**BLOFELD: BroadLy Optimal FramEwork for modeLling Disease**

Mossa Reimert, Carsten Kirkeby, Maya Gussmann, Søren Nielsen,

Anette Boklund, Jens Frederik Agger, Matt Denwood

Version 5 – 2020-12-15

**Background**

Modelling the spread of infectious diseases within Denmark has become a core component of the veterinary contingency services and is now a vital tool to be deployed in the event of a disease outbreak. Models for FMD, ASF and BT have previously been developed by individuals, who are no longer involved in deliverance of the contingency services, and although (with the exception of BT) the code for these models remains available to the group, the remaining modellers within the group do not currently feel comfortable with using these models in their current form. This currently leaves the DK-VET consortium in an extremely vulnerable position in the event of a disease outbreak in Denmark.

Over the past month, a section-wide consultation process has identified the core requirements for FMD, ASF, BT and AI models in terms of the features these models must contain in order to deliver our core obligations to FVST. This exercise has revealed the following general points:

* The models must necessarily be complex in order to capture all the features that have been identified as essential (including those relevant to EU legislation). This has important consequences for the fundamental design of the models in that they must be capable of handling this complexity without losing transparency of the core processes within the model, or the ability to isolate and test individual aspects of the model to ensure that programming errors are not driving the outputs of the models. These criteria are not met by the models that are currently available, which have been developed by adding complexity incrementally over a long period of time, rather than being originally designed to handle this complexity.
* Although there are some requirements that are unique to specific diseases, the vast majority of the requirements are identical between ASF, FMD and BT, and several requirements are also shared with AI. This ability to generalise a substantial part of the models to be consistent between diseases has huge potential advantages in terms of future maintainability of models for these four diseases, and also highlights the likely additional overlap with other pathogens of pigs and ruminants, as well as potential for partial overlap with pathogens of other host species.
* The essential data requirements are already substantial for many of the diseases, and the potential inclusion of additional desirable data sources further increases the complexity of updating data sources within the models. However, many of the same data sources are common between diseases, and other work within the section both ongoing and planned to start in 2021 will greatly expand the possibilities for automated data access from external sources.

**Project Proposal**

The most viable long-term future for meeting the identified requirements is to transition towards a unified modelling framework that meets the needs that we have discovered during the section-wide consultation process. By “modelling framework”, we do NOT mean a specific disease model, but rather the underlying code and tools that address the needs that are common to all disease models, including:

* A concept of spatially-fixed “animal groups” (either herds of cattle/sheep/goats/pigs, flocks of domestic chickens/turkeys, or groups of wildlife within the same habitat)
* A user-specified disease model (either simple e.g. deterministic SI model, or more complex e.g. stochastic SELIR model with multiple animal groups) describing spread of disease within a single spatially-fixed “animal group”
* An ability for disease to be transmitted between animal groups according to user-specified contact/spread patterns
* A concept of “disease detection” by both passive and active surveillance, according to user-specified characteristics including imperfect sensitivity and specificity and heterogeneity in test performance/application between animal groups
* A concept of “disease control” that can affect individual animal groups (by culling, vaccination, etc) as well as disease transmission between animal groups (by movement restriction, elevated biosecurity, etc)
* A concept of “macro-level” outputs of the model, including resource utilisation by FVST and economic consequences of a disease outbreak at different spatial and organisational/societal levels
* Ability to input data and parameters, for example locations of animal groups and numbers of animals, contact/spread patterns informed by animal movements
* Ability to output summary measures at different levels, including farm-level measures (on-farm prevalence), day-level measures (prevalence of affected herds within different herd categories, and/or herds in different geographical regions and/or surveillance zones), and iteration-level measures (macro-level outputs such as total outbreak size/cost)

This list of features may seem basic, but in fact there are a number of potential ways of implementing these features at the most basic level, each of which has different consequences for flexibility, usability and computational efficiency of models implemented using the resultant framework. We will spend the necessary time examining the consequences of design decisions at this most basic level to ensure that we have the optimal framework for use by each of the following user groups:

