A COMMUNITY RESOURCE FOR INTEGRATED PREDICTIVE MICROBIAL MODELLING (PMM-Lab)

Matthias Filter (*), Christian Thöns, Jörgen Brandt, Armin A. Weiser, Alexander Falenski, Bernd Appel, Annemarie Käsbohrer

Federal Institute for Risk Assessment, Unit Epidemiology and Zoonoses, Department 4 - Biological Safety; Max-Dohrn-Straße 8–10, 10589 Berlin, Germany

(*) corresponding author: matthias.filter@bfr.bund.de , Tel. +49 30 18412-2209

Abstract

The prediction of growth and inactivation of microorganisms in food matrices is a field of intensive research with increasing relevance to food safety professionals and public authorities. Today, large quantities of experimental and quality control data are available in comprehensive public or proprietary databases. In parallel the mathematical foundation for the generation of predictive microbial models has been developed and is available. Nevertheless, there is a gap in easy-to-use and transparent software solutions that enable e.g. food safety professionals to apply state-of-the art mathematical modelling concepts to their proprietary data that provide additionally an integrated data and model management solution.

The software PMM-Lab has been specifically designed to address this gap. It is outlined right from the beginning as community resource to allow broad application and joint development. PMM-Lab already provides many valuable data analysis and modelling features including modules to create, visualize, analyze, save, import, export and deploy predictive microbial models based on experimental data. Moreover, it allows consistent data management for the analyzed experimental data and generated models due to the integrated database. Technically, PMM-Lab is an extension to the open-source data integration and analysis platform KNIME (www.knime.org) from which it inherits highly beneficial properties like modularity, flexibility, scalability and extensibility.

PMM-Lab is designed to be used by laboratory scientists, food safety specialists or even modelling experts who wish to distribute their methods to a broad community. Also users trying to answer fundamental research questions based on experimental data collections or looking for structure in their growing data and model collection will find value in this software. Possible application areas are numerous including prediction of shelf life and food spoilage. PMM-Lab has been developed by the Federal Institute for Risk Assessment, Germany and is freely available at https://sourceforge.net/projects/pmmlab/ under the GNU public license.

1. Introduction

Our research objective was to develop a community resource facilitating the creation, collection, sharing and application of Predictive Microbial Models (PMMs) [1] which extends existing community resources like the ComBase database [2] on experimental microbial data. PMMs are mathematical equations that relate the number of microorganisms observed during an experiment or a process to time / duration considering the initial microorganism level and environmental conditions like competing microorganisms, food matrices, additives or other physical or chemical factors [3]. Usually, PMMs belong to the group of parametric regression models. Ultimate goal of any PMM is to make predictions on the number of microorganisms for time points and environmental conditions which have not been experimentally determined. Comprehensive overviews on currently applied model types and model equations can be found in [4–6].

2. Material and methods

On the basis of the open-source software framework Konstanz Information Miner (KNIME) [7] the new community resource "PMM-Lab" has been developed. KNIME (www.knime.org) is an open-source platform for integrated data access, data mining, statistics, data visualization and reporting. KNIME allows users to graphically assemble data analysis workflows from libraries of functional building blocks called "nodes" which are connected by non-functional "pipes" that solely determine the data and information flow within each specific workflow.

PMM-Lab extends KNIME through a library of specifically designed nodes and an integrated HSQL database (www.hsqldb.org), called PMM-Lab DB. PMM-Lab nodes comply with the KNIME interface definition and can be integrated into the KNIME platform seamlessly. Nodes were programmed in

Java using the open-source Integrated Development Environment (IDE) Eclipse (www.eclipse.org) and software libraries available within. In part, PMM-Lab nodes make use of additional open-source software libraries like the Jep Java 2.4.1 GPL library (http://www.singularsys.com/download/jep-2.4.1-ext-1.1.1-gpl.zip) which is applied in PMM-Lab nodes that parse and evaluate mathematical equations. From the Apache Commons Math library (http://commons.apache.org/proper/commons-math/) the implementation of the Levenberg-Marquardt algorithm is used in the node "Model Fitting" that performs model parameter estimation. PMM-Lab nodes generating diagrams of data and models use the JFreeChart package (http://www.jfree.org/jfreechart/).

