Teaching Wavelets in XploRe[1]

Sigbert Klinke[2], Yuri Golubev[3], Wolfgang Härdle[2] and Michael H. Neumann[4]

[2] Humboldt University of Berlin, Department of Economics, Institute of Statistics and Econometrics, Spandauer Strasse 1, D-10178 Berlin, Germany
[3] Institute for Problems of Information Transmission, Bolshoi Karetny 19, Moscow, Russia
[4] Weierstrass Institute for Applied Analysis and Stochastics, Mohrenstrasse 39, D-10117 Berlin, Germany

Summary

Teachware is a set of computer software tools for computeraided interactive teaching of certain knowledge elements. The construction of teachware for statistical knowledge is a rather young field since it heavily depends on data structures and graphical interaction possibilities. In this paper we present a teachware module for XploRe - a statistical computing environment. We focus on the situation of teaching wavelets, a technique for adaptation of spatial inhomogeneity.

Keywords: teachware, wavelets, interactive HMTL

[1]The authors and the wavelet tutorial are accessible via WWW `http://wotan.wiwi.hu-berlin.de`.
Financial support was received by the Deutsche Forschungsgemeinschaft, SFB 373 ("Quantifikation und Simulation Ökonomischer Prozesse"), Humboldt-Universität zu Berlin.

1 Introduction

1.1 Teachware

Over the past decade many programs have been written and numerous platforms have been proposed for computer assisted teaching. Especially statistical teachware has been built which aimed at conveying topics of statistical science to students. Koch & Haag (1996) list in their "Statistical Software Guide 95/96" some programs which name themselves teachware or learnware. We list some of them:

PRISTAT 1 an interactive program designed for both general purpose analysis and education. It is based on the book of Kolev (1993),

SchoolStat a MacIntosh shareware program on a spreadsheet basis,

Sila a tool for teaching students how the logic of inference in statistics works.

Many more teachware programs are available for several platforms. One of the first programs, originally designed for a small home computer, is described in Bowman & Robinson (1989), Bowman & Robinson (1990). Proenca (1995) describes a teachware in XploRe for interactive linear regression and smoothing.

Many of these teachwares are based on systems like Toolbook which are more like a *weakly dynamic* textbook. We say weakly dynamic, since desirable links to the (underlying) statistical software by construction are not an element of such teachware. Thus the teacher or the student may change certain predefined program elements.

If teachware allows for external effects by the student it risks to be not a self contained and a self explained system. Such external effects are the inclusion of student or teacher written codes or the application to an external dataset. We must therefore introduce a second level where the teachware explains also the necessary handling of the underlying statistical computing environment. That makes it necessary to jump between direct software use and the explanation of how to use it. For this task we found the HTML technology the most suited one.

1.2 Wavelets

In order to demonstrate our concept, we present a tutorial on wavelets. The application of wavelet ideas to nonparametric statistics is relatively new and has drawn much attention by statisticians. Wavelets are also used in other fields like approximation theory, sound analysis and image compression. One of their basic properties is that they provide a sparse representation of smooth functions, even if the degree of smoothness varies considerably over the domain of interest of if the function is only piecewise smooth. These favorable approximation properties, which are not shared by the classical Fourier basis, lead to a superior performance of estimators of functions with spatially inhomogeneous smoothness properties compared to classical linear estimators (kernel, spline).

Introductions to wavelets and applications may be found in the books by Härdle, Kerkyacharian, Picard & Tsybakov (1997) and Kaiser (1994). A first attempt for an interactive tool for wavelet smoothing was integrated in the teachware lessons of Proenca (1995) in XploRe 3 by Klinke (1997).

2 Structure of the System

2.1 The Users View

One of the principles of teachware is the accessibility from everywhere. A student must be able to use the system in the class as well as at home. An internet link is therefore a necessity. We offer two parallel possibilities of access:

1. browsing through the wavelet tutorial HTML-files in the internet
2. javing XploRe.

By *javing* we mean the use of Java interface in XploRe. Both entries are found by a direct internet link over the XploRe system which is availabe through

`http://wotan.wiwi.hu-berlin.de/xplore/xplore4.html`

It is required that the user has both processes started. This offers the necessary flexibility and the interaction with external phenomena, e.g. with external datasets and user written codes.

Suppose that the user has either installed XploRe or is javing it in a WWW-browser. Then Figure 2 shows the situation where XploRe has been started.

