Scottish COVID-19 Response Consortium

Monday, 27th July 2020
Scottish COVID-19 Response Consortium

Monday, 27th July 2020
Network Simulation

Jess Enright (Glasgow)
Bob Turner (Sheffield)
Rafael de Almeida (Man Group)
Chris Hughes (Man Group)
Fedor Gorokhovik (Man Group)

Language  Python
Spatial scale  Data dependent – currently Health Board or IZ
Model type  Spatially-embedded network of age-structured populations, deterministic or stochastic, compartmental model
Features  - Movement flows from data or simulated
- Age-class mixing from input matrix
- Local and long-distance spread
- Dynamic movement and contact damping
Mode  Forward simulation, daily timesteps
Basic inference implemented – ABC-SMC

Key questions
- Control related to age class or movements
- Importance of network flows, impact of changes to flows

https://github.com/ScottishCovidResponse/simple_network_sim

Will Pettersen (Glasgow)
Anthony O'Hare (Stirling)
Leighton Pritchard (Strathclyde)
Alex Konovalov (St Andrews)
Jason Dykes (City)
Jo Wood (City)
## Network Simulation

### Significant achievements over the RAMP period

<table>
<thead>
<tr>
<th>Software engineering</th>
<th>Many major improvements in design, automated testing, documentation, reproducibility, and more!</th>
</tr>
</thead>
</table>
| Modelling functionality    | - Choice to focus on network-of-populations  
- Generalisation to arbitrary regions given data  
- Generalised compartmental model  
- Time-varying activity damping  
- Stochastic model operation  
- Generating realistic behaviour  
- Enormously improved software choices! |
| Data Interface             | - Move to inputs from files instead of hard-coded or simulated  
- Using real data as input  
- Data Pipeline API integration |
| Inference and Sensitivity  | - Linear and non-linear sensitivity analysis  
- Implementation of ABC-SMC for basic inference |
## Contact Tracing

### Key questions

What are the optimal contact tracing protocols?
- Benefits of blanket vs tailored actions and advice for alertees?
- How can we reduce the burden of contact tracing whilst maintaining high levels of efficacy
  - Targeted alerts / rapid testing / release of contacts from isolation
- How much more effective does the tracing become if tracing of secondary contacts is allowed?
- What is the impact of basing alerting on a positive test rather than symptoms?

<table>
<thead>
<tr>
<th>Language</th>
<th>Java</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spatial scale</td>
<td>No explicit spatial scale but a form of ‘location’ is retained</td>
</tr>
<tr>
<td>Model type</td>
<td>Individual-based stochastic network model</td>
</tr>
<tr>
<td></td>
<td>- Age-structured individuals</td>
</tr>
<tr>
<td></td>
<td>- Weighted contacts</td>
</tr>
<tr>
<td>Key Feature</td>
<td>Flexible contact tracing scenarios</td>
</tr>
<tr>
<td>Mode</td>
<td>Forward simulation, user-defined timesteps (quarter of day)</td>
</tr>
</tbody>
</table>

Contact Tracing: Key features

Sibylle Mohr, Louise Matthews (EPIC / Glasgow)
Sam Brett, Vino Mano (Man Group),
John Nonweiler (UKAEA), Ed Townsend

Explore contact-tracing policies

Isolation measures
• Blanket 14-day isolation vs. dependent on the time of contact (0-14 days)

Alerts / testing
• targeting of alerts at individuals most likely to be infected
• testing of individuals to release their contacts from isolation

Different 'alerting' scenarios
• Alerting based on symptomatics
• Alerting based on positive test

Contact tracing
• Direct contacts only
• Also secondary contacts (contacts of direct contacts)
Contact Tracing Team

Major RAMP Achievements

Software development

- Code clean-up; codebase tested / integrated / optimised; re-factoring; added software tests / unit tests...
- Documentation, verification, integration into the SCRC environment...

Model functionality

- Multiple, sophisticated contact-tracing scenarios / secondary contacts / global lockdown policies / testing scenarios / test efficacy
- Model runs on various contact data, together with list of age-structured individuals; Random infection / targeted seeding
- Complex statistics output (R calculations; traced / isolated / infected...)

