

Fitting Polynomial Curves with Genetic Algorithms: A State of the Art Review

Search

Google: polynomial genetic (regression OR fitting)

1. [Optimal Sampling of Genetic Algorithms on Polynomial Regression](#) (2008)

This paper investigates the utility of sampling as an evaluation-relaxation technique in genetic algorithms (GAs). In many real-world applications, sampling can be used to generate a less accurate, but computationally inexpensive fitness evaluator to speed GAs up. This paper focuses on the problem of polynomial regression as an example of problems with positive dependency among genes. Via statistical analysis of the noise introduced by sampling, this paper develops facet-wise models for the optimal sampling size, and these models are empirically verified. The results show that when the population is sized properly, small sampling sizes are preferred for most applications. When a fixed population size is adopted, which is usually the case in real-world applications, an optimal sampling size exists. If the sampling size is too small, the sampling noise increases, and GAs would perform poorly because of an insufficiently large population. If the sampling size is too large, the GA would spend too much time in fitness calculation and cannot perform well either within limited run duration.

2. **Paywalled** [Genetic polynomial regression as input selection algorithm for non-linear identification](#)

The performance of non-linear identification techniques is often determined by the appropriateness of the selected input variables and the corresponding time lags. High correlation coefficients between candidate input variables in addition to a non-linear relation with the output signal induce the need for an appropriate input selection methodology. This paper proposes a genetic polynomial regression technique to select the significant input variables for the identification of non-linear dynamic systems with multiple inputs. Statistical tools are presented to visualize and to process the results from different selection runs. The evolutionary

approach can be used for a wide range of identification techniques and only requires a minimal input and a priori knowledge from the user. The evolutionary selection algorithm has been applied on a real-world example to illustrate its performance. The engine load in a combine harvester is highly variable in time and should be kept below an allowable limit during automatic ground speed control mode. The genetic regression process has been used to select those measurement variables that have a significant impact on the engine load and that will act as measurement variables of a non-linear model-based engine load controller.

3. Paywalled [Adaptive polynomial regression for colorimetric scanner calibration using genetic algorithms](#)

Device independent colour representation is a common operation in order to assure colour consistency throughout the colour measurement and reproduction pipeline, i.e., from input to output devices. Different methods exist to convert device dependent colour spaces into device independent colour representation, usually relying on a priori fixed model structures and without imposing limitations on maximum approximation errors. This tends to lead to colour mismatches above the JND in certain regions of the device independent space. Accounting for maximal approximation errors is crucial for some colour application domains, such as in automatic visual inspection and high quality colour reproduction applications. In this paper, a novel colorimetric calibration algorithm for RGB input devices is introduced, using a polynomial modeling strategy. In order to adaptively identify the adequate model structure, i.e. polynomial order and terms, an OCV approach is followed to asymptotically approach the optimal bias-variance trade-off. Furthermore, a weighted least squares approach is applied to, indirectly, limit maximal approximation errors. These weights as well as the optimal model structure are identified using a GA. The performance of the proposed approach is compared with respect to state of the art colorimetric calibration algorithms. The achieved results show that the algorithm exhibits good generalization capabilities and that it is compatible with high quality colour reproduction for all the calibration set sizes tested. Furthermore, better colour space mappings can be obtained by the proposed method compared to the state of the art algorithms.

4. [Using Multiobjective Genetic Programming to Infer Logistic Polynomial Regression Models \[and\] Experimental Supplement](#) (2002)

In designing non-linear classifiers, there are important trade-offs to be made between predictive accuracy and model comprehensibility or

complexity. We introduce the use of Genetic Programming to generate logistic polynomial models, a relatively comprehensible non-linear parametric model; describe an efficient twostage algorithm consisting of GP structure design and Quasi-Newton coefficient setting; demonstrate that Niched Pareto Multiobjective Genetic Programming can be used to discover a range of classifiers with different complexity versus “performance” trade-offs; introduce a technique to integrate a new “ROC (Receiver Operating Characteristic) dominance” concept into the multiobjective setting; and suggest some modifications to the Niched Pareto GA for use in Genetic Programming. The technique successfully generates classifiers with diverse complexity and performance characteristics.