The tutorial itself and the WWW-browser should be used in parallel such that the user can read the text and immediately execute the commands, see Figure 3. Besides loading the library `twave` the user has not to type anything and the whole dialog is now menu-driven. The tutorial is *strongly dynamic* since it has links to the executable XploRe codes for wavelet analysis. The

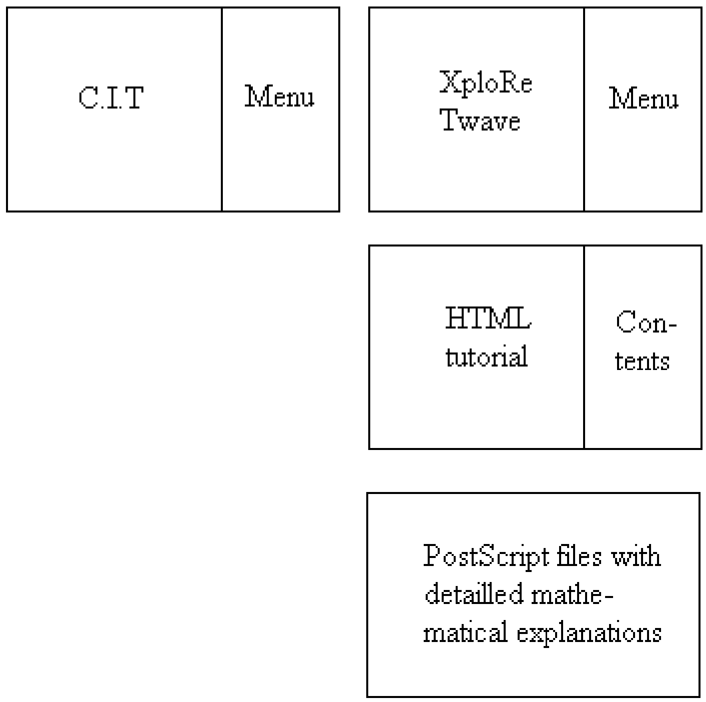

Figure 1: The Wavelet tutorial in XploRe consists of three parts: the XploRe library `twave`, the HTML tutorial and the Postscript files for detailled mathematical explanations. In contrast C.I.T just consists of the program and additional manuals.

static underlying mathematical formulation is described in a PostScript file as explained in Figure 1.

A current drawback of HTML 2.0 as a basis language is that it does not support the typesetting of formulas. However HTML 3.0 will support also formula typsetting. In practice this has a big advantage, it forces the developper to describe the properties in words rather than in formulas. This is especially important if we are teaching to unexperienced students which are not so familiar with the mathematical notation.

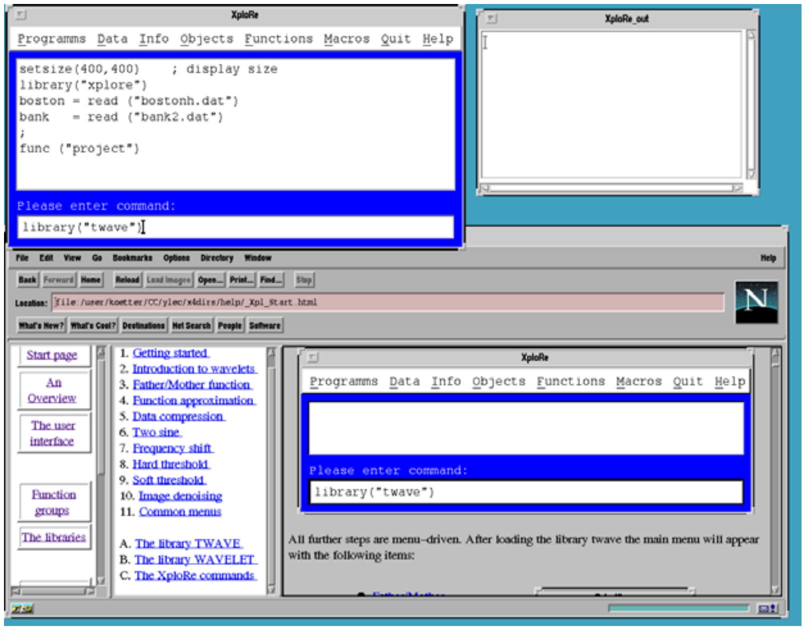

Figure 2: The screen shows in the left upper corner the XploRe command window. The window in the right upper corner is the XploRe output window. The window in the bottom of the screen is the WWW-browser (here: Netscape) with the Wavelet tutorial. The window consists of three frames: the XploRe help system frame (left), the wavelet tutorial overview (middle) and the wavelet tutorial itself (right).