* **Architects**: those implementing the framework itself, and making fundamental design decisions will impact the way in which all users interact with the framework. These individuals will have a complete understanding of the internal implementation of the framework along with all complexities this entails. Future modification and/or expansion of the framework will require input from one or more of these individuals.
* **Developers**: those implementing disease models, spread models, detection, intervention and macro-level summary layers within the framework. These individuals will be required to take on a moderate level of programming complexity depending on the nature of their disease model, but will not be required to understand the full complexity of the framework implementation in order to use it. The possible implementation options available to developers will be somewhat constrained by the architects in order to ensure consistency between different models, but all viable use cases will be possible within the framework and extensive documentation will be provided by the architects to facilitate the development of new models. This will ensure that development of new models within the framework will be more straightforward than development of equivalent models independently. Developers will be able to use a number of different programming languages according to their preference, using the relevant API provided by the architects. Additional requirements such as provision for unit testing will need to be implemented by the developers in order to ensure integrity of new code, as dictated by the architects. Collaboration on model development, as well as transfer of “ownership” of a model from one individual to another, will be greatly facilitated by the consistency and modularity that will be intentionally enforced by the framework.
* **Practitioners**: those using models provided by developers, who are required to update data sources, modify parameter arguments passed to models, activate/deactivate different disease spread, disease detection, disease control and macro-level elements within the models, and run models according to needs for different outputs. These individuals will require only a very basic level of programming ability and will be heavily supported by use-case documentation provided within the framework. A critical advantage for these users is that experience with using the framework for one disease will be directly transferable to using the framework for another disease, which will greatly reduce the cognitive burden of using different disease models in an outbreak situation. These individuals will not be required to understand the internal implementation of either the framework itself or the disease model they are working with, including technical decisions such as choice of different programming languages for different aspects of the model.
* **Stakeholders**: those depending on the outputs of specific disease models to make policy decisions, and therefore must be able to completely trust that the models produce consistently correct output without being able to understand the internal implementation, but do not necessarily need to run the models themselves. The modular design of the framework combined with the enforced use of unit testing for each disease model will provide this security and therefore reduce the potential criticism of our models from interested parties with more extensive software development experience. Although this has not yet been an issue within the veterinary world we highlight the advantage of addressing the potential criticisms before they are publicly made in the event of a disease outbreak.

We note that a number of disease modelling frameworks have already been developed by other groups, including InterSpread (<https://www.interspreadplus.com>), EcoEpi (<https://www.ecoepi.eu/CSFWB/>), disease-specific models such as EuFMD (<http://www.fao.org/eufmd/en/>), and tools such as those provided by Ausvet (<https://epitools.ausvet.com.au>). Re-using these frameworks would likely be simpler than developing a new framework specifically for our purposes, but it should be noted that in this case the role of “architect” as described above is taken by an outside agency, and modellers within AWDC are limited to the roles of “developer” and/or “practitioner” with limited ability to influence the capabilities of the framework to meet our (frequently Denmark-specific) requirements. For this reason, we do not believe that the use of an externally provided framework is a viable option. Similarly, we do not necessarily expect that our framework will be immediately applicable to other international groups with potentially different requirements, although we are hopeful that the scope of the framework may be broadened to include international collaboration in the future. International collaboration on projects such as DigiVet can be used to explore the potential for future project proposals in this area.

**Relevant Experience**

We note that the design of this framework is primarily a computer science task rather than a disease modelling task, and should therefore be informed by best practices in computer science and software engineering whilst simultaneously being guided by the requirements identified for each disease. We will therefore draw on experience from a wide variety of projects to help us achieve these goals, including:

