1

Wavelet thresholding on non-equispaced data

Maarten Jansen

Abstract

This paper tackles the problem of non-equispaced data smoothing using the non-linear, non-parametric estimation technique of wavelet thresholding. Wavelets on irregular grids can be constructed, using the so-called lifting scheme. The lifting scheme is *grid-adaptive* in that it provides smooth basis functions on the given data set. We demonstrate however that this scheme pays little attention to the numerical condition of the transform. This bad condition originates from the fact that a straightforward application of a lifting scheme on irregular point sets mixes up scales within one level of transformation. Experiments illustrate that stability is a crucial issue in threshold or other shrinking algorithms. Unstable transforms blow up the bias in the smoothing curve. We propose a *scale-adaptive* scheme that avoids scale-mixing, leading to a smooth and close fit.

1.1 Introduction

A classical ('first generation') wavelet transform assumes the input to be a regularly sampled signal. In many applications however, data are not available on a regular grid, but rather as non-equidistant samples. Used as an input to one of the standard wavelet shrinkers, irregular point set leads to a wiggly output, since the wavelet algorithm seeks for outputs which are smooth on equidistant grids. Therefore, if the input at hand is not equidistant, the irregularity of the grid is reflected in the output curve.

Nearly all existing wavelet based regression of non-equispaced data combines a traditional equispaced algorithm for fitting with a "translation" of the input into an equispaced problem [9, 11, 7, 4, 1, 13, 12, 3, 2]. In general, preprocessing the data changes their statistics. White noise, for instance, becomes coloured, or stationarity may be lost. Also, working in two phases (pre- or post-processing, before or after the actual smoothing) may reduce the overall potential: a good result may become suboptimal after post-processing.

Therefore, we opt for an integrated approach, in which the basis functions are adapted to the irregular point set. One approach to doing this involves "squeezing" techniques, starting from (pieces of) standard wavelets [8]. We

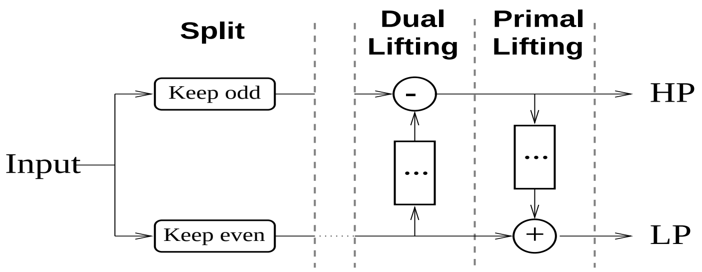


Figure 1.1. A lifting implementation of one level in a wavelet transform. Lifting operations come in three types: splitting, dual lifting and primal lifting. Just as in a classical wavelet transform, the output consists of detail, high pass (HP), or wavelet coefficients and low pass coefficients (LP), which are further processed in the following steps.

take a different way, based on the so called lifting scheme [15]. This scheme constructs a wavelet transform on a given data grid in consecutive steps, starting from a trivial transform. Every step adds new, smoothness properties. This gradual construction motivates the name *lifting*. Apart from a few publications that we know of, the use of lifting [6, 10, 16] in smoothing is quite new. The next section reviews the basic concepts of lifting and thresholding. Next we explore and analyse the stability problems following from a straightforward implementation of the lifting scheme on an arbitrary grid. We propose a relaxed scheme that overcomes this instability, leading to a smooth and close fit. Simulations and real data examples illustrate that the modified scheme even adapts automatically to the case of simultaneous observations.

1.2 Lifting and thresholding

A classical wavelet decomposition proceeds in consecutive steps — or levels — all corresponding to a particular scale. The action within one level can always be implemented using lifting [5]. The first stage in a lifting scheme is splitting the data into an even and odd sequence. This corresponds to the decimation in the classical implementation. Next in the lifting scheme is an alternating sequence of so called dual and primal lifting steps, as in Figure 1.1. This paper concentrates on decompositions with one dual and one primal lifting step. The first, dual, step can be interpreted as a *prediction* of the odd points by neighbouring evens. The difference between prediction and true value is detail information, it is the wavelet coefficient. In the Haar transform, this prediction is simply the value of the left even neighbour. Other popular predictions are linear or cubic interpolation in the evens.

