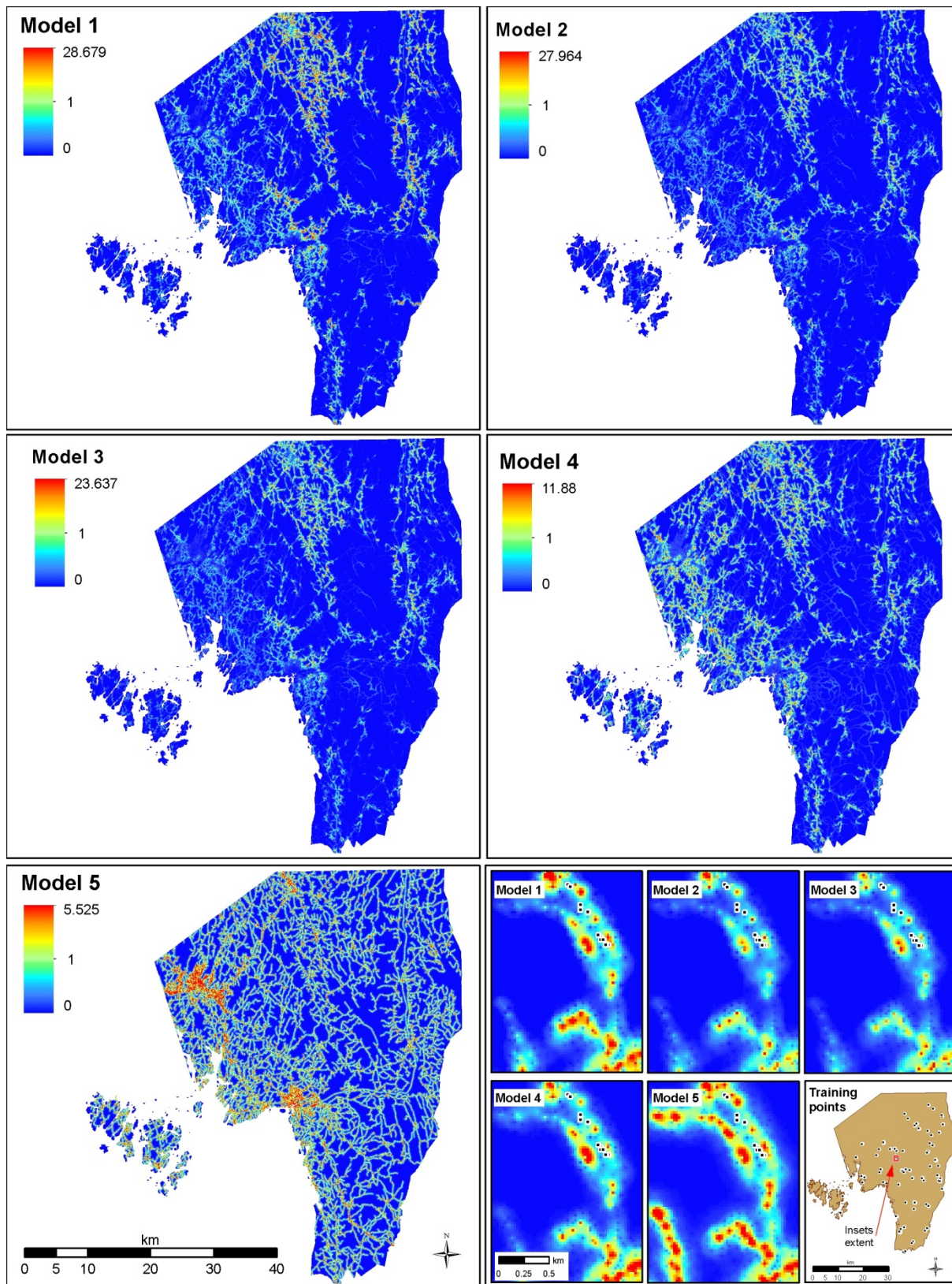


Fig. A4.1. Map representations of predictions from five MaxEnt models for *Scorzonera humilis* in SE Østfold, SE Norway, using the PRO (probability-ratio output). The five models are: Model 1 = the model with 1 EV (DIR); Model 2 = the model with 2 EVs (DIR + DIA); Model 3 = the model with 3 EVs (DIR + DIA + ALT); and Model 4 = the model with 4 EVs (DIR + DIA + ALT + SOE); Model 5 = the final model with 4 EVs and interactions between three pairs of EVs. Insert maps show mapped predictions for the small area.