* Experiences from the MAPRA project, particularly in relation to the design of a BT model for use by EFSA (the “EFSABT” model). This model was implemented within a framework that provides modularity of different aspects of the model, so that users can for example heavily modify aspects of the between-farm spread model (by adjusting R code) without needing to understand or modify the underlying disease model (written in C++). We note that this framework is not sufficiently complex for our needs and cannot therefore be simply re-used, but the experiences gained from developing the framework are highly relevant. Some of the relevant output from this project is attached as an appendix, and the full report is available from: <https://efsa.onlinelibrary.wiley.com/doi/abs/10.2903/sp.efsa.2016.EN-1112>
* The currently active PhD project on ASF has highlighted the difficulty in transferring “ownership” of the existing disease models between individuals. Lessons learned during this process will be leveraged to ensure that the framework we develop minimises these difficulties.
* Between members of this group we have several decades’ worth of experience in developing and using disease models for different diseases including FMD, BT, BVD, CSF, ASF, mastitis, paraTB, and numerous other examples. Our experiences will be used to ensure that past mistakes are not repeated.
* Our involvement in work on Covid19 modelling has also highlighted similar issues with collaborative development of disease models, particularly when under extreme time pressure i.e. in the event of an outbreak. We have learned a large number of lessons from this process, some of which are expressed in a recent publication: ﻿<https://www.frontiersin.org/article/10.3389/fvets.2020.00513/full> We also note the widespread criticism from “real” programmers of disease models developed by other disease modelling groups, for example the headline “Neil Ferguson’s Imperial model could be the most devastating software mistake of all time” and detailed criticism given by <https://analyticsindiamag.com/the-most-devastating-software-mistake-of-all-time-why-is-the-imperial-model-under-criticism/> - a desire to avoid similar criticism on our future work is a strong motivating factor behind development of a more robust modelling framework.
* Some members of the group have experiences in contributing to large-scale collaborative software development projects, most notably including JAGS (<http://mcmc-jags.sourceforge.net>), and a number of R packages that are used outside the section including runjags (<https://CRAN.R-project.org/package=runjags>), ExtendR (<https://github.com/extendr/extendr>), and vtree (<https://github.com/nbarrowman/vtree>). Although not directly related to disease modelling, this experience of software development is highly relevant to design and creation of a modelling framework, in particular the experience gained from integration of multiple programming languages within a single software framework.

**Integration with Existing Projects**

There are no fewer than four existing research projects with synergistic potential for our proposed framework development, and we intend to fully exploit these potentials to enhance delivery of all projects.

Damo

The Damo project aims to update the data sources used by the FMD and ASF models, and to re-run these models using the new data sources to explore the differences in model outputs compared to outputs based on the older data. The original plan was for the model runs to be completed by Tariq Halasa, who has extensive experience with these models and could therefore be expected to address any unexpected issues arising from use of the existing models with new data. This option is no longer available, and delivery of the Damo project may well necessitate several additional months of work to fully understand and address unexpected issues in the existing models that are illuminated by the updated inputs. We believe that investment of such a large amount of time in the existing model is not a good use of limited time resources within the section. However, we consider that spending time on a re-implementation of the existing model within a more suitable framework is a much more productive use of resources, and will have major secondary benefits to the delivery of the current Damo project as well as any future extensions of this work. We therefore propose the use of some of the existing Damo budget to finance the development of the modelling framework.

African swine fever

The ongoing PhD project on the topic of ASF modelling will initially be set back due to the diversion of resources to the framework development, but the development of the framework and later use of this framework for the continued development of the ASF model will be highly beneficial to the outcome of the project as a whole. Therefore, although delivery of the model to FVST may occur later than planned, the delivered product will be much stronger as a result.

DigiVet

The externally-financed DigiVet project aims to develop a holistic solution to livestock data accessibility and interoperability by developing common data pipelines that can be used for accessing similar datasets in the UK, SE, NO, EE and DK. Use of these datasets to inform modelling of infectious diseases comprises one of the three case studies within the project, and there is therefore a substantial crossover in terms of the data requirements identified by the section-wide consultation and the planned objectives of the DigiVet project, not least due to the direct involvement of FVST on the DigiVet project. It is therefore sensible for work on the modelling framework to occur in parallel with the early activities of DigiVet, so that we can ensure that the eventual data pipeline outputs from the DigiVet project will be completely compatible with the disease models being used within the section. We also note that the funded project requires a time commitment of 4 months per year from seniors within the section, the majority of which was to be used by MD. By transferring some of this work burden to MMR, this will reduce the time pressure on MD during 2021.

GOLDFINGER

A substantial amount of work has already been invested in streamlining data transfer from SEGES, partly for the benefit of the veterinary contingency activities. By integrating the future activities of GOLDFINGER with both DigiVet and work on the modelling framework, we will greatly facilitate future updates of data sources involving datasets held by SEGES. In particular, we will be able to ensure that the implementation details of the data extraction solution currently being developed in collaboration with SEGES produce datasets that are fully compatible with the inputs required by the modelling framework.

