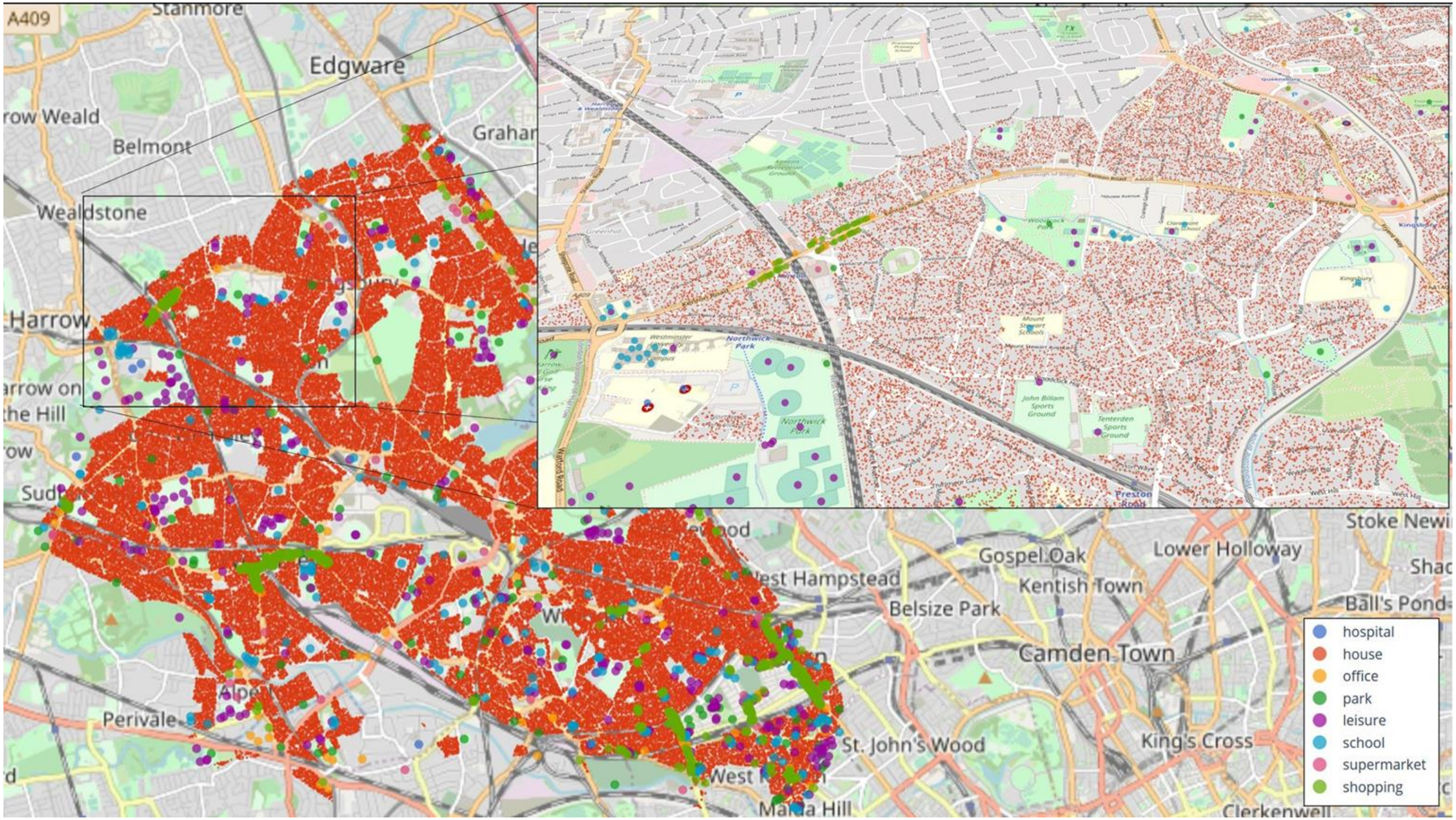
Motivation

- FACS aims to support the emergency response against the COVID-19 outbreak.
  - Rapidly test public health interventions (e.g., as lockdowns) at local levels.
  - Provide reliable and reproducible forecasts of the spread of SARS-CoV2 in the local regions (most existing models are nationwide).
- Modelling at local levels has several important advantages
  - Helps inform location specific public health interventions, which are now increasingly common.
  - Can be used to inform individual healthcare organizations about required capacity and expected patient flows.