## Appendix 5: R code for selected procedures, applied to the example data set

The R code included the file *MaxEntForwSel.R* is an extract of a series of R scripts (R Development Core Team 2013) and functions under development (S. Mazzone & R. Halvorsen, in prep.), for implementing the alternative procedures laid out by Halvorsen (2013), of which several are explained in the present paper. The *MaxEntForwSel.R* file includes R code for the forward stepwise variable selection procedure, parameterised for application with our example data (the *MaxEntForwSelScorzHum.Rdata* file provided). The data file contains all objects necessary to inspect the example data set and perform forward stepwise variable selection.

The script included makes use of two main input objects (as well as other, ‘minor’ objects), provided by the *MaxEntForwSelScorzHum.Rdata* file: a ‘starting’ parameters object (MIAPar), and a ‘starting’ data object (M3\_N\_SWD). These two objects will be used to produce the main output objects for this particular stage of the modelling process, namely the ‘best set of’ in the form of both new ‘model parameter’ objects and ‘data objects’, which in turn will be used in the next stage.

These objects are in the form of R-lists, each containing several R-dataframes, and will store two kinds of information: a set of ‘parameter’ lists and a set of ‘data’ lists. The parameter lists (such as MIAPar) specify all key parameters internal to R required by the scripts themselves, statistical model parameters (such as the internal model performance assessment criterium and other specifications of the model selection procedure), modelling tools-specific parameters (such as required by Maxent.jar, Windows or R-specific options and settings), and also data-specific parameters (such as location and number and types of response and explanatory variables) that are specified throughout the process.

The R object type ‘list’ is used throughout as lists give the flexibility necessary for building an increasing trail of information, that will vary in size and dimension. Lists open for making use of the tag (names) structure that gets built along the way, to create new objects and link them, and access them accordingly. Access to ‘embedded metadata’ is then made possible at different stages of the modelling process.

The data lists are specific to each stage in the modelling process, and contains R-dataframes of rows and columns of actual data value. Thus, for example, in the case of M3\_N\_SWD, the first column refers to the response variable (a 1 or a name) and each row

refers to the number of records contained by the data. The second and third are the geographic coordinates of that location. The other columns are the respective derived variables.

Each script is structured to work individually as well as part of subsequent iteration cycles and stages, and comprises distinct but ‘re-usable’ sections of code that break down the task and produce the required outcome, which in turn will be used in other sections of the process.

Script parts (concepts and grouped lines of code) are, as far as possible, ‘recycled’ across each script or set of scripts.

The scripts contain the following main sections, that build, run, parameterise and rank MaxEnt models during the forward stepwise variable selection procedure:

1. Parameter set-up and initialisation section:
  - a. Preparation of R-specific objects (lists) to be used throughout the modelling process; stores both data and parameters accordingly
  - b. ‘Initialization’ of the input parameters provided, both in the MIAPar, as well as according to the parameters of the Input Data
2. Section specific ‘commented out Inspection codes’
  - a. These are simply a series of R-functions such as *str(r-object name)*, or *head(r-object name)* that enable the user to inspect the process along the way. They have an # in front and are, accordingly, not run unless actively enabled by the user by removing the #. They are provided both for convenience and as guidance. Users are urged to temporarily remove the # to explore the process and the data along the way.
3. Preparing the appropriate set of derived variables to use in the null model comparison
  - a. The first iteration uses the full set of derived variables prepared (DVs) for each individual explanatory variable (EV), in which each DV is individually tested.
  - b. In the next iteration, two DV are compared, etc..
4. Preparing the maxent.jar batch file, which in turns runs maxent.jar.