3. Results:

The PMM-Lab plug-in extends the KNIME software framework such that a plethora of predictive microbial modelling tasks can easily be accomplished. This software is primarily intended to support domain experts in their efforts to create, document and share predictive microbial models in a transparent fashion. Additionally, the software can help teaching good modelling practice as well as easing model verification, validation and deployment. To provide users with an opportunity to participate actively in the PMM-Lab community a software repository including Wiki and issue tracking system has been set up at http://sourceforge.net/projects/pmmlab.
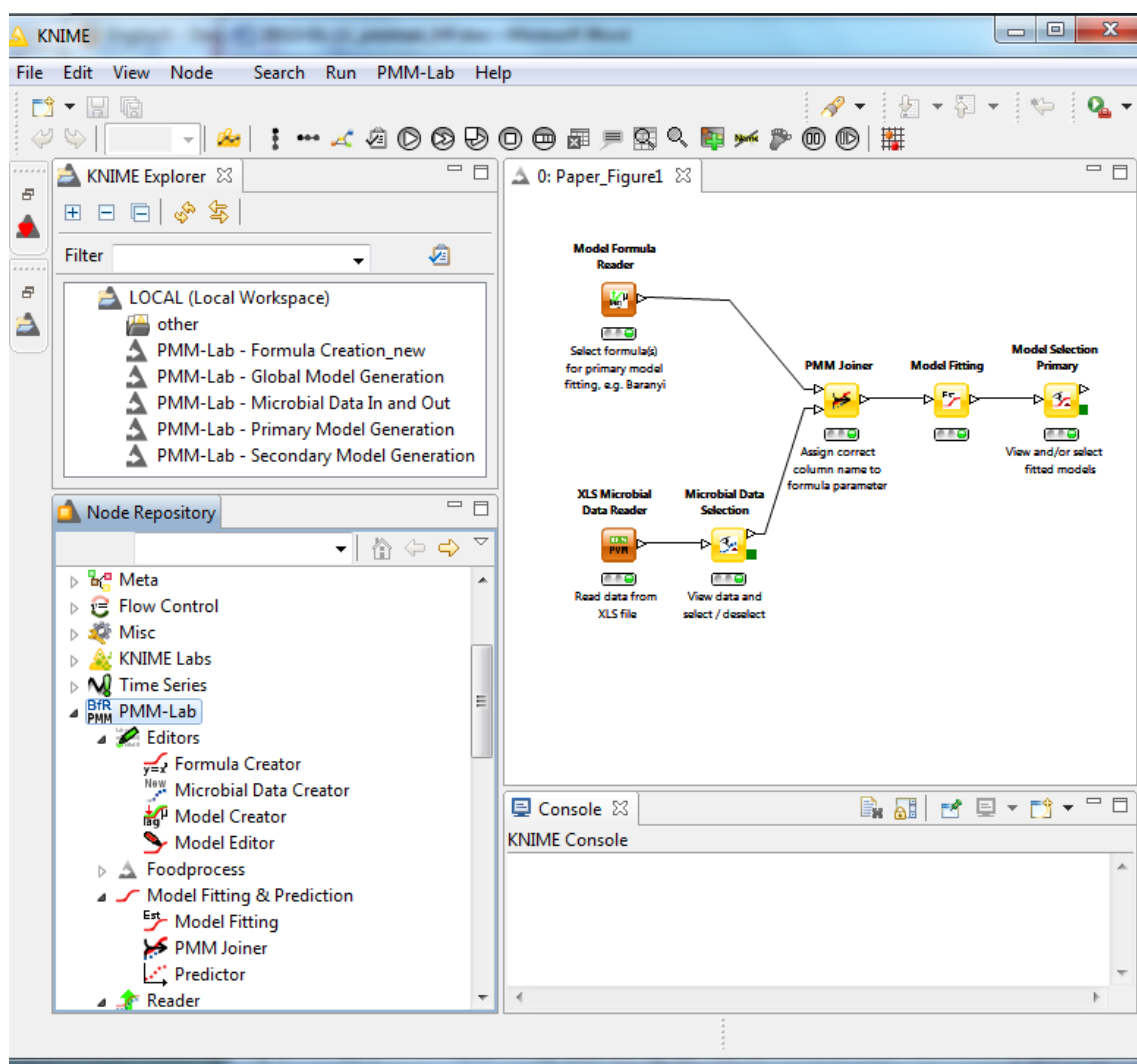