Whereas the dual step computes the detail *coefficients*, the primal lifting or *update* changes the values of the low-pass coefficients. It thereby actually creates the wavelet *basis functions*. This mechanism is visualised for the Haar case in [10, Section 2.6.1]. It can also be understood by running the inverse transform. The inverse transform starting from a zero input, except for a single coefficient, converges towards the wavelet basis function associated with that coefficient. This is generally true (not only for lifted transforms) and is known as the subdivision scheme [16]. It is the classical trick to make plots of wavelet functions. Applied to lifting, we see that the first step of the inverse transform is undoing the last update (primal lifting). This update filter shows a response to the only non-zero detail (wavelet) coefficient in the HP-branch. Next, this response is being subtracted from the zero sequence in the LP-branch. Suppose this filter response has two non-zeros $A_{j,k}$ and $B_{j,k}$. We then obtain the same basis function if we omit the last update step, and run the inverse transform with a single non-zero in the HP-branch *plus* a sequence with two non-zeros, $-A_{j,k}$ and $-B_{j,k}$, in the LP-branch. As a consequence, the wavelet function $\psi_{j,k}$ at scale j and location k equals

$$\psi_{j,k} = \psi_{j,k}^{[0]} - B_{j,k}\varphi_{j,k} - A_{j,k}\varphi_{j,k+1},$$

where $\varphi_{j,k}$ is the scaling function at the same scale and location, and $\psi_{j,k}^{[0]}$ is the wavelet basis function without update step. The same trick can be used to see that this wavelet function is nothing but the odd scaling function $\varphi_{j+1,2k+1}$ at the previous, fine scale $j+1$. Indeed, a single non-zero in the HP-branch would proceed unchanged and with no effect to the LP-branch in the next subdivision step. All together, we have

$$\psi_{j,k} = \varphi_{j+1,2k+1} - B_{j,k}\varphi_{j,k} - A_{j,k}\varphi_{j,k+1}. \tag{1.1}$$

The two degrees of freedom $A_{j,k}$ and $B_{j,k}$ can be used to make the wavelet have a zero integral and one more vanishing moment. Both dual and primal lifting steps are by no means limited to equidistant point sets.

A classical wavelet transform has strong decorrelating properties. It uses the correlation between neighbouring samples to obtain a sparse representation of the noise free signal. The main part of the coefficients is close to zero and the essential information is captured by a limited number of large, important coefficients. Replacing the small coefficients with an absolute value below a certain threshold with zero reduces the noise without affecting the noise free signal too much. A central issue in this kind of smoothing procedures is how to find a suitable value for the smoothing parameter, in this case the threshold λ. This article opts for a minimum mean square error (MSE) approach. The expected MSE (also known as *risk*) is a sum of bias and variance. The variance stands for the noise: it decreases when the threshold grows. The bias on the other hand increases when the threshold grows. The minimum MSE threshold is the best trade-off between variance and bias in ℓ_2-norm sense.

Figure 1.2 has an experiment with the well known test function 'heavisine':

$$f(x) = 4\sin(4\pi x) - \text{sign}(x - 0.3) - \text{sign}(0.72 - x)$$

sampled on 2048 points. Those points were drawn from a uniform density on $[0, 1]$. Such a uniform selection does not have huge gaps in absolute terms, but observations may be arbitrarily close to each other. This implies that extra small scale phenomena are observed within a relatively wide "gap". To this signal, white and stationary noise was added with a standard deviation of $\sigma = 0.3$. Figure 1.2(middle) shows the output from