Future projects

The key outcome of this proposed project is to facilitate development of models in future projects, including modelling projects funded directly the DK-VET consortium as well as other internally funded and externally funded projects. It is impossible to generate a complete list of all future projects that will benefit from this work, but a partial list includes:

* Avian Influenza model development
* Bluetongue model development
* Future development/refinement of the FMD and ASF models in response to bestilling from FVST
* Development of a model for Salmonella Dublin in cattle

Although primarily envisaged as a between-herd disease spread model, the framework we propose will also include within-herd disease spread models of arbitrary complexity. It would therefore also be possible to use the framework for the development of within-herd models (e.g. mastitis) by excluding the between-herd spread elements. A future project may be useful to explore the utility of the framework (or perhaps an expanded framework) for these purposes.

We note that the proposed framework would have a strong emphasis on usability for users with a range of programming abilities, and that a simplified disease model example implemented within this framework could potentially be used for teaching purposes both for veterinary students as well as other groups of individuals (virologists, FVST, etc) as part of future knowledge transfer activities within the section. Previously this has been done using models implemented in InterSpread Plus to give people some hands-on experience of “playing around” with the model and exploring the effect of increasing zone size etc, but the options for using existing models in this way are somewhat limited. Having more direct control over the user interface would create additional possibilities for creating improved teaching tools as part of future activities.

**Detailed Project Description**

WP1: Assessment of Requirements

This consultation WP aims to ensure that all existing and future needs for disease modelling are captured by the framework that will be developed. There will be three WP outputs (the first two of which will involve wide consultation within AWDC):

* Biological requirement sheet for each of FMD, ASF, BT and AI, to ensure that the framework is capable of capturing all important aspects of the disease.
* Technical requirement sheet to ensure that the framework is capable of meeting the language preferences and programming levels of developers and users within AWDC.
  + Development plan for the framework itself, giving an overview of the planned implementation as well as technical details of any limitations and requirements that will be imposed on users and developers. Drafted by MD and MMR after consideration of the potential design decisions, and iterated based on comments and concerns raised by CK/MG (representing anticipated developers), and SSN/AB/JFA (representing anticipated users).

Responsible: MD (1 month)

Contributors: MMR (1 month), CK, MG, SSN, AB, JFA (2 weeks each)

Funding: Section funds or arbejdsprogrammet 2021 (MMR), core duties (others)

WP2: Framework Development

This technical WP will produce a completed modelling framework, complete with documentation for architects, unit tests, and the completed user interface. The framework will be disease-agnostic but will use simplified disease processes for development and testing purposes. Periodic feedback will be obtained from anticipated developers to ensure that the framework is fit for purpose. There will be two outputs from this WP:

* + A software package implementing a disease-agnostic model framework, from which specific models can be derived by importing and re-exporting the package. The package will be self-contained, cross-platform, and hosted on a central repository site with version control and continuous integration services.
  + A peer-reviewed publication detailing the motivation, features and implementation of the modelling framework.

Co-responsible: MMR (3 months) and MD (1.5 months)

Contributors: CK and MG (2 weeks each)

Funding: Either section funds or arbejdsprogrammet 2021 (MMR), core duties (others)

WP3: Data Interoperability

This WP will run alongside DigiVet and GOLDFINGER, with the majority of the work and outcomes being delivered as needed by these projects. We note that the international nature of the DigiVet project may well result in external parties becoming interested in our framework, and perhaps contribution via the centralised repository. This is particularly relevant for colleagues tackling similar problems in the UK, SE, NO and EE. The main outcome of relevance for the modelling framework is to ensure that design decisions made during development of the framework within WP2 are compatible with the data pipelines being made available as outputs of DigiVet and GOLDFINGER. There are therefore no specific outputs for this WP other than those specified by DigiVet and GOLDFINGER.

Responsible: MD (3 months)

Contributors: MMR (5 months), SSN (2 months)

Funding: External via DigiVet (MD and MMR) and GOLDFINGER (MD and SSN) projects.