Inference

- Sensitivity analysis for model parameters / policy parameters
- Comparison between CTM model and ODE model

Visualisation

- Multiple static and dynamic visualisation of different output data
  - Network data / compartmental outputs / infection maps

And so much more!!

- Data API for Java is taking shape; new documentation / inferences / visualisation...

<table>
<thead>
<tr>
<th>RSE core team</th>
<th>Sam Brett</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Vino Mano (Man Group)</td>
</tr>
<tr>
<td></td>
<td>John Nonweiler (UKAEA)</td>
</tr>
<tr>
<td></td>
<td>Edward Townsend</td>
</tr>
<tr>
<td>Inference / model</td>
<td>Rikiya Takahashi</td>
</tr>
<tr>
<td>assessment</td>
<td>Fedor Gorokhovik</td>
</tr>
<tr>
<td></td>
<td>(Man Group)</td>
</tr>
<tr>
<td>Visualisation team</td>
<td>Cagatay Turkay (Warwick)</td>
</tr>
<tr>
<td></td>
<td>Dan Archambault (Swansea)</td>
</tr>
<tr>
<td></td>
<td>Kai Xu (Middlesex)</td>
</tr>
<tr>
<td>Contact tracing</td>
<td>Louise Matthews (UoG)</td>
</tr>
<tr>
<td>options</td>
<td>Jamie Prentice (UoG)</td>
</tr>
</tbody>
</table>

https://github.com/ScottishCovidResponse/Contact-Tracing-Model
Spatial Simulation

Claire Harris (EPIC / BioSS)

James Cook (UKAEA)

https://github.com/ScottishCovidResponse/Simulation.jl

<table>
<thead>
<tr>
<th>Language</th>
<th>Julia</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spatial scale</td>
<td>1km² grid, Scotland / UK</td>
</tr>
<tr>
<td>Model type</td>
<td>Explicitly spatial, stochastic, compartmental model</td>
</tr>
</tbody>
</table>
| Features       | - Airborne and environmental spread of virus  
|                 | - Local spread vs. long-distance commuting  
|                 | - Impact of climate / environment on transmission  
|                 | - Demographic structure, population density  
|                 | - Inference and visualisation frameworks |
| Mode           | Forward simulation, daily timesteps |

Key questions

- What effect does long term pollution have on susceptibility / severity of COVID-19?
- What is the effect of weather/environment on the virus?
- What are the relative contributions of direct/indirect transmission?

RSE team: Alex Robson, Bella Wu, Eric Perim Martins and Sean Lovett (Invenia)

Epidemiology: Glenn Marion and Iain McKendrick (BioSS), Louise Matthews and Richard Reeve (Glasgow)

Inference: Ben Swallow and Simon Babayan (Glasgow)

Visualisation: Aidan Slingsby (London City)

Data: Sonia Mitchell and Ciaran McMonagle (Glasgow), Stephen Catterall (BioSS) and Jeremy Walton (Met Office)
Major developments

• **Codebase:**
  • Integrated, tested, documented, optimised
  • Python wrapper for data pipeline
  • Fully deterministic version of code

• **Epidemiology:**
  • Mostly everything!

• **Data:**
  • Data flowing from the pipeline
  • Demographics/Epi params/Weather/Pollution

• **Inference:**
  • ABC-SMC interface in R that can recover initial parameters

• **Visualisation:**
  • Fully interactive Java visualisation, with multiple modes, scaling and spatiotemporal dynamics
# Individual-based Simulation

## Model purposes
- Detailed simulation of different communities
- Completely flexible reopening/reclosing of businesses and compartments of society
- Comparison between different communities / areas
- Scenarios for opening schools
- Flexible addition of new societal compartments (could add theatres, cinemas etc)
- Shielding populations
- Strategies for roll-out of vaccination or new tests
- Role of superspreaders
- Informing other models with network structure