5. Paywalled [Optimal Polynomial Regression Models by using a Genetic Algorithm](#)

For a highly nonlinear response the accuracy of a linear or quadratic polynomial regression model is in general poor. Therefore higher order regression models might be used instead. However, a fully expanded high-order regression model cannot usually be modeled, but a model with only some of the potential high-order terms might be enough. Such a regression model is proposed here. For such a model, the determination of which high-order terms that would yield the most accurate regression model is an optimization problem in itself. The aim of this work is to solve that optimization problem in order to find the optimal polynomial regression model. Different regression models are commonly used to approximate the behavior of an unknown response in a given design domain. The regression models are usually obtained from a design of experiments, the corresponding responses and the constitution of the regression model. In this work a new approach is proposed, where the constituents of a polynomial regression model are of arbitrary order. A genetic algorithm is used to find the optimal terms to be included in the so-called optimal polynomial regression model. In the genetic algorithm the genetic operators rank selection, elitism, multi-point crossover and two methods of mutation are applied. One mutation method where single genes are changed randomly and one where segments of genes changes place within the individual is considered. One generation of the genetic algorithm consists of 100 individuals and the algorithm stops when it has reached 1000. The objective for the genetic algorithm is to minimize the sum of squared errors of the predicted responses. Thus, the fitness of the genetic algorithm is the sum of squared errors. In practice the genetic algorithm generates an optimal set of exponents (chromosome) of the design variables for the specified number of terms in the regression model, where each term is a product of a regression coefficient and the design variables. Several example problems are presented to show the performance and

accuracy of the optimal polynomial regression model. The examples consider explicit functions as well as practical engineering problems such as finite element models that are approximated using the optimal polynomial regression model. For all examples the optimal polynomial regression model is compared with ordinary regression models. Results show a greatly improved performance for optimal polynomial regression models compared to traditional regression models. Furthermore, the optimal polynomial regression model is also beneficial since its derivatives trivial. This is of great importance for subsequently applied optimization algorithms.

6. [Accelerated Genetic Programming of Polynomials](#) (2001?)

THE ABSTRACT IS NOT "PASTABLE"; THIS PAPER IS OLD BUT CONTAINS LOTS OF LINKS TO THE PRE-HISTORY

7. [Method for the identification of explicit polynomial formulae for the friction in turbulent pipe flow](#) (1999)

The paper describes a new regression method for creating polynomial models. The method combines numerical and symbolic regression. Genetic programming finds the form of polynomial expressions, and least squares optimization finds the values for the constants in the expressions. The incorporation of least squares optimization within symbolic regression is made possible by a rule-based component that algebraically transforms expressions to equivalent forms that are suitable for least squares optimization. The paper describes new operators of crossover and mutation that improve performance, and a new method for creating starting solutions that avoids the problem of under-determined functions. An example application demonstrates the trade-off between model complexity and accuracy of a set of approximator functions created for the Colebrook–White formula.

8. [A genetic algorithm approach to curve fitting](#) (1995)
9. [gapolyfitn](#) (2009 -- 2012)

This function implements a method of using genetic algorithms to optimise the form of a polynomial, i.e. reducing the number of terms required in comparison to a least-squares fit using all possible terms, as described in the following paper:
(PAYWALLED, NOT SCIENCE) Clegg, J. et al, "The use of a genetic

algorithm to optimize the functional form of a multi-dimensional polynomial fit to experimental data", 2005 IEEE Congress on Evolutionary Computation, 928-934, Edinburgh, September, 2005

10. [Polynomial Curve Fitting with Varying Real Powers](#) (2011)

In this study, the polynomial curve fitting is expanded with real powers by combining the genetic algorithm and the traditional least squares estimator. In general, integer values are used as the power of variables in traditional polynomials. However, the usage of real powers decreased the approximation error on polynomial curve fitting. In addition, the number of parameters is also decreased, when it is compared to the traditional polynomials. But, the number of adapted parameters of the proposed method is bigger than the parameters of traditional polynomials with same conditions. Although the improvements in polynomial curve fitting, the algorithm can be used for the positive input space to avoid the complex outputs. Ill. 5, bibl. 12, tabl. 4 (in English; abstracts in English and Lithuanian).