2.2 The Developers View

From the developers view the whole teaching system can be decomposed into two parts:

1. Programming the single task and combining them to a system
2. Writing the HTML pages and the PostScript files

The macros are based on two commands in XploRe 4:

`fwt`	the fast wavelet transform and
`invfwt`	the inverse fast wavelet transform.

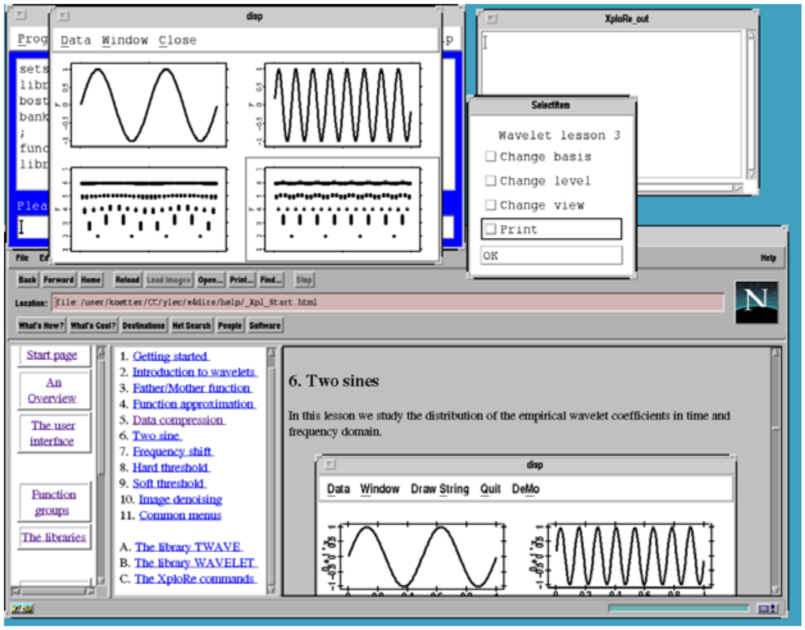

Figure 3: The screen shows in the left upper corner the XploRe command window overlayed by a display of the "two sines" lesson of **twave** library. The WWW-browser in the bottom of the screen shows the "two sines" lesson in the browser. In this lesson we look at the approximation of two sines with different frequencies by different mother wavelet coefficients. Whereas the browser shows the initial picture with the two sine functions and an approximation with the Daubechies-4 basis, we have changed already in XploRe to the Haar basis. In the right upper corner we see the XploRe output window and the actual menu for this lesson.

The library **wavelet** computes the constants necessary for the different wavelet bases. In the teaching system the Haar basis, the Daubechies-4 and the Symmlet-4 are used. By offering only these three bases we create the incentive to learn more about other bases, e.g. by trying them or reading about them in the underlying static information, see Figure 1.

Analoguous to the generation of the wavelet bases in **waveletmain** when the library **wavelet** is called, the library **twave** executes during loading the macro **twavemain** which starts the teaching system. **twavemain** calls a macro named **twlesson** which either interactively starts a wavelet lesson (call **twlesson(NaN)**) or a specific lesson (call **twlesson(NaN|i)** with $i = 1, ..., 8$).

`twlesson` displays in an endless loop the basic menu which allows to select one or more lesson for execution. Then one or more of the macros `twles1`, ..., `twles8` is executed which itself displays a submenu which allows different manipulations.

In nearly each lesson a display is shown with two vertical windows and a lesson dependent number of horizontal windows (usally two or three). The upper window always shows a plot of the problem and the lower window a representation of the mother wavelet coefficients which are appropriate for the problem.

We provide four basic views to the wavelet coefficients:

The standard view which shows the coefficients for each resolution level in a line as a vertical bar. The height of the bar is determined by the size of the coefficient.

The ordered coefficients which shows the coefficients as the absolute size of the coefficient. Not all coefficients will be shown.

The circle coefficient shows the coefficients as in the standard view, but uses circles instead of bars. The radius of the circle depends on the absolute size of the coefficient. The circle is drawn in red if the coefficient is used in construction the wavelet function and it is drawn in blue otherwise.