Figure 1:
Screenshot of the KNIME-GUI with integrated PMM-Lab node library.
The customizable KNIME-GUI comprises in its standard version of 4 views displaying
a) the KNIME Explorer - giving an overview on all workflows in the active workspace (upper-left),
b) the Node Repository - including the PMM-Lab nodes (lower left),
c) the Workflow Editor - as the central interactive workflow generation, modification and execution view (upper right)
d) the Console - which displays warnings or error messages to the user.

The depicted workflow in the Workflow Editor view can be used to create primary models based on data read in from a XLS-file (node XLS Microbial Data Reader) and a formula selected from the internal PMM-Lab DB "Model formulas" table (node: Model Formula Reader). The assignment of variables from the selected formula to variables from the read-in experimental data has to be performed manually by the user in the configuration menu of the PMM Joiner node. Subsequently model parameters are estimated by the execution of the Model Fitting node and fitting results can be visualized and manually selected in the Model Selection Primary node for further processing or saving to the PMM-Lab DB.

The PMM-Lab database (PMM-Lab DB) is an integral component of PMM-Lab. It comprises of a set of relational database tables specifically designed to capture all information related to PMMs, e.g. references, experimental data, formulas, model parameter estimates etc. Database tables are grouped into four main categories: System tables, Basic tables, Tenacity, PMMs (see Figure 2).
A separate GUI has been developed to access and edit the database tables (accessible via the menu option "PMM-Lab -> Open DB Gui…"). Additionally it is possible to create several distinct local database instances (via the menu option "PMM-Lab -> DB Settings") and access these databases independently from within the workflow, if desired.
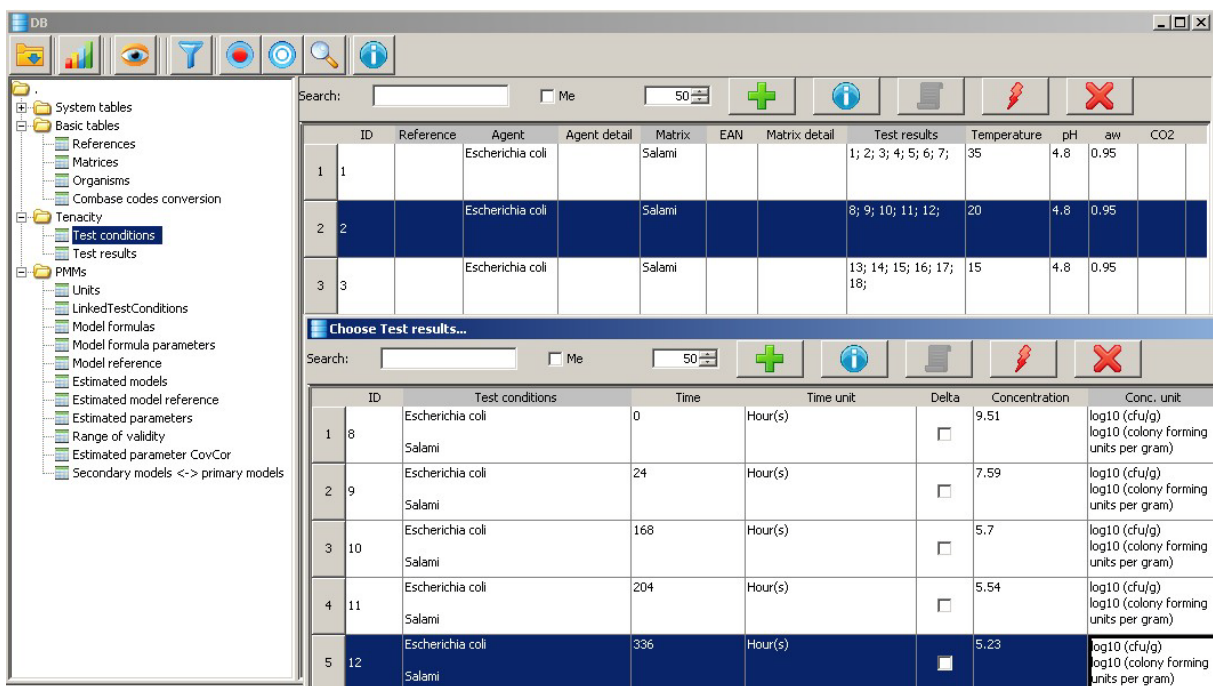