11. [A symbolic data-driven technique based on evolutionary polynomial regression](#) (2006)

This paper describes a new hybrid regression method that combines the best features of conventional numerical regression techniques with the genetic programming symbolic regression technique. The key idea is to employ an evolutionary computing methodology to search for a model of the system/process being modelled and to employ parameter estimation to obtain constants using least squares. The new technique, termed Evolutionary Polynomial Regression (EPR) overcomes shortcomings in the GP process, such as computational performance; number of evolutionary parameters to tune and complexity of the symbolic models. Similarly, it alleviates issues arising from numerical regression, including difficulties in using physical insight and over-fitting problems. This paper demonstrates that EPR is good, both in interpolating data and in scientific knowledge discovery. As an illustration, EPR is used to identify polynomial formulæ with progressively increasing levels of noise, to interpolate the Colebrook-White formula for a pipe resistance coefficient and to discover a formula for a resistance coefficient from experimental data. 02)

fitting) (2011 OR 2012 OR 2010 OR 2009)

1. [Use of genetic algorithms with multivariate regression for determination of gelatine in historic papers based on FT-IR and NIR spectral data](#) (2010)

Quantitative non-destructive analysis of individual constituents of historic rag paper is crucial for its effective preservation. In this work, we examine the potentials of mid- and near-infrared spectroscopy, however, in order to fully utilise the selectivity inherent to spectroscopic multivariate measurements, genetic algorithms were used to select spectral data derived from information-rich FT-IR or UV-vis-NIR measurements to build multivariate calibration models based on partial least squares regression, relating spectra to gelatine content in paper. A selective but laborious chromatographic method for the quantification of hydroxyproline (HYP) has been developed to provide the reference data on gelatine content. We used 9-fluorenylmethyl chloroformate (FMOC) to derivatise HYP, which was subsequently determined using reverse-phase liquid chromatographic separation and fluorimetric detection. In this process, the sample is consumed, which is why the method can only be used as a reference method.

The sampling flexibility afforded by small-size field-portable spectroscopic instrumentation combined with chemometric data analysis, represents an attractive addition to existing analytical techniques for cultural heritage materials.

2. **Paywalled** [Heat error modeling methods of NC machine tool machining holes or slots of door hardware based on genetic algorithms](#) (2011)

According to special accuracy require of holes or grooves of door hardware, the principle of NC machining of machining holes or grooves of door hardware was put forward. Heat Error is an important factor of machining accuracy of machine tool. The way of improving machine tool accuracy by using thermal deformation compensation is a necessary intelligent module of high quality NC machine tool. The heat error modeling principle and methods based on genetic algorithms for machining holes or grooves of door hardware on the NC machine tool were raised, which are the key for compensating heat error. The most optimized heat error can be obtained by error modeling based on genetic algorithms, which can compensating the heat error of machine tool for machining holes or grooves of door hardware.

3. **Paywalled** [Modeling of a liquid epoxy molding process using a particle swarm](#)

Modeling of manufacturing processes is important because it enables manufacturers to understand the process behavior and determine the optimum operating conditions of the process for a high yield, low cost and robust operation. However, existing techniques in modeling manufacturing processes cannot address the whole common issues in developing models for manufacturing processes: a) manufacturing processes are usually nonlinear in nature; b) a small amount of experimental data is only available for developing manufacturing process models; c) outliers often exist in experimental data; d) explicit models in a polynomial form are often preferred by manufacturing process engineers; and e) models with satisfactory prediction accuracy are required. In this paper, a modeling algorithm, namely, the particle swarm optimization-based fuzzy regression (PSO-FR) approach, is proposed to generate fuzzy nonlinear regression models, which seek to address all of the common issues in developing models for manufacturing processes. The PSO-FR first employs the operations of particle swarm optimization to generate the structures of the process models in nonlinear polynomial form, and then it employs a fuzzy coefficient generator to identify outliers in the original experimental data. Fuzzy coefficients of the process models are determined by the fuzzy coefficient generator in which the experimental data excluding the outliers is used. The effectiveness of the PSO-FR approach is evaluated by modeling the manufacturing process liquid epoxy molding process which is a commonly used technology for microchip encapsulation in electronic packaging. Results were compared with those based on the commonly used modeling methods. It was found that PSO-FR can achieve better goodness-of-fitness than other methods. Also, the prediction accuracy of the model developed based on the PSO-FR is better than the other methods.

4. **Paywalled** [Iterative two-step genetic-algorithm-based method for efficient polynomial B-spline surface reconstruction](#) (2012)

Surface reconstruction is a very challenging problem arising in a wide variety of applications such as CAD design, data visualization, virtual reality, medical imaging, computer animation, reverse engineering and so on. Given partial information about an unknown surface, its goal is to construct, to the extent possible, a compact representation of the surface model. In most cases, available information about the surface consists of a dense set of (either organized or scattered) 3D data points obtained by using scanner devices, a today's prevalent technology in many reverse engineering applications. In such a case, surface reconstruction consists of two main stages: (1) surface parameterization and (2) surface fitting.