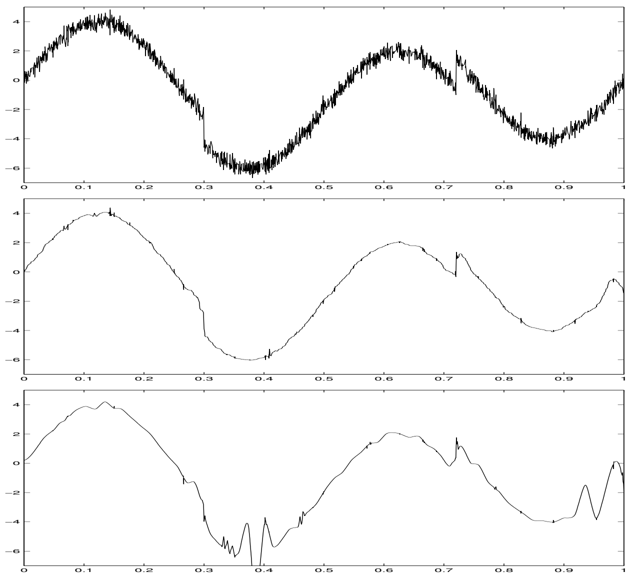


Figure 1.2. (top) A noisy version of the 'heavisine' test function. This is a piecewise continuous signal, a typical member of the class of signals for which wavelets are the optimal approach. (middle) Classical wavelets cannot take the grid structure into account. Remapping the estimation to this grid makes its irregularity show up in the result. (bottom) Second generation wavelets, based on lifting, are smooth on the actual grid but lead to a tremendously biased reconstruction.

thresholding the 5 finest resolution level of a classical wavelet transform. The wavelet transform here was constructed by lifting with cubic prediction followed by 2-taps-update, but no grid structure was taken into account. A much smoother result follows if the transform is based on the grid. Figure 1.2(bottom) however shows that the reconstruction from these second generation wavelets contains strange artifacts, or — in statistical terminology — a strange bias.

1.3 The problem: instability

A classical wavelet transform guarantees a norm semi-equivalence between the input and the wavelet coefficients: if $\boldsymbol{w}$ is the wavelet transform of $\boldsymbol{y}$, then the ℓ_2 norms of these vectors satisfy:

$$c \cdot \|\boldsymbol{w}\| \leq \|\boldsymbol{y}\| \leq C \cdot \|\boldsymbol{w}\|,$$

with $0 < c, C < \infty$ independent of the vector length. This relates to the concept of Riesz bases. Loosely speaking, a Riesz basis, also known as stable basis, is a basis in which the basis vectors or functions cannot be arbitrarily close to each other. This notion becomes important in vector spaces with infinite dimension, or, as in our case, when dealing with situations where the dimension is finite but arbitrarily large. The constants c and C are closely related to the condition number of the wavelet transform matrix.

The extension to non-equispaced data through lifting gives no guarantee for the preservation of this comfortable Riesz basis background. High condition numbers mean that a small modification of wavelet coefficient values may result in an unpredictable effect on the output. In the case of thresholding, this means that a small coefficient may carry substantial signal information. Not only do some individual coefficients have a wide impact, the interaction between the coefficients may be unpredictable, due to the fact that the transform is far from orthogonal. This bad condition can only be detected with a global analysis: basis functions my be pairwise far from each other and yet all together, they constitute an arbitrarily oblique basis.

1.4 Stable transforms with interpolating prediction

Since the instability is so hard to detect and localise, we examine which steps in the transform initiate the problems. The lifting scheme as such does not pay any particular attention to stability. If the update coefficients in (1.1) are large, say if,

$$B_{j,k} \gg \frac{\|\psi_{j,k}^{[0]}\|}{\|\varphi_{j,k}\|},$$

the lifted wavelet $\psi_{j,k}$ nearly falls within the vector space spanned by its neighbouring scaling functions at the same scale. This creates a detail space which is far from orthogonal to the coarse scaling space. When the scaling functions are further decomposed into a wavelet basis at coarser scales, the immediate correlation between basis functions becomes hidden.

The Haar transform remains orthogonal, so in that case, the update step (with a single non-zero element in its filter) has definitely a stabilising effect: without update, the transform would not be orthogonal. (The Haar transform on irregular data is sometimes called — rather confusingly — *Unbalanced Haar*)

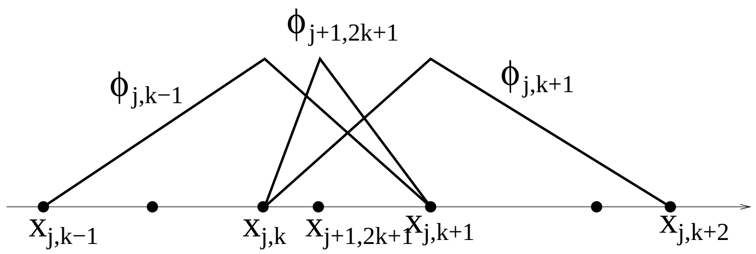


Figure 1.3. Scaling basis functions for linear prediction.