## Language
Java

## Spatial scale
Defined area (e.g., Lothian HB, Orkney)

## Model type
Individual-based stochastic simulation

## Features
- Individual hourly movements and mixing
- Implicit network structure
- Full age and household structure
- Mixing and transmission in distinct community compartments
- Health boards and care homes

## Mode
Forward simulation, hour time steps

## Fitting
Relies on parameters and visual comparison to data

## Data
Most key parameters are now evidenced, but would be good to complete the list

---


Paul Bessell (Edinburgh), John Nonweiler (UKAEA), Blair Archibald (Glasgow), Bob Taylor
Individual-based Simulation

Major RAMP Achievements

Software development

- Simple terribly written code has been transformed into a very efficient and complete epidemiological model
- Complete testing environment with the model
- 1 run of a 200 day simulation of 910,000 population on DiRaC is ~22 mins
- Full software documentation

Model functionality

- Fully age structured disease progression
- Age structured household structure and implicit contact network
- Fully age-structured non-Covid hospital utilisation
- Disease transmission during travel
- Build-your-own lockdown release scenarios

Fitting

- (For Lothian at least) Model broadly fits the mortality, hospitalisation and R values)
- Will run it through Thibaud’s model

Further steps

- Write up the epidemiological documentation
- Get better contact parameter data
### Model purposes
- Inference of key epidemiological parameters
- Impact of scales (national vs regional vs health board)
- Prediction of community level infection
- Quantification of number of primary infection
- Optimization lifting movement restrictions strategies

### Language
<table>
<thead>
<tr>
<th></th>
<th>C++</th>
</tr>
</thead>
</table>

### Spatial scale
- Flexible

### Model type
- Age-structured compartmental stochastic model

### Features
- Different transmission behaviour between community and HCW
- Impact of bed capacity on hospitalisation rate

### Mode
- Inference – ABC-smc
- Forward simulation

---

**RSE team:** Sanket Gadgil, Kristian Zarębski, Qingfeng Xia (UKAEA), Andy Bennett

**Epidemiology:** Mark Bronsvoort, Paul Bessell (Edinburgh), Louise Matthews (Glasgow)

**Uncertainty quantification:** Ben Swallow (Glasgow), Ian Vernon (Durham), and the rest of #UQ team

**Data:** Stella Mazeri (Edinburgh)
Major RAMP Achievements

Software development
- Simple terribly written code has been transformed into a very efficient and complete epidemiological model
- Complete testing environment with the model
- Full software documentation
- Full epidemiological documentation
- Sensitivity analysis

Model functionality
- Both inference and forward prediction implemented
- Fully age structured disease progression
- Role of HCW in the case reporting

Fitting
- Model broadly fits the mortality, hospitalisation and R values for Scotland
- Inference framework validated

Further steps
- Uncertainty quantification
- Coding and assessing lockdown strategies
- Fit on all health board in UK
**BEEPmbp**

Chris Pooley (BioSS / EPIC)
Glenn Marion (BioSS/ EPIC)

https://github.com/ScottishCovidResponse/CoronaPMCMC

<table>
<thead>
<tr>
<th>Language</th>
<th>C++</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spatial scale</td>
<td>Middle Layer Super Output Areas (MSOA) or output areas (OA), Scotland</td>
</tr>
<tr>
<td>Model type</td>
<td>Individual-based, spatial, stochastic compartmental model</td>
</tr>
<tr>
<td>Features</td>
<td>- Individual risk factors, e.g. age, gender</td>
</tr>
<tr>
<td></td>
<td>- Area risk factors, e.g. population density</td>
</tr>
<tr>
<td></td>
<td>- Time variation in transmission</td>
</tr>
<tr>
<td></td>
<td>- Calculation of model evidence (ME)</td>
</tr>
<tr>
<td></td>
<td>- Optimized parallel implementation</td>
</tr>
<tr>
<td>Mode</td>
<td>Inference using multi-temperature MCMC with model based proposals</td>
</tr>
</tbody>
</table>

**Key questions**
- How is $R$ changing as a function of time?
- What determines the spatial spread of disease?
- What are the effects of government policies?
To Do…
- Comprehensive validation.
- Results for age structured populations with sex.
- Incorporation into data pipeline.