Both tasks are critical in order to recover surface geometry and topology and to obtain a proper fitting to data points. They are also pretty troublesome, leading to a high-dimensional nonlinear optimization problem. In this context, present paper introduces a new method for surface reconstruction from clouds of noisy 3D data points. Our method applies the genetic algorithm paradigm iteratively to fit a given cloud of data points by using strictly polynomial B-spline surfaces. Genetic algorithms are applied in two steps: the first one determines the parametric values of data points; the later computes surface knot vectors. Then, the fitting surface is calculated by least-squares through either SVD (singular value decomposition) or LU methods. The method yields very accurate results even for surfaces with singularities, concavities, complicated shapes or nonzero genus. Six examples including open, semi-closed and closed surfaces with singular points illustrate the good performance of our approach. Our experiments show that our proposal outperforms all previous approaches in terms of accuracy and flexibility.

5. **Paywalled** [A hybrid stock selection model using genetic algorithms and support vector regression](#) (2012)

In the areas of investment research and applications, feasible quantitative models include methodologies stemming from soft computing for prediction of financial time series, multi-objective optimization of investment return and risk reduction, as well as selection of investment instruments for portfolio management based on asset ranking using a variety of input variables and historical data, etc. Among all these, stock selection has long been identified as a challenging and important task. This line of research is highly contingent upon reliable stock ranking for successful portfolio construction. Recent advances in machine learning and data mining are leading to significant opportunities to solve these problems more effectively. In this study, we aim at developing a methodology for effective stock selection using support vector regression (SVR) as well as genetic algorithms (GAs). We first employ the SVR method to generate surrogates for actual stock returns that in turn serve to provide reliable rankings of stocks. Top-ranked stocks can thus be selected to form a portfolio. On top of this model, the GA is employed for the optimization of model parameters, and feature selection to acquire optimal subsets of input variables to the SVR model. We will show that the investment returns provided by our proposed methodology significantly outperform the benchmark. Based upon these promising results, we expect this hybrid GA–SVR methodology to advance the research in soft computing for finance and provide an effective solution to stock selection in practice.

6. **Paywalled** [Development of Product Design Models Using Fuzzy Regression Based Genetic Programming](#)(2012)

To develop a functional model which relates customer requirements to the design attributes of a new product, a set of customer survey data is usually used. As mentioned in Chapter 5, customer survey data is usually fuzzy in nature, as human feeling is usually fuzzy, and also nonlinearities are unavoidable in the relationships between customer requirements and design attributes. However, the development of explicit functional models has not been addressed by previous studies in modelling the relationships between customer requirements and design attributes. Also, those previous modelling methods can address either nonlinearity or fuzziness only. To overcome the deficiencies of the above approaches, this chapter presents a fuzzy regression based genetic programming method, namely FR-GP, to generate functional models which represent this nonlinear and fuzzy relationships between customer requirements and design attributes.

7. [link A Novel hybrid genetic algorithm for kernel function and parameter optimization in support vector regression](#) (2009)

This study developed a novel model, HGA-SVR, for type of kernel function and kernel parameter value optimization in support vector regression (SVR), which is then applied to forecast the maximum electrical daily load. A novel hybrid genetic algorithm (HGA) was adapted to search for the optimal type of kernel function and kernel parameter values of SVR to increase the accuracy of SVR. The proposed model was tested at an electricity load forecasting competition announced on the EUNITE network. The results showed that the new HGA-SVR model outperforms the previous models. Specifically, the new HGA-SVR model can successfully identify the optimal type of kernel function and all the optimal values of the parameters of SVR with the lowest prediction error values in electricity load forecasting.

8. [Comparative study of response surface methodology, artificial neural network and genetic algorithms for optimization of Soybean hydration](#) (2010)

The present investigation deals with the modelling and optimization of soybean hydration for facilitating soybean processing and it focuses on maximization of mass gain, water uptake and protein retention in the bean. Process variables considered for optimization were: soybean to water ratio (1:2.48 obtained with response surface methodology, RSM, and 1.19 obtained with artificial neural network and genetic algorithm,

ANN/GA), time (2.0 h using RSM and 8.0 h using ANN/GA) and temperature (40.0 °C using RSM and 45.1 °C using ANN/GA). The findings in this first report on optimization of soaking conditions for soybean hydration employing response surface methodology, hybrid artificial neural network and genetic algorithms reveal a substantially better alternative to the time-consuming soaking process, extensively practiced in industries, in terms of process time economy. Reasonably accurate neural network model (regression coefficient of 0.9443) was obtained based on the experimental data. The optimized set of process conditions was predicted through genetic algorithm, and the effectiveness of the ANN/GA model, validated through experiments, was indicated by significant correlations (R^2 and mean squared error (MSE) being 0.9380 and 5.9299, respectively). RSM also resulted in accurate models for predicting percentage mass gain, percentage water uptake and percentage protein retention (R^2 and MSE in the range of 0.889–0.9297 and 0.80–4.94, respectively).