An update step with two non-zeros aiming at two vanishing moments (zero integral and first moment) satisfies:

$$A_{j,k} = \frac{M_{j+1,2k+1}}{M_{j,k}} \frac{\overline{x}_{j,k+1} - \overline{x}_{j+1,2k+1}}{\overline{x}_{j,k+1} - \overline{x}_{j,k}}$$

$$B_{j,k} = \frac{M_{j+1,2k+1}}{M_{j,k+1}} \frac{\overline{x}_{j+1,2k+1} - \overline{x}_{j,k}}{\overline{x}_{j,k+1} - \overline{x}_{j,k}}.$$

In these expressions, $M_{j,k}$ stands for the scaling function integrals:

$$M_{j,k} = \int_{-\infty}^{\infty} \varphi_{j,k}(x)dx,$$

and $\overline{x}_{j,k}$ stands for the first normalised moment:

$$\overline{x}_{j,k} = \frac{\int_{-\infty}^{\infty} x\varphi_{j,k}(x)dx}{\int_{-\infty}^{\infty} \varphi_{j,k}(x)dx}.$$

In the case of linear interpolating prediction, the scaling basis functions are piecewise linear (first order splines) as depicted in Figure 1.3. The classical update step aims at vanishing moments only. Unlike the Haar update, this update step is not a full orthogonalisation. Nevertheless, it does have a stabilising effect:

Theorem 1.1 *In a lifting scheme with linear prediction and update with two non-zero elements, both used for vanishing moments, the unlifted detail basis function (i.e. the odd scaling function at fine scale) is closer to the space spanned by the coarse scale scaling functions than the lifted detail basis function.*

Proof:
The projection of $\varphi_{j+1,2k+1}$ onto $\text{Span}\{\varphi_{j,k}, \varphi_{j,k+1}\}$ can be written as:

$$\begin{aligned} \varphi_{j+1,2k+1} &= \varphi^{\perp}_{j+1,2k+1} + \varphi^{//}_{j+1,2k+1} \\ &= \varphi^{\perp}_{j+1,2k+1} + b_{j,k}\varphi_{j,k} + a_{j,k}\varphi_{j+1,k} \end{aligned}$$

The wavelet function can then be decomposed into:

$$\begin{aligned}
\psi_{j,k} &= \varphi_{j+1,2k+1} - B_{j,k}\varphi_{j,k} - A_{j,k}\varphi_{j,k+1} \\
&= \varphi^{\perp}_{j+1,2k+1} + (b_{j,k} - B_{j,k})\varphi_{j,k} + (a_{j,k} - A_{j,k})\varphi_{j+1,k} \\
&=: \varphi^{\perp}_{j+1,2k+1} + \psi^{//}_{j,k}.
\end{aligned}$$

So, both have the same orthogonal component. If $\|\psi^{//}_{j,k}\| \leq \|\varphi^{//}_{j+1,2k+1}\|$, then it follows that the angle between $\psi_{j,k}$ and $\mathrm{Span}\{\varphi_{j,k}, \varphi_{j,k+1}\}$ is greater than that between $\varphi_{j+1,2k+1}$ and $\mathrm{Span}\{\varphi_{j,k}, \varphi_{j,k+1}\}$. For normalised scaling functions, we have:

$$\|\varphi^{//}_{j+1,2k+1}\| = a^2_{j,k} + b^2_{j,k} + 2a_{j,k}b_{j,k}\cos\alpha_{j,k},$$

where α is the angle between $\varphi_{j,k}$ and $\varphi_{j,k+1}$. We now verify that all four $a_{j,k}, b_{j,k}, A_{j,k}, B_{j,k}$ are positive and that the latter two are bounded by:

$$A_{j,k} \leq 2a_{j,k} \text{ and } B_{j,k} \leq 2b_{j,k}. \tag{1.2}$$

To check this, and without loss of generality, we put: $x_{j,k} = 0$, $x_{j,k+1} = 1$, $x_{j+1,2k+1} = z$, $x_{j,2k} = 1 + \Delta$ and $x_{j,k-1} = -\Gamma$. Thanks to the closed form of the scaling functions, all four quantities $a_{j,k}, b_{j,k}, A_{j,k}, B_{j,k}$ can then be expressed as functions of z, Δ, Γ and the inequalities (1.2) follow for all real values of these independent variables. These inequalities suffice to prove the requested norm inequality. □

This result is a relative one in the sense that it compares the stability of the basis after primal lifting with that before primal lifting. This basis may be unstable in the first place, if two data points are arbitrarily close to each other. In that case however, even the orthogonal Haar has difficulties: the non-updated Haar basis can be arbitrarily unstable and in the limit, the updated basis function would be zero, since that is the only function with zero integral on an interval with three data points of which two coincide.

