Presentation of final solution
– LifeWatch showcase

DA P6 Climate Change Data Challenge Application:
Göteborg University, Department of Marine Sciences
Title
Biodiversity Virtual Laboratory – A platform for integrating disparate data and predicting ecological responses to climate change

Authors
Matthias Obst\textsuperscript{1}, Gunnar Anderson\textsuperscript{2}, Alan Williams\textsuperscript{3}, Francisco Quevedo Fernandez\textsuperscript{4}, Anders Lindström\textsuperscript{2}, Finn Bacall\textsuperscript{3}, Alex Hardisty\textsuperscript{4}

Affiliations
\textsuperscript{1}Göteborg University, Department of Marine Sciences; \textsuperscript{2}Swedish National Veterinary institute; \textsuperscript{3}School of Computer Science, University of Manchester, Manchester, UK; \textsuperscript{4}School of Computer Science and Informatics, Cardiff University, Queens Buildings, 5 The Parade, Cardiff CF24 3AA, UK
Demonstration of the BioVeL Infrastructure

In the following presentation we demonstrate the seamless integration of the CORINE land cover dataset into BioVeL’s ecological modelling environment.

The BioVeL virtual laboratory (www.biovel.eu) is a platform for executing Taverna workflows. These workflows orchestrate complex and distributed tasks for data access and analysis.

In our demonstration we use workflow technology to connect the CORINE data set with climatic data and carry out scalable and semi-automated analytical cycles to predict the spread of mosquito-borne diseases under various climate scenarios.

The demonstration features 3 principal steps:
Step 1: The CORINE dataset was customized to the needs of the scientists
Step 2: The customized data set was made accessible in the workflow environment
Step 3: A number of predictive modelling experiments were performed using the CORINE and climatic datasets
Step 1: Customizing the dataset

The CORINE land cover dataset was transformed according to the needs of the scientists. This resulted in 44 individual data layers (see below) with continuous variables for the various land types. All transformation was done using ArcGIS.

<table>
<thead>
<tr>
<th>Layer no.</th>
<th>LABEL1</th>
<th>LABEL2</th>
<th>LABEL3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Artificial surfaces</td>
<td>Urban fabric</td>
<td>Continuous urban fabric</td>
</tr>
<tr>
<td>2</td>
<td>Artificial surfaces</td>
<td>Urban fabric</td>
<td>Discontinuous urban fabric</td>
</tr>
<tr>
<td>3</td>
<td>Artificial surfaces</td>
<td>Industrial, commercial and transport units</td>
<td>Industrial or commercial units</td>
</tr>
<tr>
<td>4</td>
<td>Artificial surfaces</td>
<td>Industrial, commercial and transport units</td>
<td>Road and rail networks and associated land</td>
</tr>
<tr>
<td>5</td>
<td>Artificial surfaces</td>
<td>Industrial, commercial and transport units</td>
<td>Port areas</td>
</tr>
<tr>
<td>6</td>
<td>Artificial surfaces</td>
<td>Industrial, commercial and transport units</td>
<td>Airports</td>
</tr>
<tr>
<td>7</td>
<td>Artificial surfaces</td>
<td>Mine, dump and construction sites</td>
<td>Mineral extraction sites</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>44</td>
<td>Water bodies</td>
<td>Marine waters</td>
<td>Sea and ocean</td>
</tr>
</tbody>
</table>

Overview over 44 data layers with information on land use types created from the CORINE land cover data set
Step 1: Customizing the dataset

Mosquitos can be stationary or mobile. Stationary species have an access radius to surrounding habitats of max 300m, while mobile species have an access radius of max 3000 m.

For each grid cell we calculated the proportion of the land type that is accessible in a 300m as well as 3000 m radius. This information was incorporated into the grid cells in each of the 44 data layers.

Although these data layers are now highly customized, they still have generic value. The 44 variables can now be used by other scientists who wish to investigate correlations between land types and abundance of terrestrial species (with similar habitat access range) across the European continent.

Example. Data layer with variable no. 27, as extracted and transformed from the CORINE land use data set. The map shows continuous values describing access to moors and heathland with a mobility range of 300m across the Scandinavian shield.
Step 2: Making the dataset accessible

Because of its generic value, we made the data set publicly available in the BioVeL modelling environment.