Figure 2:
Screenshot of the PMM-Lab DB GUI displaying
a) database tables (left pane),
b) 3 sample datasets from table "Test conditions" (upper right pane),
c) entries of the table "Test results" linked with dataset ID:2 in table "Test conditions" (pane "Choose Test results…")

Use case 1: Integration of data from heterogeneous information sources
The integration of information from different independent sources is often a first major hurdle in many predictive microbial modelling tasks. PMM-Lab provides means to effectively accomplish this task in a reproducible and transparent way. As illustrated in Figure 3 for experimental microbial data the user can easily
- reading in datasets downloaded from the ComBase online database via the ComBase Reader node
- enter data and experimental details manually via the Microbial Data Creator node
- reading in datasets from Excel files via the XLS Microbial Data Reader node.

All data processing or data modification steps performed within a workflow, e.g. those in the Microbial Data Editor node, can be saved as part of the workflow. In this way the "evolution" of a dataset can be documented transparently for all entries in the PMM-Lab DB.
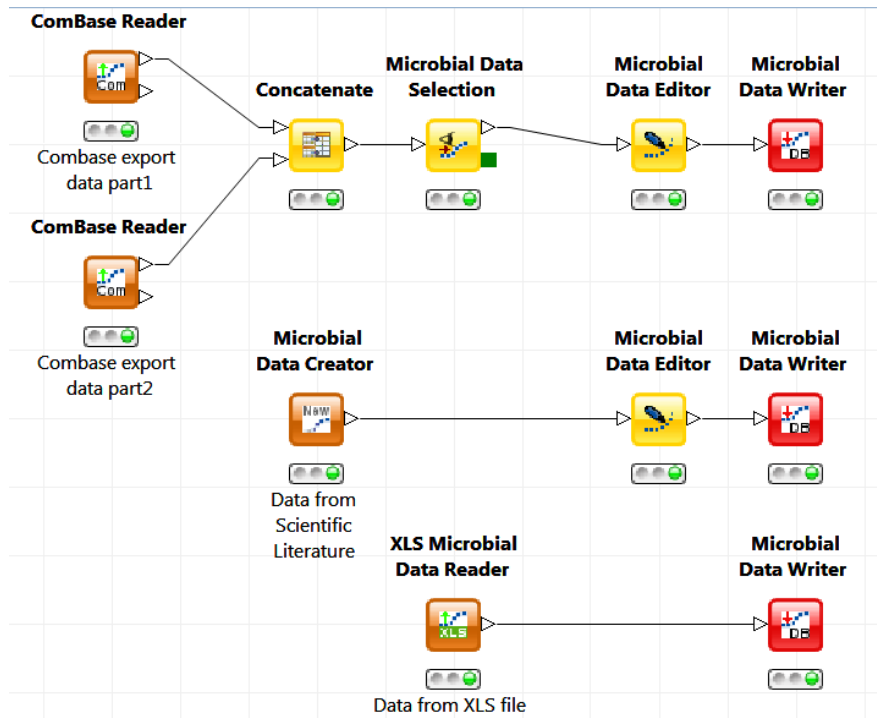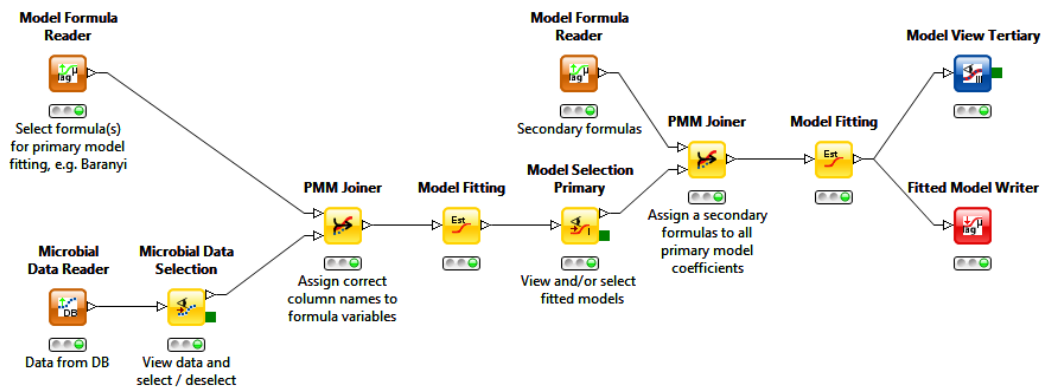
Figure 3:
A workflow that reads in data from three different sources and saves them into the user's PMM-Lab DB