Agent-based **Simulation Development Approach**

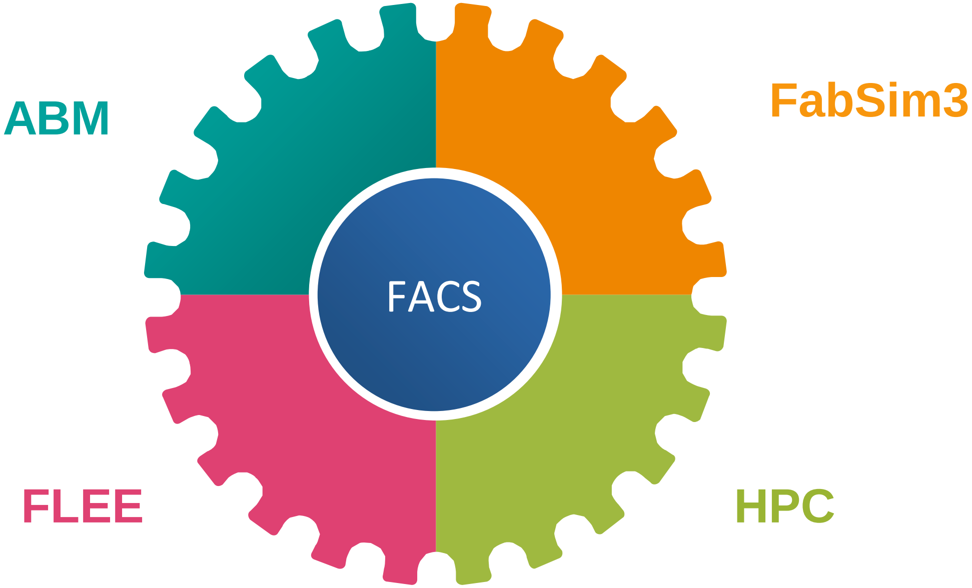


Location Graph



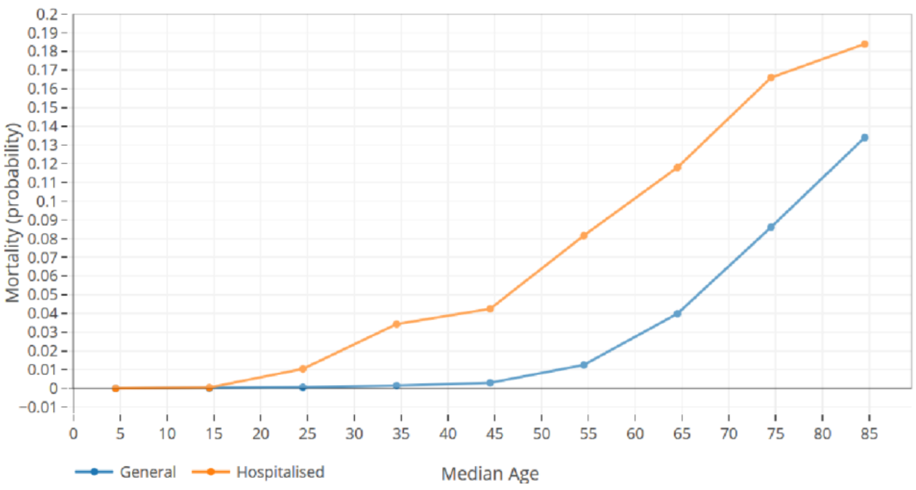
Solution

- FACS inherits features of a comprehensive simulation framework from its ancestors:
  - FLEE: specializes in ABS complex dynamics e.g., agent movements https://flee.readthedocs.io/
  - FabSim3: to simulate a large population of agents using remote supercomputers https://fabsim3.readthedocs.io/



