MULTISCALE VARIANCE-STABILIZING TRANSFORM FOR MIXED-POISSON-GAUSSIAN PROCESSES AND ITS APPLICATIONS IN BIOIMAGING

B. Zhang[a], M. J. Fadili[b], J.-L. Starck[c], J.-C. Olivo-Marin[a]*

[a]UAIQ URA CNRS 2582
Institut Pasteur 75724 Paris France

[b]GREYC UMR CNRS 6072
14050 Caen France

[c]DAPNIA/SEDI-SAP CEA-Saclay
91191 Gif-sur-Yvette France

ABSTRACT

Fluorescence microscopy images are contaminated by photon and readout noises, and hence can be described by Mixed-Poisson-Gaussian (MPG) processes. In this paper, a new variance stabilizing transform (VST) is designed to convert a filtered MPG process into a near Gaussian process with a constant variance. This VST is then combined with the isotropic undecimated wavelet transform leading to a multiscale VST (MS-VST). We demonstrate the usefulness of MS-VST for image denoising and spot detection in fluorescence microscopy. In the first case, we detect significant Gaussianized wavelet coefficients under the control of a false discovery rate. A sparsity-driven iterative scheme is proposed to properly reconstruct the final estimate. In the second case, we show that the MS-VST can also lead to a fluorescent-spot detector, where the false positive rate of the detection in pure noise can be controlled. Experiments show that the MS-VST approach outperforms the generalized Anscombe transform in denoising, and that the detection scheme allows efficient spot extraction from complex background.

***Index Terms*—** variance stabilizing transform, Mixed-Poisson-Gaussian process, wavelet, fluorescence microscopy

1. INTRODUCTION

Fluorescence microscopy is a widely used technique to image biological specimens. The resulting images are corrupted by photon and camera readout noises. The stochastic data model is thus a Mixed-Poisson-Gaussian (MPG) process. For many applications such as denoising and deconvolution, it would be rather complicated to directly deal with such processes since every sample exhibits an infinite Gaussian mixture distribution. A commonly used technique is to first apply a variance stabilizing transform (VST), e.g., the generalized Anscombe transform (GAT) [1], to Gaussianize the data so that each sample is near-normally distributed with an asymptotically constant variance. The VST allows to apply standard denoising and deconvolution methods on the transformed data. Then, the final estimate is obtained by inverting the VST on the processed data.

In this paper, we propose a new VST to Gaussianize a low-pass filtered MPG process. This transform can be considered as a generalization of the GAT and a recently proposed VST for Poisson data [2]. Then, this VST is combined with the isotropic undecimated wavelet transform (IUWT) [1] leading to a multiscale VST (MS-VST). The usefulness of MS-VST is demonstrated for image denoising and spot detection in fluorescence microscopy. In the first case, we detect significant Gaussianized wavelet coefficients under the control of a false discovery rate (FDR) [3]. A sparsity-driven iterative scheme is proposed to properly reconstruct the final estimate. In the second case, we show that a slight modification of the denoising algorithm leads to a fluorescent-spot detector, where the false positive rate of the detection in pure noise can be controlled. Experiments show that the MS-VST approach outperforms the GAT in denoising, and that the proposed detection scheme allows efficient spot extraction from complex background.

2. VST FOR A FILTERED MPG PROCESS

A MPG process $\mathbf{x} := (X_i)_{i\in\mathbb{Z}^d}$ is defined as:

$$X_i = \alpha U_i + V_i, \quad U_i \sim \mathcal{P}(\lambda_i), \quad V_i \sim \mathcal{N}(\mu, \sigma^2) \qquad (1)$$

where $\alpha > 0$ is the overall gain of the detector, U_i is a Poisson variable modeling the photon counting, V_i is a normal variable representing the readout noise, and all $(U_i)_i$ and $(V_i)_i$ are assumed mutually independent. Given a discrete filter h, we note a filtered MPG process as $Y_i := \sum_j h[j]X_{i-j}$. We will use X and Y to denote any one of X_i and Y_i respectively. We further denote by τ_k the quantity $\sum_i (h[i])^k$ for $k \in \mathbb{N}^*$.