Use case 2: Generation of new PMMs

The development of new predictive microbial models on the basis of experimental data is a challenging task due to the complex nature of the underlying biological systems. Additionally, technical difficulties may result from the heterogeneity and incompatibility of the modelling software applied in the community. PMM-Lab provides support for both issues due to its very flexible and modular structure which can effectively be applied to accomplish even complex modelling tasks as illustrated in Figure 4.
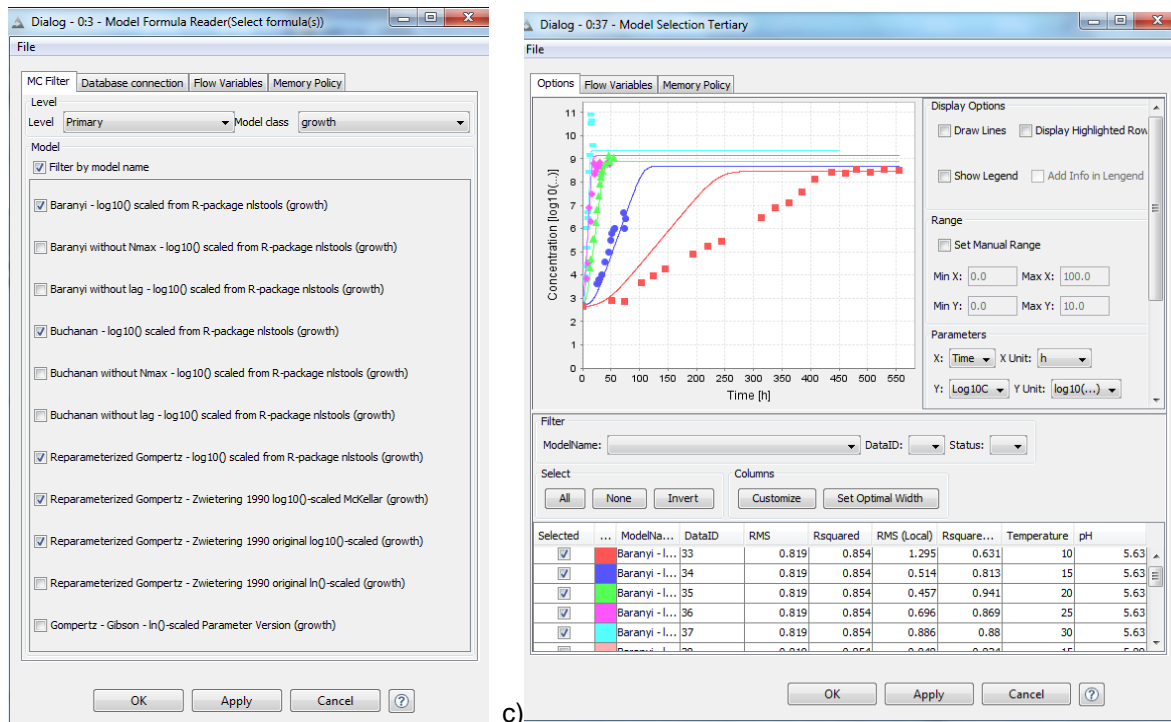