Source: https://github.com/djgroen/facs

Epidemiological Inputs



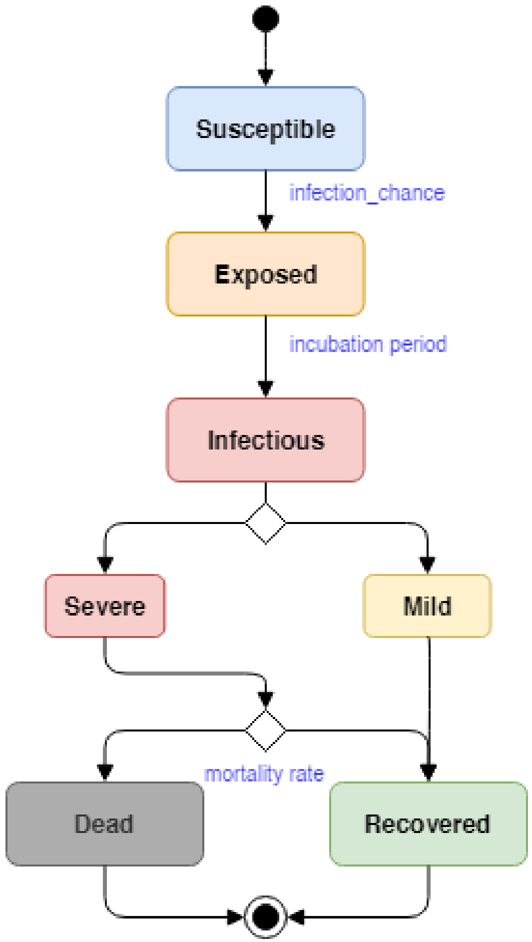
- Parameter/Assumptions

  Infection rate = 0.07
  Incubation Period = 3.5 days
  Illness Duration:
  (i) Mild, no hospitalization, 14 days to recovery
  (ii) Severe, hospitalization, 20 days to recovery (if surviving)
  Time period to hospitalization = 12 days (average)
  Mortality rate = 1.6%
  Mortality per Age = 0 - 0.134
  Period to either mortality or recovery for severe cases: 8 days starting from the day of hospitalization (ICU)
  Period to recovery for mild cases: 8.5 days (average)

In-Building Disease Model

$$\mathbb{P}_{inf} = \left(\frac{LSs}{OD_{loc}} \times \frac{LSi}{A_{con}}\right) \times \frac{IR}{360} \times M$$

- **P_{inf}** is the probability of the susceptible person become infected
- **LS_s** is the length of stay of the susceptible person (in minutes)
- **LS_i** is the length of stay of all infectious persons together (in minutes)
- **OD_{loc}** is the opening duration of the location on that given day
- **A_{con}** is area of the location (in m2)
  IR is the infection rate (i.e., 0.07)
  M is a static contact rate multiplier (1.0 if no interventions)

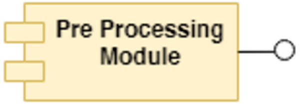


FACS Code Overview

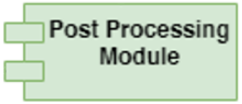
- Main functionalities:

  Provide a disease modelling suite to non-programmers.
  Provide an open-ended API for the modellers and programmers to use it for further scientific research and development on disease spread.

  Additional components:



Pre-Processing Module
- Use to prepare data files
- Location Graph
- Disease Model (YAML)
- Demography data



Post-Processing Module
- Visualization
- Validation

FabCovid19
- FabSim plugin
- Ensemble runs
- Replication runs

FACS Overview

- Main execution platforms
  - Manual local execution (e.g. by script)
  - Using FabSim3 (FabCovid19 plugin) on a supercomputer.
    - Optimized for large ensemble simulations, to investigate different scenarios and account for stochasticity
    - Also allows users and developers to perform sensitivity analysis
  - Using Cloudify and CKAN (to be demonstrated later)