9. **Paywalled** [Hybrid modelling based on support vector regression with genetic algorithms in forecasting the cyanotoxins presence in the Trasona reservoir \(Northern Spain\)](#) (2013)

Cyanotoxins, a kind of poisonous substances produced by cyanobacteria, are responsible for health risks in drinking and recreational waters. As a result, anticipate its presence is a matter of importance to prevent risks. The aim of this study is to use a hybrid approach based on support vector regression (SVR) in combination with genetic algorithms (GAs), known as a genetic algorithm support vector regression (GA–SVR) model, in forecasting the cyanotoxins presence in the Trasona reservoir (Northern Spain). The GA–SVR approach is aimed at highly nonlinear biological problems with sharp peaks and the tests carried out proved its high performance. Some physical–chemical parameters have been considered along with the biological ones. The results obtained are two-fold. In the first place, the significance of each biological and physical–chemical variable on the cyanotoxins presence in the reservoir is determined with success. Finally, a predictive model able to forecast the possible presence of cyanotoxins in a short term was obtained.

10. [link](#) [Obtaining linguistic fuzzy rule-based regression models from imprecise data with multiobjective genetic algorithms](#) (2009)

Backfitting of fuzzy rules is an Iterative Rule Learning technique for obtaining the knowledge base of a fuzzy rule-based system in regression problems. It consists in fitting one fuzzy rule to the data, and replacing the

problems. It consists in fitting one fuzzy rule to the data, and replacing the whole training set by the residual of the approximation. The obtained rule is added to the knowledge base, and the process is repeated until the residual is zero, or near zero. Such a design has been extended to imprecise data for which the observation error is small. Nevertheless, when this error is moderate or high, the learning can stop early. In this kind of algorithms, the specificity of the residual might decrease when a new rule is added. There may happen that the residual grows so wide that it covers the value zero for all points (thus the algorithm stops), but we have not yet extracted all the information available in the dataset. Focusing on this problem, this paper is about datasets with medium to high discrepancies between the observed and the actual values of the variables, such as those containing missing values and coarsely discretized data. We will show that the quality of the iterative learning degrades in this kind of problems, because it does not make full use of all the available information. As an alternative to sequentially obtaining rules, we propose a new multiobjective Genetic Cooperative Competitive Learning (GCCL) algorithm. In our approach, each individual in the population codifies one rule, which competes in the population in terms of maximum coverage and fitting, while the individuals in the population cooperate to form the knowledge base.

arXiv all:((polynomial AND genetic) AND (regression OR fitting)),

2009 -- 2013

1. [Comparative analysis of model independent methods for exploring the nature of dark energy](#) (2013)

We make a comparative analysis of the various independent methods proposed in the literature for studying the nature of dark energy, using four different mocks of SnIa data. In particular, we explore a generic PCA approach, the Genetic Algorithms, a series of approximations like Padé power law approximants, and various expansions in orthogonal polynomials, as well as cosmography, and compare them with the usual fit to Λ CDM. We find that, depending on the mock data, some methods are more efficient than others at distinguishing the underlying model, although there is no universally better method.

2. [Genetic Exponentially Fitted Method for Solving Multi-dimensional Drift-diffusion Equations](#) (2013)

A general approach was proposed in this article to develop high-order

A general approach was proposed in this article to develop high-order exponentially fitted basis functions for finite element approximations of multi-dimensional drift-diffusion equations for modeling biomolecular electrodiffusion processes. Such methods are highly desirable for achieving numerical stability and efficiency. We found that by utilizing the one-one correspondence between continuous piecewise polynomial space of degree $k+1$ and the divergence-free vector space of degree k , one can construct high-order 2-D exponentially fitted basis functions that are strictly interpolative at a selected node set but are discontinuous on edges in general, spanning nonconforming finite element spaces. First order convergence was proved for the methods constructed from divergence-free Raviart-Thomas space RT_0 at two different node sets