Both sets of layers with 300m and 3000m access range to the 44 different land types were deposited at one of BioVeL’s service providers in Brazil. For details please visit https://www.biodiversitycatalogue.org/services/37

Now the layers are seamlessly accessible through the Ecological Niche Modelling workflow (ENM). For details on the workflow, please visit http://purl.ox.ac.uk/workflow/myexp-3355.20

The ENM workflow can now be executed on the BioVeL portal (http://portal.biovel.eu) and has access to the new layers

If you want to try yourself, follow the guidelines starting on the next slide...
Step 3: Run modelling experiments with the CORINE data set

1. Register and login at http://portal.biovel.eu

2. Choose ‘Ecological niche modelling workflow’
Step 3: Run modelling experiments with the CORINE data set

3. Find the tutorials and documentation
4. Find the original Taverna workflow
5. Start the workflow

This workflow takes as input a file containing species occurrence points to create a model with the openModeller Web Service. Algorithm, environmental layers and mask are selected during the workflow. The model is tested (internal test and optional cross validation external test) and then projected one or more times. All points from the input file are used to create a single model, even if there are differences in the scientific names. Cross validation calculates the mean AUC. Model projections can be downloaded from the links in the workflow output. They are geotiff files with suitability values ranging from 0 to 255 (nadata=255).

For more information about the input file format, please check the documentation of the corresponding parameter. The default occurrence points are from a marine species called Gammarus tigrinus, so it is necessary to choose marine environmental layers during the modelling procedure to use it.

Workflow requirements: When running on Taverna workbench, this workflow requires Internet connection and workbench version >= 2.5.

Please note that ecological niche modelling experiments can take a long time to run depending on the parameters - sometimes several hours. This may happen with high resolution environmental layers, thousands of occurrence points and heavy algorithms, such as ANN and GARP BS. Cancelling a workflow run may not cancel the corresponding job on the server side, so if this procedure is repeated the server may get overloaded.

More information and documentation about this workflow can be found here: https://wiki.biovel.eu/display/doc/Ecological+Niche+Modelling+%28EN%29+Workflow

Related runs:
- Ecological niche modelling workflow sweep
- Ecological niche modelling workflow sweep
  - 5 layers (meanTemp, MeanPrec, 3 forest) - low res - PILOT
  - 5 layers (low res: Temp, Prec, highres: 3 forest layers)
- Ecological niche modelling workflow sweep
- Ecological niche modelling workflow (v5) run 30 Aug 2015 07:52:37 UTC
- Ecological niche modelling workflow (v5) run 30 Aug 2015 06:57:12 UTC
- Ecological niche modelling workflow (v5) run 30 Aug 2015 06:53:28 UTC
- Ecological niche modelling workflow (v5) run 30 Aug 2015 06:48:28 UTC
- Ecological niche modelling workflow (v5) run 29 Aug 2015
Step 3: Run modelling experiments with the CORINE data set

The ecological niche modelling (ENM) will guide you through a serious of interactions where the algorithm, parameters, environmental layers, and geographical masks are selected. An example input file is provided by default. Your model is tested and projected under the climate scenarios that you select. The model projection outputs are geotiff files with habitat suitability values.

Try yourself following the tutorial at http://tinyurl.com/ng59nv6

Ecological niche modelling workflow as developed for climate studies. For details see http://dx.doi.org/10.1111/jbi.12395
Step 3: Run modelling experiments with the CORINE data set

In step 10 of the tutorial (http://tinyurl.com/ng59nv6) you will be able to select the CORINE data (a) together with climatic and environmental data (b).

Various climate scenarios can be selected by choosing climate layers for 2050, 2100, etc.

*the layer names are currently replaced to identify the variables, e.g. 'corine_land_cover_original_r300m_code_14' will be named 'Corine14_Rice fields' by 1st of September
Step 3: Run modelling experiments with the CORINE data set

As an example we created and projected a simple model using CORINE and climate data for one mosquito species (*Aedes cantans*). The experiment is documented here: [http://portal.biovel.eu/runs/9868](http://portal.biovel.eu/runs/9868)

To see the model projections, click on ‘BioSTIF_link’ and activate the projections on the left hand panel in the upcoming GIS interface.

*Predicted changes (present-2050) in habitat suitability for the disease carrying mosquito *Aedes cantans* as calculated from the ENM workflow using CORINE and climatic data*
Step 3: Run modelling experiments with the CORINE data set

Once the parameters are fixed, the analytical cycles can be executed for a large number of species using the ’Data sweep’ function.
Step 3: Run modelling experiments with the CORINE data set

As an example we created a simple Sweep analysis (i.e. batch run execution) for 4 species (Aedes cantans, A.cinereus, A. communis, and Coquillettidia richiardii). The experiment is documented here: https://portal.biovel.eu/sweeps/238
A fully documented analysis will be published as a Research Object (www.researchobject.org) in the Zenodo online repository (https://zenodo.org) once the scientific study is completed. Here is an example of an earlier LifeWatch use case using a marine data set: https://zenodo.org/record/14839#.VeQ4Fc5FGbg
Acknowledgements and financial disclosure

The analysis as well as underlying tool development was supported by the EU’s Seventh Framework Program project BioVeL (www.biovel.eu, grant no. 283359) and by the Swedish Research Council through the Swedish LifeWatch project (http://swedishlifewatch.se, grant no. 829-2009-6278)