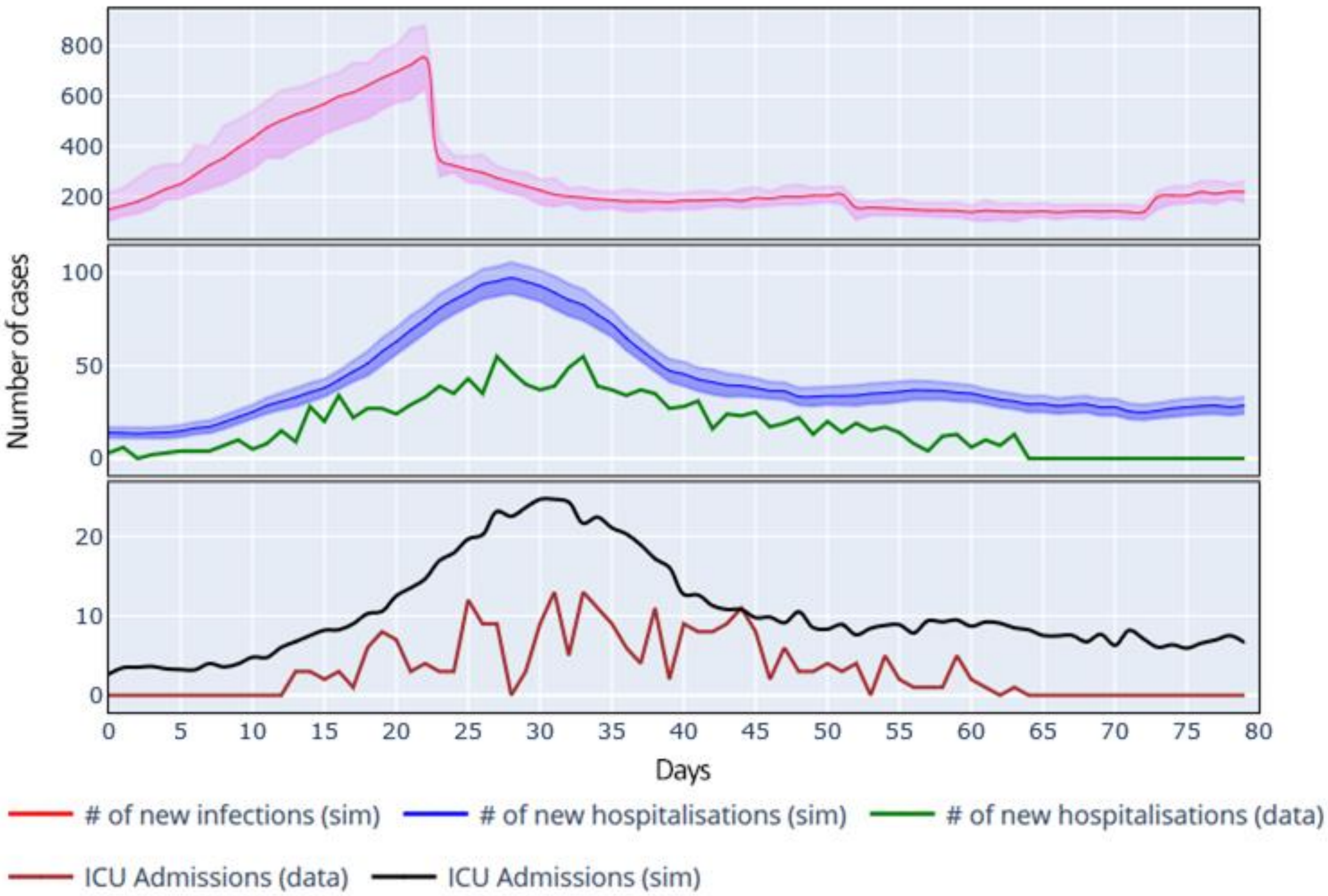
- Typical execution time
  - 1 to 3 hours (depending on the platform)
- Typical ensemble size
  - `#scenarios x #locations x 25`
- Code is sequential, but parallelization efforts are underway!

Supported locations and interventions

- London Boroughs
  - Brent
  - Ealing
  - Harrow
  - Hillingdon
  - Kensington & Chelsea
  - Hammersmith & Fulham
  - Westminster
  - Camden
- Madrid city-scale
- Islamabad city-scale