For each model iteration, the script creates:

- a. the appropriate ‘model run’ directory for each iteration;
- b. writes out as .csv (SWD) files, the selected DV and response variable to train the model in this working directory;
- c. creates the results output folder
- d. saves all this information in the form of a windows batch file (.bat) as recommended by Phillips in his MaxEnt tutorial (Phillips 2011); then

- e. runs this bat file (which in turn runs the Maxent.jar software), and waits for it to finish before, starting again on the next iteration.
5. Storing key information for each MaxEnt model parameterised in a ‘model parameter tracking table’, provided in the form of a dataframe. This dataframe contains the starting parameters, new information from the modelling results, the results of applying the model performance assessment criteria. In this script also the stepwise subset selection takes place; information about selected variables and models are saved to a new ‘selected tracking table’, which in turn is used to perform the next cycle of iterations. This section contains:
    - a. loading maxent.jar results and reading them into the appropriate parameter object;
    - b. model subset selection and parameterisation;
    - c. applying the model selection criteria and obtaining the resulting models; and
    - d. selection and saving of appropriate results in the appropriate R list previously initialised
  6. Series of code that cleans up unnecessary objects and saves the entire session or just portions of it in .Rdata format files which can be used as input to the next phases.

The scripts are extensively commented. The comment lines are meant to guide the process for interested users, and to provide metadata for the modelling process.

To run the script with the test data provided, copy the *MaxEntForwSel.R* and *MaxEntForwSelScorzHum.Rdata* files as well as the executable Maxent software file, *Maxent.jar*) into one and the same working directory. The *Maxent.jar* needs to be placed in the working directory for the script to work without prior editing. The script is currently coded to run on Windows machines, as it uses Windows batch (.bat) files and makes use of “\” to mark directories. Make sure you have a current version of java.

Open the R.script provided, select all lines of code and source. You will see one pop-up window asking for the location of the working directory, i.e., the directory in which you’ve placed the data and files. After that, just sit back and watch!

## Appendix references

- Akaike, H. 1973. Information theory as an extension of the maximum likelihood principle. – In: Petrov, B. N. and Csaki, F. (eds), Second international symposium on information theory, Akademiai Kiado, pp. 267–281.
- Della Pietra, S. et al. 1997. Inducing features of random fields. – IEEE Trans. Pattern Anal. Mach. Intell. 19: 1–13.
- Dudík, M. and Phillips, S. J. 2009. Generative and discriminative learning with unknown labeling bias. – Adv. neural Inf. Process. Syst. 21: 401–408.
- Dudík, M. et al. 2007. Maximum entropy density estimation with generalized regularization and an application to species distribution modeling. – J. Machine Learning Res. 8: 1217–1260.
- Edwardsen, A. et al. 2011. A fine-grained spatial prediction model for the red-listed vascular plant *Scorzonera humilis*. – Nord. J. Bot. 29: 495–504.
- Elith, J. et al. 2011. A statistical explanation of MaxEnt for ecologists. – Divers. Distrib. 17: 43–57.
- Halvorsen, R. 2012. A gradient analytic perspective on distribution modelling. – Sommerfeltia 35: 1–165.
- Halvorsen, R. 2013. A strict maximum likelihood explanation of MaxEnt, and some implications for distribution modelling. – Sommerfeltia 36: 1–132.
- Hastie, T. et al. 2009. The elements of statistical learning: data mining, inference, and prediction. – Springer.
- Jaynes, E. T. 1957a. Information theory and statistical mechanics. – Phys. Rev. 106: 620–630.
- Jaynes, E. T. 1957b. Information theory and statistical mechanics. II. – Phys. Rev. 108: 171–190.
- Kullback, S. 1959. Information theory and statistics. – Wiley.
- Myers, R. H. et al. 2002. Generalized linear models with applications in engineering and the sciences. – Wiley.
- Phillips, S. J. and Dudík, M. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. – Ecography 31: 161–175.
- Phillips, S. J. et al. 2004. A maximum entropy approach to species distribution modeling. – In: Anonymous (ed), Proceedings of the 21st international conference on machine learning, ACM Press, pp. 655–662.

- Phillips, S. J. et al. 2006. Maximum entropy modeling of species geographic distributions. – *Ecol. Modelling* 190: 231–259.
- R Development Core Team 2013. R: a language for statistical computing, Version 3.0.2. – Available from <http://cran.r-project.org>, R Foundation for Statistical Computing.
- Schwarz, G. 1978. Estimating the dimension of a model. – *Ann. Statist.* 6: 461–464.
- Shipley, B. 2010. Community assembly, natural selection and maximum entropy models. – *Oikos* 119: 604–609.
- Sokal, R. R. & Rohlf, F. J. 1995. *Biometry*, ed. 3. – Freeman.
- Zuur, A. F. et al. 2007. *Analysing ecological data*. – Springer.