a)

Figure 4:

Screenshot of a PMM generation workflow and some of the associated node configuration menus.
a) A workflow that creates a secondary model based on primary models fitted to experimental data from the PMM-Lab DB, b) the configuration menu of the Model Formula Reader node – allowing users to specify formulas to be used in subsequent model fitting process (in this case primary model generation), c) the configuration menu of the Model Selection Tertiary node – allowing users to create charts and statistics on the generated secondary model.

Use case 3: Sharing PMMs and workflows

A key motivation in this research has been to support the exchange of PMMs including the workflows and data associated to them. To accomplish this, PMM-Lab makes use of built-in KNIME features: workflows including all associated data can be saved as .zip files ("File -> Export KNIME workflow…"). These files can then be distributed to other researchers e.g. vial email or as attachment to a scientific publication. Interested researchers in turn can easily import these workflows ("File -> Import KNIME workflow…") including the contained data into their own local PMM-Lab software installation. As PMM-Lab is freely available, free of charge and platform independent, this approach should be applicable without any constraints by anyone.

4. Summary:

PMM-Lab is a new free and open-source community resource that extends the KNIME data analysis platform with specific features relevant for predictive microbial modelling. Beside these specific features the full potential of the PMM-Lab solution unfolds in the interaction with other KNIME node libraries that are freely available as well, e.g. "Statistics", "Mining", "R", "Community Nodes" or "Reporting". Moreover, because of its modular structure the PMM-Lab software has the potential to serve as a data exchange and integration platform for other software solutions available in the field. It is, for example, easily possible to extend the PMM-Lab node library with specific import / export nodes for third party software as it has been done for the ComBase database system. The PMM-Lab software is designed as a community resource, so any direct contributions, comments, bug reports or feature requests are highly welcomed and appreciated.

Reference List

1. Gonzales-Barron UA (2011) Predictive Microbial Modelling. In: Handbook of Food Safety Engineering. Wiley-Blackwell. pp. 108-152.

2. Baranyi J, Tamplin ML (2004) ComBase: A common database on microbial responses to food environments. Journal of Food Protection 67: 1967-1971.

3. Lebert I, Lebert A (2006) Quantitative prediction of microbial behaviour during food processing using an integrated modelling approach: a review. Int J Refrid 29: 968-984.

4. Munoz-Cuevas M, Metris A, Baranyi J (2012) Predictive modelling of Salmonella: From cell cycle measurements to e-models. Food Research International 45: 852-862.

5. Coroller L, Kan-King-Yu D, Leguerinel I, Mafart P, Membre JM (2012) Modelling of growth, growth/no-growth interface and nonthermal inactivation areas of Listeria in foods. Int J Food Microbiol 152: 139-152. S0168-1605(11)00574-5 [pii];10.1016/j.ijfoodmicro.2011.09.023 [doi].

6. Brul, S., Gerwen, S. van, and Zwietering, M. H. (2007) Modelling microorganisms in food. Cambridge, UK: Woodhead Publishing Limited.

7. Berthold MR, Cebron N, Dill F, Fatta GD, Gabriel TR, Georg F, Meinl T, Ohl P, Sieb C, Wiswedel B (2006) KNIME: The Konstanz Information Miner. 58-61.