Model parameter inference
- Data: National hospital admissions/populations and regional death data.
- Distributions for certain transitions.

Variation in R
- Captured time variation in disease transmission before/after lockdown.

Population density
- Higher disease transmission in areas of high population density

Different area factors, e.g. deprivation, pollution.
- Sensitivity analysis.
Open Epidemiology

Requirements
- Reproducible
- Traceable
- Validated
- Open source
- Open data
Open Epidemiology

Requirements
- Reproducible
- Traceable
- Validated
- Open source
- Open data

https://github.com/ScottishCovidResponse
Enabling software teamwork

Github 33 repositories, 14 teams, 13 projects

RSE

Alys Brett (UKAEA)

Team processes: Lead RSEs

Zulip chat:
Now reached >5000 weekly messages

May June July
Software Quality

• Improving code clarity and software structure
• Automated testing and code analysis
Improving reproducibility and trust

- Data pipeline integration
- Managing configuration, randomness, build, release run processes

Software Checklist

https://github.com/ScottishCovidResponse/SCRC-internal/blob/master/software_checklist.md

- Framework for evaluating software engineering aspects of model readiness
- Completed for current status of all models
Simple Network Sim

- Synchronised docs and code using docstrings & typehints to making it easy to get into the codebase
- Extensive automation of unit and regression tests and code-checking through CI - lets us focus on adding features and fixing bugs
- Adopted industry standard Python data science tools `pandas` and `numpy` - peer-assured variable-types and functions
- Tight control of execution using `conda` virtual environment to help ensure reproducibility
- Great multidisciplinary working and input to current code from a range of contributors: academics, quantitative researchers, postdocs, engineers

Simulation.jl

- Software teamwork practices (first multi-developer team)
  - Issue boards & labelling, contribution and review guidelines etc
- Automated testing
  - Travis & Appveyor for CI
  -Codecov for test coverage
- Extensive testing of new epidemiology code
- Reproducibility improvements
  - "Deterministic hijacking" of the stochastic model to allow regression testing
  - Run scripts, TOML files for code & dependancy versions
- Self-contained example run scripts serve as end to end tests
- Integration of data pipeline
Contact Tracing Model and Covid Simulation Model

- Many model refinements and new features
- Initial code clean-up then incremental improvements to design & implementation
- Automated tests and CI
  - 90% unit test coverage, regression tests, basic automated epidemiological/model-validation tests
- Documentation for user, developers & to understand algorithms
- Improved reproducibility and trustworthy results
  - refactoring RNG usage
  - moving parameters to an input file, adding logging of git hash and inputs
  - R scripts for validation, version numbering and release processes
- Enabled parallel runs of models on DiRAC HPC
- Implementing the Java data pipeline
  - Close to end-to-end model run (using CSV instead of HDF5 files initially)
New functionality: Multiple model structures, forward prediction / parameter inference modes

Reproducibility: Parameterised config of random number generation, Logging inputs, outputs, versions

Software engineering and code quality
• Extensive refactoring and architecture improvements
• CMake build system
• Unit and regression test frameworks
• CI automation of: tests, documentation generation (Doxygen), code coverage, static analysis
• Created project dashboard
  https://scottishcovidresponse.github.io/Covid19_EERAModel

Development of C++ interface to Python data pipeline code, and integration with model (in progress)
• Fixed configuration in header files replaced with TOML configuration file
• Storing code provenance information and making runs reproducible
• Improved code structure, i.e. making it more modular
• Implemented CI (automated testing), as well as regression testing (flag whether a code change affects the numerical results)
• Added automated source code analysis to catch bugs - found several
**Imperial CovidSim Model**

**Peter Hill** (York)

**Integration with SCRC data pipeline**
- Via a Python wrapper layer using pipeline API
- No need to touch original code
- Can treat chosen version as a black box whilst verifying inputs & outputs