The partial sum shows the approximation of the wavelet function by adding sequentially one resolution level after another.

The views are produced by the macros `waveint1` to `waveint4`.

3 Content of the System

The teaching system is composed by 8 lessons which cover the most interesting facts about wavelets.

The tutorial itself is based on 14 topics, see Figure 2.

Front page. It allows the user to jump back to the beginning of the tutorial.

Getting started. Here is described how XploRe is started and how the library `twave` has to be called.

Introduction to wavelets. It introduces the user of the system to wavelets in general.

Father/Mother function. In this lesson the student makes himself familiar with the basis functions used in the wavelet analysis. These functions are basically obtained by dyadic translations and dilations of two specific functions, a so-called scaling function and a wavelet. The obtained basis functions are called father and mother functions, respectively. One can choose between functions from three different wavelet bases, the classical Haar basis, the Daubechies 4 basis and the Coiflet 2 basis. Since we use a periodic wavelet basis, the father wavelets may look a bit different from commonly known father wavelets corresponding to a basis on the whole real line.

Function approximation. The ability of wavelet bases to provide parsimonious approximations for smooth functions is the key to a favourable behaviour of statistical estimators based on a wavelet expansion. This lesson demonstrates how certain smoothness features of a function translate into sparsity in the space of coefficients. For example, for a piecewise constant function the coefficients corresponding to the mother wavelets are equal to zero, except for those coefficients which correspond to mother wavelets whose support contains a jump point of the function.

Data compression. This lesson describes the ability to compress certain functions into a small number of significant coefficients. For certain functions, some of them with spatially inhomogeneous smoothness properties, the ordered coefficients with respect to a wavelet basis are compared with the ordered coefficients with respect to the classical Fourier basis. It can be clearly seen that the wavelet bases have a superior ability to compress functions with inhomogeneous smoothness properties into a small number of coefficients.

Two sines. In this lesson the student studies the distribution of the empirical wavelet coefficients in time and frequency domain. For two sine functions with different frequencies the coefficients are displayed with respect to their spatial position and their location in resolution scale. The power of the wavelet coefficients moves to the finer resolution scales if the frequency of the function increases.

Frequency shift. The goal of this lesson is to demonstrate that the wavelet transform reflects the properties of the signal simultaneously in frequency and time domains. We consider the signal composed from two sine waves having different frequencies on the time intervals $[0, 0.5]$ and $(0.5, 1]$, respectively. It can be again seen how the power of the wavelet coefficients moves to the finer scales when the frequency of the sine becomes higher.

Hard thresholding. In statistical applications like nonparametric regression, only noisy data about the underlying function are given. Therefore, empirical versions of the wavelet coefficients are equal to the true coefficients plus some contribution by the noise. With nonparametric wavelet methods, the smoothing operation is usually performed in the domain of coefficients. Whereas a linear downweighting of the coefficients is appropriate in the case of functions functions with spatially homogeneous smoothness properties, functions with considerably inhomogeneous smoothness properties require different, essentially nonlinear regularization methods. Quite a popular method to "denoise" data is hard thresholding, that means all coefficients which are in absolute value above a certain threshold are untouched, whereas the other coefficients are set to zero:

$$H(x) = xI(\text{abs}(x) > t),$$

where t is the threshold. In this lesson the effect of hard thresholding can be studied for different choices of the threshold parameter.

Soft thresholding. Along with hard thresholding, soft thresholding procedures are used in many statistical applications. In this lesson we study the so-called wavelet shrinkage procedure for recovering the regression function from noisy data. The only difference between the hard and the soft thresholding procedure is the choice of the nonlinear transform on the empirical wavelet coefficients. For soft thresholding the following nonlinear transform is used:

$$S(x) = \text{sign}(x)(\text{abs}(x) - t)I(\text{abs}(x) > t),$$

where t is the threshold.

Translation invariant. The stationary wavelet transform described in Coifman & Donoho (1995) and Nason & Silverman (1995) is implemented here. It is well-known that nonlinear wavelet estimators are not translation-invariant: if we shift the underlying data set by a small amount, apply nonlinear thresholding and shift the estimator back, then we usually obtain an estimator diffferent from the estimator without the shifting and backshifting operation. To get rid of this, we average over several estimators obtained by shifting, nonlinear thresholding and backshifting. By Jensen's inequality, the squared loss of this new estimator is not larger and usually smaller than the average of the squared losses of these individual estimators. In the context of spectral density estimation, Neumann (1996b) observed in simulations a considerable improvement by this method over the standard wavelet estimator.