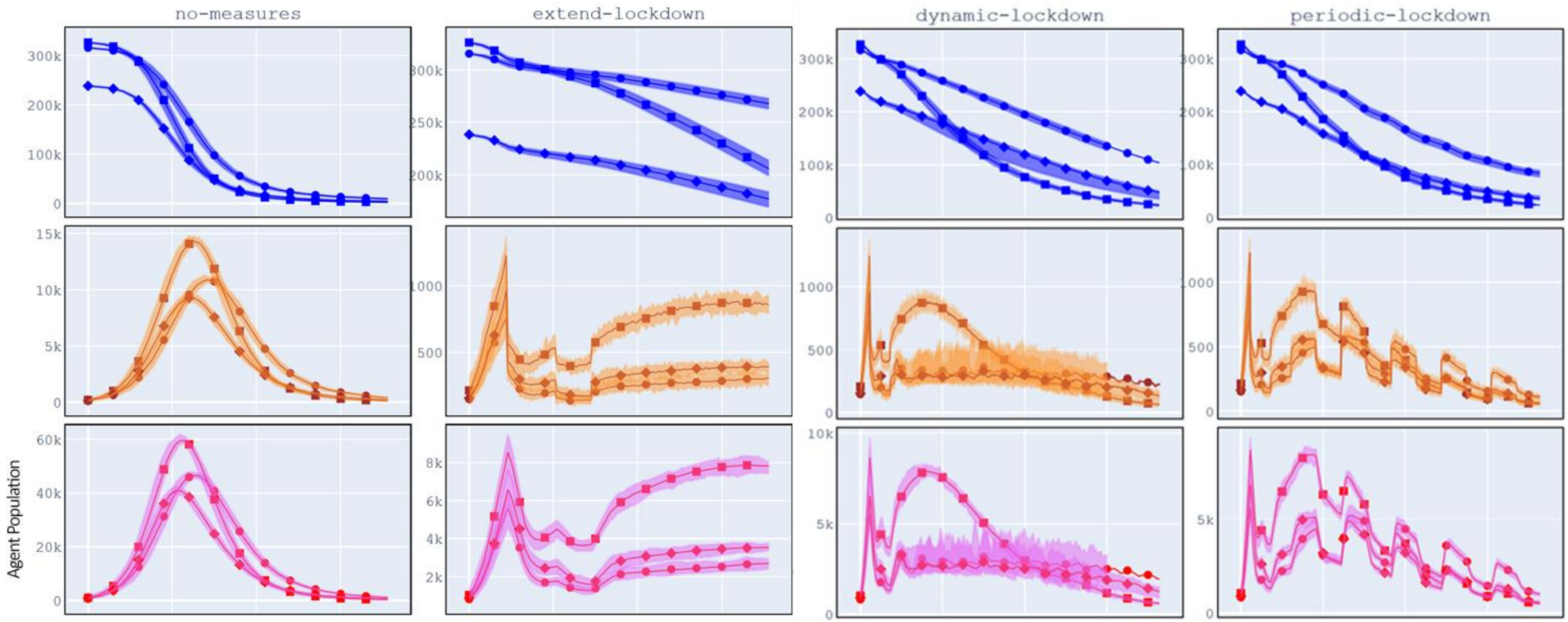
- Public Health Interventions Supported
  - Default (none).
  - Case isolation, household quarantine.
  - Social distancing (1-2 meters).
  - Schools, shop, amenity closures.
  - Work from home directive.
  - Mask wearing directives (voluntary or enforced).
  - Reduced transportation use.
  - Test and tracing.
  - Rudimentary vaccine mechanisms.
  - Shielding of elderly people.
- Under Development
  - Airflow-dependent transmission model within confined spaces.
  - Street-level interactions and interventions.

Model Validation



Simulation Results

Susceptible, exposed and infectious populations over time.



Future Plans

- Apply simulation to other countries/cities
- Parallelize the code.
  - Initialization takes non-negligible amount of time, so that is a first target for parallelization.
  - Resolving location visits is also expensive, so we will most likely go for a similar approach as used in Flee (equally distributed locations + distributed agents).
- Integrate with other simulators for hybrid modeling (e.g., with discrete event hospital models)
- Work on improving the quality of forecasting through sensitivity driven simulation development and code updates.
- Validate the code more widely.

HiDALGO

https://hidalgo-project.eu

THANK YOU !

Acknowledgement: This work has been supported by the HiDALGO project and has been partly funded by the European Commission's ICT activity of the H2020 Programme under grant agreement number: 824115. This paper expresses the opinions of the authors and not necessarily those of the European Commission. The European Commission is not liable for any use that may be made of the information contained in this paper.

Dennis Hoppe
HLRS
Nobelstraße 19
70569 Stuttgart, Germany
Email: hoppe@hlrs.de
Web: https://hlrs.de

Derek Groen
Brunel University London
Wilfred Brown Building 210
UB8 Uxbridge, London, United Kingdom
Email: derek.groen@brunel.ac.uk
Web: https://www.brunel.ac.uk

Coordinator of HiDALGO
F. Javier Nieto
ATOS Spain
Capuchinos de Basurto, 6
48013 Bilbao, Spain
Email: francisco.nieto@atos.net
Web: http://www.atosresearch.eu