**LSHTM Model**

**Kristian Zarebski** (UKAEA)

- **Rationalised parameter initialisation**
  - external config at start of run
- **Faster runs** with single scenario selected via parameter
- **Integration with the SCRC data pipeline**
  - R frontend for python to interact with pipeline API
  - Inputs converted to HDF5 & TOML - now moving to data pipeline locations
- **Collaboration with Swansea RSEs**
Initial specification of Data Registry: with Aaron Reeves (SRUC)

Specification of data and metadata workflows

Literature review: procedure and associated data extraction tool

Dataset assessment: 'intrinsic quality', 'representativeness' and 'fitness for purpose' assessment procedure

Meta-analysis and distribution analysis: Jim Lewsey, Houra Haghpanahan, Ryan Field, Janine Illian, Caitlin Fox, Mahi Siddika, Megan Laxton, Martje Rave

Data brokering: Richard Bailey, Ciaran McMonagle

Specialist Dataset expertise: Ruth, Ciaran, Jeremy Walton, Denise Brown, Stephen Catterall, Amira Elayouty

Data source and provenance IDing and wrangling: Sonia Mitchell, Ruth, Ciaran, Denise

Data team: Richard, Ciaran, Sonia, Ryan, Jeremy Gribble, Hannah Jevans, Nardus Mollentze, Denise
Data pipeline

Requirements
- Reproducible
- Traceable
- Open source
- Open data

API
- C/C++
- Java
- Python
- Julia
- R

Data and metadata flow
- Enter source data into pipeline
- Download inputs to model
- Run model using Data Pipeline API
- Upload outputs to pipeline
- Reporting on model outputs

Component
- Provided by user
- Core pipeline / SCRC
- From other source
- Generated by pipeline

https://github.com/ScottishCovidResponse
Data pipeline for input

https://github.com/ScottishCovidResponse/SCRCdataAPI
https://github.com/ScottishCovidResponse/SCRCdata

> 31,000 papers uploaded to the registry

8 data products uploaded to the registry

4 upload templates

Working on processing additional data products, generating upload templates, and checking robustness of code

Sonia Mitchell
Ciaran McMonagle
Nardus Mollentze
Ryan Field
Data pipeline for modelling

Richard Blackwell, Chris Hughes, Vino Mano, Rafael de Almeida (Man Group), Ian Hinder (Manchester), Kristian Zarębski (UKAEA), Alex Robson, Sean Lovett (Invenia)

https://github.com/ScottishCovidResponse/data_pipeline_api
Data registry

Jonathan Hollocombe and Andrew Lahiff (UKAEA)

https://github.com/ScottishCovidResponse/data-registry

- Available at data.scrc.uk
- Provides web-page view of stored data and information about registry API.
- API can be seen at data.scrc.uk/api
- More info available at data.scrc.uk/docs
Epidemiology
- Based on consensus by Irish Epidemiological Modelling Advisory Group (IEMAG)

Inference and model assessment

Inference: quantify epidemic and effect of interventions
- focus on simulation-based methods (model agnostic + parallelisable)
- But also data-augmentation MCMC methods (BEEPmbp)
- And emulation for Uncertainty Quantification and model fitting

Model assessment: better models
A key issue moving forwards
- statistical fit (evidence, residuals, DIC etc.) and predictions
- model misspecification
- validation within and btw model/code-base

Now in more detail from
- Ben Swallow (UoG)
- Rikiya Takahashi and Fedor Gorokhovik (Man Group)

Inference* + SA > quantify uncertainty > model assessment > better models
*focus on Bayesian inference
Inference and model assessment

Ben Swallow (UoG)

Inference: quantify unknown epidemiological parameters

- Again focus on simulation-based methods (easily parallelisable)
- Uses an ABC-SMC algorithm to estimate 6 parameters on 1 km² grid of Scotland
- **quantify uncertainty** in model output and parameters
- Flexible to varying data outputs and thresholds

Model testing

- Using data from model simulations at known parameter values we are able to retrieve the true values
- Stochastic model so repeated runs at the same parameter values
- Tested and seems all parameters retrieved using either deaths or symptomatic cases

Issues/open questions

- Link to data pipeline
- Incorporate priors beyond uniform
- Consolidate multiple compartment summaries