3. [A Unified Framework for Trees, Multi-Dimensional Scaling and Planar Graphs](#) (2010)

Least squares trees, multi-dimensional scaling and Neighbor Nets are all different and popular ways of visualizing multi-dimensional data. The method of flexi-Weighted Least Squares (fWLS) is a powerful method of fitting phylogenetic trees, when the exact form of errors is unknown. Here, both polynomial and exponential weights are used to model errors. The exact same models are implemented for multi-dimensional scaling to yield flexi-Weighted MDS, including as special cases methods such as the Sammon Stress function. Here we apply all these methods to population genetic data looking at the relationships of "Abrahams Children" encompassing Arabs and now widely dispersed populations of Jews, in relation to an African outgroup and a variety of European populations. Trees, MDS and Neighbor Nets of this data are compared within a common likelihood framework and the strengths and weaknesses of each method are explored. Because the errors in this type of data can be complex, for example, due to unexpected genetic transfer, we use a residual resampling method to assess the robustness of trees and the Neighbor Net. Despite the Neighbor Net fitting best by all criteria except BIC, its structure is ill defined following residual resampling. In contrast, fWLS trees are favored by BIC and retain considerable strong internal structure following residual resampling. This structure clearly separates various European and Middle Eastern populations, yet it is clear all of the models have errors much larger than expected by sampling variance alone.

4. [Multivariate forecast of winter monsoon rainfall in India using SST anomaly as a predictor: Neurocomputing and statistical approaches](#) (2009)

In this paper, the complexities in the relationship between rainfall and sea surface temperature (SST) anomalies during the winter monsoon

surface temperature (SST) anomalies during the winter monsoon (November-January) over India were evaluated statistically using scatter plot matrices and autocorrelation functions. Linear as well as polynomial trend equations were obtained and it was observed that the coefficient of determination for the linear trend was very low and it remained low even when polynomial trend of degree six was used. An exponential regression equation and an artificial neural network with extensive variable selection were generated to forecast the average winter monsoon rainfall of a given year using the rainfall amounts and the sea surface temperature anomalies in the winter monsoon months of the previous year as predictors. The regression coefficients for the multiple exponential regression equation were generated using Levenberg-Marquardt algorithm. The artificial neural network was generated in the form of a multiplayer perceptron with sigmoid non-linearity and genetic-algorithm based variable selection. Both of the predictive models were judged statistically using the Willmott index, percentage error of prediction, and prediction yields. The statistical assessment revealed the potential of artificial neural network over exponential regression.

arXiv all:((algorithm AND genetic) AND (regression OR fitting)),

2009 -- 2013

Yields 98 results, too many to include here. Most report applications, some are directed to other non-linear models, a few are repeated and two are concerned with **bayesian inference**:

1. [Evolutionary Stochastic Search for Bayesian model exploration](#)(2010)

Implementing Bayesian variable selection for linear Gaussian regression models for analysing high dimensional data sets is of current interest in many fields. In order to make such analysis operational, we propose a new sampling algorithm based upon Evolutionary Monte Carlo and designed to work under the "large p, small n" paradigm, thus making fully Bayesian multivariate analysis feasible, for example, in genetics/genomics experiments. Two real data examples in genomics are presented, demonstrating the performance of the algorithm in a space of up to 10,000 covariates. Finally the methodology is compared with a recently proposed search algorithms in an extensive simulation study.

2. [abc: an R package for Approximate Bayesian Computation \(ABC\)](#) (2011)

Many recent statistical applications involve inference under complex

models, where it is computationally prohibitive to calculate likelihoods but possible to simulate data. Approximate Bayesian Computation (ABC) is devoted to these complex models because it bypasses evaluations of the likelihood function using comparisons between observed and simulated summary statistics. We introduce the R abc package that implements several ABC algorithms for performing parameter estimation and model selection. In particular, the recently developed non-linear heteroscedastic regression methods for ABC are implemented. The abc package also includes a cross-validation tool for measuring the accuracy of ABC estimates, and to calculate the misclassification probabilities when performing model selection. The main functions are accompanied by appropriate summary and plotting tools. Considering an example of demographic inference with population genetics data, we show the potential of the R package. R is already widely used in bioinformatics and several fields of biology. The R abc package will make the ABC algorithms available to the large number of R users. abc is a freely available R package under the GPL license, and it can be downloaded at [this http URL](#)