WP4: FMD Model

Development of the modelling framework will be greatly facilitated by a concrete use case. We will therefore develop an FMD model in parallel with the framework, which will duplicate the features of the existing FMD model and therefore be available for use to ensure delivery of the Damo project outputs. As much as possible of the existing code (from both ASF and FMD models) will be re-used, but a need for some re-coding is anticipated. There will be three outputs from this WP:

* + A like-for-like replacement of the existing FMD model within the new framework, with improved possibilities for future development and use of the model
  + Ensuring delivery of the Damo project deliverables using a combination of the existing and redeveloped FMD models
  + A short publication highlighting the features of the re-implemented model as well as computational advantages compared to other FMD models that are available

Responsible: MMR (2 months)

Contributors: MD (2 weeks)

Funding: Damo project

WP5: Training / knowledge transfer

Following completion of the modelling framework and first use-case in the redeveloped FMD model, we will hold separate training sessions for potential developers and users of our framework in order to facilitate immediate and widespread uptake of the modelling framework within the section and beyond. We will also provide coding support as part of these activities, and resolve any issues that arise within the framework as a result of the expansion to additional use cases. This WP is not essential to successful completion of the rest of the project, but is fundamental to the uptake of the modelling framework outside of the core developers, which is the most important long-term outcome of the project. There are three WP outcomes:

* + Completion of the framework documentation for developers, users and stakeholders, with multiple iterations of the documentation informed by feedback from within AWDC.
  + A one-week training session delivered by the framework architects for the benefit of potential developers, so that they are able to immediately start using the framework to develop their own models. Part of this training session will also include coding support over a slightly longer period to minimise any stress or confusion caused by new programming concepts used within the framework.
  + A one-day training session delivered by the entire project team for the benefit of potential users and stakeholders both within and outside AWDC. This will focus on the “high-level” features of models implemented within the framework, as well as highlighting the robustness delivered by the modularisation and unit testing within the framework, but will not be technical in nature.

Responsible: MD (1 month)

Contributors: MMR, CK, MG, (1 month each), SSN, AB, JFA (2 weeks each)

Trainees: All interested parties within or affiliated to the section, DK-VET and FVST

Funding: Arbejdsprogrammet 2022



**Budget**

We have worked hard to devise a project that can be met mostly from existing project resources for 2021. However, some additional commitment of resources is needed to guarantee the remainder of the time that MMR needs away from his PhD. We are aware that the arbejdsprogrammet for 2021 may now be fixed, so it might be that resources must be committed from the section budget, but we believe that this represents good value for money for the section overall. Delivery of the training WP5 will require a project to be funded as part of the 2022 arbejdsprogrammet.

Budget for 2021:

* + No additional salary costs are being requested for MD, CK, SSN, JFA, AB, MG; this work will be done as part of their commitment to core activities and external projects
  + No travel or consumables expenses are being requested (where necessary, expenses will be met from the DigiVet and GOLDFINGER projects)
  + Salary costs for 10 months of a scientific assistant (step 6) are required for Mossa
    1. 120.000 DKK will be used directly from the DigiVet project
    2. 150.000 DKK will be used directly from the Damo project
    3. **An additional commitment of 145.000 DKK is needed for 2021**

Budget for 2022:

* + One month of funding for Mossa (42.000 DKK) will be used directly from DigiVet
  + A project is requested under the arbejdsprogrammet for 2022 to hold a 1-week training event for developers within the section (around 5-10 people), and a second 1-day training event for users & stakeholders (max 20 people):
    1. Salary of one month each for MD, MG, CK, MMR to cover planning/preparation/delivery of these events
    2. Salary of two weeks each for AB, JFA, SSN to cover planning/delivery/participation in the 1-day event
    3. Expenses budget for tea/coffee/lunch for up to 10 people for 1 week
    4. Expenses budget for tea/coffee/lunch for up to 20 people for 1 day
    5. Additional small budget to cover ancillary costs
    6. Total costs approx.:
       1. 1-month salaries: 42+46+60+60 = 208.000 DKK (approx)
       2. 2-week salaries: 30+30+40 = 100.000 DKK (approx)
       3. One-week course: 12.500 DKK (assuming we use G8)
       4. One-day course: 7.500 DKK (assuming room hire)
       5. Ancillary: 2.000 DKK

Note: We propose that the budgets for 2021 and 2022 be considered independently, and that a formal (and more precise) request for resources from arbejdsprogrammet 2022 be made at a later date, contingent on work beginning on this project on 1st March 2021.