Image denoising. As an application of higher-dimensional wavelet methodology, we also included a lesson on image denoising. It was shown in Neumann & von Sachs (1997) that the commonly used isotropic d-dimensional basis does not lead to optimal estimators if the function to be estimated has different degrees of smoothness in different directions. In contrast, estimators based on a certain anisotropic basis can attain optimal rates of convergence in such anisotropic smoothness classes as well as if the "effective dimension" of the function is smaller than its nominal dimension; cf. Neumann & von Sachs (1997) and Neumann (1996a). We applied nonlinear thresholding in a two-dimensional anisotropic wavelet basis to an image corrupted by some additive noise. A considerable improvement over the noisy image can be observed.

Common menus. describes some common menus, e.g. choice of wavelet basis, choice of function, printing.

The library TWAVE. We list the macros of the library `twave`.

The library WAVELET. We list the necessary macros of the library `wavelet`.

The XploRe commands. We describe the basic XploRe commands.

4 Future Work

Some future work may consist in the integration of more interactivity in the system. The visualization of the father and mother wavelet functions can be directed by the cursor keys instead of using a menu entry.

We need a stronger integration of the help system and tutorial with the software. We may need help buttons in the software which immediately display some help text to the actual lesson, e.g. a part of the tutorial. Vice versa we may want to start a specific lesson from the tutorial on mouseclick.

Kötter (1996) developed a JAVA-interface for XploRe. This interface allows javing XploRe from a specific lesson and to perform all necessary operations. The software itself is already prepared to allow such operations.

Another possible extension is the inclusion of further wavelet lessons. To improve the understanding of the local approximation ability of wavelets one could include an interactive manipulation of wavelet coefficients such that the student gets immediate feedback.

References

Bowman, A. & Robinson, D. (1989). C.I.T.: Introduction to Statistics, *Software*, IOP Publishing Ltd.

Bowman, A. & Robinson, D. (1990). C.I.T.: Regression & ANOVA, *Software*, IOP Publishing Ltd.

Coifman, R. & Donoho, D. (1995). Translation-invariant de-noising, *in* A. Antoniadis (ed.), *Lectures Notes in Statistics: Wavelets and Statistics*, Springer, pp. 125–150.

Härdle, W., Kerkyacharian, G., Picard, D. & Tsybakov, A. (1997). *Wavelets, approximation and statistical applications*, Springer.

Kaiser, G. (1994). *A friendly guide to wavelets*, Birkhäuser, Boston.

Klinke, S. (1997). *Data Structures in Computational Statistics*, Physica Verlag.

Koch, A. & Haag, U. (1996). The statistical software guide '95/96, *Computational Statistics & Data Analysis* **21**(2): 231–256.

Kolev, N. (1993). *Applied Statistics 1*, Stopanstvo, Sofia.

Kötter, T. (1996). *Entwicklung statistischer Software*, PhD thesis, Humboldt University of Berlin, Department of Economics, Institute of Statistics and Econometrics.

Nason, G. & Silverman, B. (1995). The stationary wavelet transform and some statistical applications, *in* A. Antoniadis (ed.), *Lectures Notes in Statistics: Wavelets and Statistics*, Springer, pp. 281–299.

Neumann, M.H. (1996a). Multivariate wavelet thresholding: a remedy against the curse of dimensionality?, *Preprint no. 229*, Weierstrass Institute, Berlin.

Neumann, M.H. (1996b). Spectral density estimation via nonlinear wavelet methods for stationary non-Gaussian time series. *J. Time Ser. Anal.* **17**: 601–633.

Multivariate wavelet thresholding: a remedy against the curse of dimensionality?, *Preprint no. 229*, Weierstrass Institute, Berlin.

Neumann, M.H. & von Sachs, R. (1997). Wavelet thresholding in anisotropic function classes and application to adaptive estimation of evolutionary spectra, *The Annals of Statistics*.

Proenca, I. (1995). Interactive graphics for teaching simple statistics, *in* W. Härdle, S. Klinke & B. Turlach (eds), *XploRe - an interactive statistical computing environment*, Springer, pp. 113–140.