The proof heavily depends on the closed form of the basis functions. Higher order prediction schemes do not lead to closed forms (the basis functions corresponding to cubic interpolation are definitely not cubic splines), and hence proving results for those schemes will be hard.

For the classical higher order lifting, there is little to prove anyway, since these schemes turn out to be arbitrarily bad with respect to numerical condition. Bad condition is the combined effect of splitting, prediction and update. Polynomial prediction is nothing but the Deslauriers-Dubuc subdivision scheme. This scheme as such is not unstable, but it may lead to large update coefficients. Small update coefficients are necessary for good condition. The key to bounded update coefficients is the values of the normalised moments of the scaling functions. Linear prediction leads to positive basis functions, and the normalised moments are then the 'mean values' of these functions. From Figure 1.3, it follows that those 'mean values' appear in a good order, which makes the update coefficients to be bounded.

Higher order prediction schemes however, lead to negative side lobes, which can be heavy if a large gap without data is followed or preceded by a densely sampled region. The even-odd-splitting disregards this mixture of scales. To avoid this scale mixture, an alternative splitting scheme takes out all odd points that are too far away from immediate neighbours [16]. Prediction of these points is deferred to coarser scales. The effect of this reorganisation is tremendous, as Figure 1.4 illustrates.

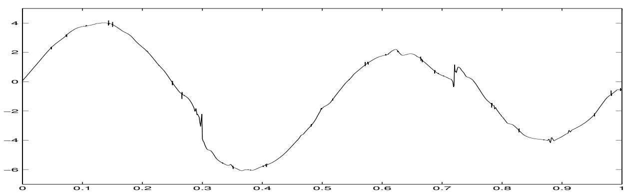


Figure 1.4. Reorganisation of the splitting operation prevents mixing of scale. This leads to more stable transforms and reduces the output bias.

1.5 Average interpolating prediction

So far, this paper has concentrated on lifting schemes with one, interpolating, prediction step. A slightly more complicated scheme interprets the input not as point samples, but rather as short time averages. The odd indexed averages are predicted by an *average interpolating polynomial.* This is constructed by first taking the mean of every even-odd input pair. These mean values are obviously also the averages on the joint even-odd-intervals. Next, we look for polynomials that average out at those means on a fixed number of consequent intervals. The average of such a polynomial on its middle, odd, interval is then used as prediction for the input value on that interval. This operation can be decomposed into a Haar transform followed by a dual lifting step. One more update step is then necessary to create more primal vanishing moments (Haar giving us the first).

Experiments as in Figure 1.5 show that this construction is more robust to irregular grids than interpolating prediction. On "very" irregular grids (i.e. when samples are far from uniformly distributed on the observation interval), this approach however equally fails. Reorganising this transform to obtain more stable smoothing is subject of current research. One possible method is to use some of the degrees of freedom in the last update step for a local pseudo-orthogonalisation [6]. This obviously applies in the context of polynomial prediction as well. We could also consider alternative splitting schemes, where no prediction is performed on a joint even-odd interval if this interval is considerably wider than its neighbouring joint intervals (e.g. twice as wide).

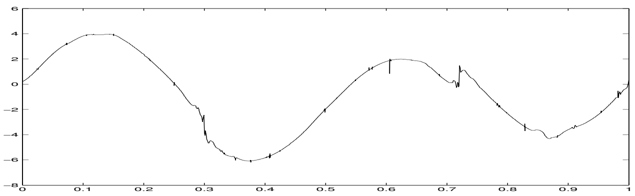
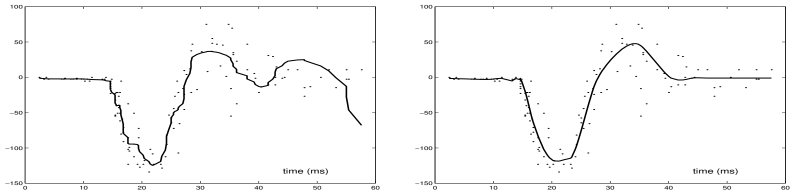


Figure 1.5. Average interpolating prediction is far less sensitive to instabilities due to scale mixing.