Bayesian inference: simulated data

Red solid line = true parameter value
Blue dashed line = posterior mean
Inference and model assessment

Rikiya Takahashi and Fedor Gorokhovik (Man Group)

Sensitivity Analysis

- Reparametrisation of exposure-related disease parameters in Contact Tracing Model
- Non-linear regression and SHAP analysis for total number of death and peak number of severely symptomatic persons
  - SHAP: SHapley Additive exPlanations (Lundberg & Lee, 2017)
- Efficient frontier in optimisation of the policy parameters under uncertainty

Open Issues Real-data-based posterior in policy optimiser, distributional policy parameter recommendations instead of point estimates

Inference

- ABC-SMC used to infer infection probability and initial infections in Simple network sim (contact multipliers to come)
- To be used in conjunction to stochastic mode to have more realistic posterior std dev
- Fitted on historical death per healthboard in Scotland
- Other parameters sources from external sources to be as credible as possible
- Wide priors chosen from Beta, Lognormal distribution

Open Issues Improved rejection mechanism, parallelization of inference

VIS

27 VIS/VA volunteers, plus Min Chen (Oxford U.)

https://github.com/ScottishCovidResponse/rampvis-ui

Dr. Alfie Abdul-Rahman, KCL
Dr. Daniel Archambault, Swansea
Dr. Benjamin Bach, Edinburgh
Dr. Rita Borgo, KCL
Professor Min Chen, Oxford
Professor Jason Dykes, City
Dr. Hui Fang, Loughborough
Elif Firat, Nottingham
Dr. Radu Jianu, City
Professor Nigel W. John, Chester
Dr. Saiful Khan, Harus Security
Dr. Robert S Laramee, Nottingham
Dr. Stella Mazeri, Edinburgh
Benjamin Nash
Dr. Phong H. Nguyen, Red Sift Ltd.
Dr. Helen C. Purchase, Glasgow
Dr. Dylan Rees, Swansea
Dr. Panagiotis Ritos, Bangor
Professor Jonathan Roberts
James Scott-Brown, Edinburgh
Dr. Aidan Slingsby, City
Dr. William Teahan, Bangor
Dr. Thomas Torsney-Weir, Swansea
Dr. Cagatay Turkay, Warwick
Dr. Franck Vidal, Bangor
Qiru Wang, Nottingham
Tianci Wen, Bangor
Professor Jo Wood, City
Dr. Kai Xu, Middlesex

Call for VIS volunteers
22 volunteers
1 May

Generic support team
(4 core developers + 5 others)

Analytical support team
(3 people) + text mining (1 person)

Four modelling support teams
(total: 12 people)

Disseminative visualization team
(3 people)

VI server architecture
2 dashboards 30+ plots
New UI, portal, search

Similarity metrics + VIS
model-specific VIS designs

30 July Generic support team reporting
9 July Analytical support team reporting
9 July Modelling support teams reporting

a survey of existing provision
an initial design of storytelling VIS
27 VIS/VA volunteers, plus Min Chen (Oxford U.)

https://github.com/ScottishCovidResponse/rampvis-ui

rampvis-api

scrc-vis-analytival

scrc-vis-modelling

https://sites.google.com/view/rampvis/
Open Epidemiology

Requirements
- Reproducible
- Traceable
- Validated
- Open source
- Open data

Responsibility
- SCRC data pipeline
- SCRC modelling platform
- SCRC visualisation
- SCRC model assessment
- Policy and government interface

Action
- Modelling platform assessment
- Policy pathway
- Input data into model
- Output results from model
- Visualisation pathway

Data → Data registry

Model 1
Model 2
Model 3
Model M

Run model

Policy decisions
- Policy briefs
- Scored model library

Score model and outputs

Completed Scorecards

https://github.com/ScottishCovidResponse
Administration

Lynn Esson (Glasgow) and Yasmin Abdalla (Edinburgh)

Rachel O’Brien (Glasgow)
Mark Meenan (Glasgow)

Requirements
- Extreme patience