1.6 Real data example

We end this discussion with a real data set, in which some measurements occurred simultaneously [14]. The classical second generation transform crashes when three or more abscis points coincide (the return values were ∞). Thanks to the separate handling of phenomena at different scales, no problem arises in the improved splitting scheme, as shown in the output in Figure 1.6.



(a) (b)

Figure 1.6. Fitting the motorcyclists' acceleration data. (a) Using a classical first generation wavelet transform (cubic interpolating prediction, 2-taps update, no grid information) leads to non-smooth fitting. (b) The result from the proposed procedure is smooth and relatively unbiased.

1.7 Conclusions

This paper has analysed the stability of multiscale data smoothing on irregular point sets. The lifting scheme itself provides a *grid-adaptive* smoothing scheme, but it tends to mix up phenomena at different scales, which leads to unstable — and hence biased — curve fitting. This paper proposes a modified, *scale*-adaptive approach, which even deals automatically with simultaneous observations.

Acknowledgement

Maarten Jansen is postdoctoral research fellow of the Fund for Scientific Research Flanders (Belgium) (FWO).

References

[1] A. Antoniadis, G. Grégoire, and W McKeague.Wavelet methods for curve estimation.*J. Amer. Statist. Assoc.*, 89:1340–1353, 1994.

[2] A. Antoniadis, G. Grégoire, and P. Vial.Random design wavelet curve smoothing.*Statistics and Probability Letters*, 35:225–232, 1997.

[3] A. Antoniadis and D.-T. Pham.Wavelet regression for random or irregular design.*Computational Statistics and data analysis*, 28(4):333–369, 1998.

[4] T. Cai and L.D. Brown.Wavelet shrinkage for nonequispaced samples.*Annals of Statistics*, 26(5):1783–1799, 1998.

[5] I. Daubechies and W. Sweldens.Factoring wavelet transforms into lifting steps.*J. Fourier Anal. Appl.*, 4(3):245–267, 1998.

[6] V. Delouille, J. Simoens, and R. von Sachs.Smooth design-adapted wavelets for nonparametric stochastic regression.Technical report, Université Catholique de Louvain, 2001.

[7] B. Delyon and A. Juditsky.On the computation of wavelet coefficients.*J. of Approx. Theory*, 88:47–79, 1997.

[8] G. Donovan, J. Geronimo, and D. Hardin.Squeezable orthogonal bases: Accuracy and smoothness.Technical report, Department of Mathematics, Vanderbilt University, 2001.preprint.

[9] P. Hall and B. A. Turlach.Interpolation methods for nonlinear wavelet regression with irregularly spaced design.*Annals of Statistics*, 25(5):1912 – 1925, 1997.

[10] M. Jansen.*Noise reduction by wavelet thresholding*, volume 161 of *Lecture Notes in Statistics*.Springer, 2001.

[11] A. Kovac and B. W. Silverman.Extending the scope of wavelet regression methods by coefficient-dependent thresholding.*J. Amer. Statist. Assoc.*, 95:172–183, 2000.

[12] M. Pensky and B. Vidakovic.On non-equally spaced wavelet regression.Preprint, Duke University, Durham, NC, 1998.

[13] S. Sardy, D.B. Percival, A.G. Bruce, H-Y. Gao, and W. Stuetzle.Wavelet denoising for unequally spaced data.*Statistics and Computing*, 9:65–75, 1999.

[14] B. Silverman.Some aspects of the spline smoothing approach to nonparametric regression curve fitting.*Journal of the Royal Statistical Society, Series B*, 47:1–52, 1985.

[15] W. Sweldens and P. Schröder.Building your own wavelets at home.In *Wavelets in Computer Graphics*, ACM SIGGRAPH Course Notes, pages 15–87. ACM, 1996.

[16] E. Vanraes, M. Jansen, and A. Bultheel.Stabilized lifting steps in noise reduction for non-equispaced samples.In M. A. Unser, A. Aldroubi, and Laine A. F., editors, *Wavelet Applications in Signal and Image Processing IX*, volume 4478 of *SPIE Proceedings*, pages 105–116, July 2001.