To simplify the following analysis we assume that $\lambda_i = \lambda$ within the support of h. It can be verified that the variance of Y ($\mathrm{Var}\,[Y]$) is an affine function of the Poisson intensity λ. To stabilize $\mathrm{Var}\,[Y]$, we seek a transformation $Z := T(Y)$ such that $\mathrm{Var}\,[Z]$ is (asymptotically) constant, irrespective of the value of λ. We define:

$$T(Y) := b \cdot \mathrm{sgn}(Y + c)|Y + c|^{1/2}, \quad b \neq 0,\ c \in \mathbb{R} \qquad (2)$$

*This work is funded by CNRS and Institut Pasteur of France. E-mail: {bzhang,jcolivo}@pasteur.fr

Lemma 1 indicates that the square-root transform (2) is indeed a VST for stabilizing and Gaussianizing a low-pass filtered MPG process.

Lemma 1 (square root as VST [4]) *If $\tau_1 \neq 0$, then we have:*

$$T(Y) - b \cdot sgn(\tau_1)\sqrt{|\tau_1|\alpha\lambda} \xrightarrow[\lambda\to+\infty]{\mathcal{D}} \mathcal{N}\left(0, \frac{\alpha b^2 \tau_2}{4|\tau_1|}\right) \quad (3)$$

This result holds for any $c \in \mathbb{R}$. However, the convergence rate in (3) varies with the value of c (b is only a normalizing factor), and we want to determine its optimal value.

2.1. Optimal parameter of the VST

Without loss of generality, suppose that $\tau_1 > 0$, then $\Pr(Y + c > 0)$ can be made arbitrarily close to 1 as $\lambda \to +\infty$. So in our asymptotic analysis below, we will essentially consider the VST in the form $T(Y) = bT_0(Y) = b\sqrt{Y+c}$. Expanding $T_0(Y)$ by Taylor series about the point $Y = \mathbb{E}[Y]$ up to the 4th order term, and by applying the expectation one can calculate the asymptotic expectation and variance of $T(Y)$:

$$\mathbb{E}[b_1 T_0] \approx \sqrt{\lambda} + \underbrace{\frac{4\tau_1(\tau_1\mu + c) - \tau_2\alpha}{8\tau_1^2\alpha}}_{C_{\mathbb{E}}} \lambda^{-1/2} \quad (4)$$

$$\mathrm{Var}[b_2 T_0] \approx 1 + \underbrace{\frac{8\tau_1^2\tau_2(\sigma^2 - \alpha\mu) - 4\tau_1\alpha(2\tau_2 c + \tau_3\alpha) + 7\tau_2^2\alpha^2}{8\alpha^2\tau_1^2\tau_2}}_{C_{\mathrm{Var}}} \lambda^{-1} \quad (5)$$

where $b_1 = (\tau_1\alpha)^{-\frac{1}{2}}$ and $b_2 = 2(\frac{\tau_1}{\alpha\tau_2})^{\frac{1}{2}}$. These settings normalize respectively the asymptotic expectation and variance to $\sqrt{\lambda}$ and 1, both values being independent of the filter h. Then the optimal c is found by minimizing the following bias-variance tradeoff (controlled by η):

$$c^* := \arg\min_{c\in\mathbb{R}} E_\eta(c) := \eta C_{\mathbb{E}}^2 + (1-\eta)|C_{\mathrm{Var}}|,\ \eta \in [0,1] \quad (6)$$

With no prior preference for either bias or variance, η can be set to $1/2$. Note that $C_{\mathbb{E}}$ is squared to give an equivalent asymptotic rate for the tradeoff terms in (4) and (5). It can be shown that (6) admits a unique solution, which can be explicitly derived out as a function of τ_k, μ, σ, α and η. This VST reduces to the GAT if h = Dirac filter δ and $\eta = 0$.

In practice, if μ, σ, and α are unknown a priori, they can be estimated by matching the first four cumulants of X with the k-statistics [5] of the samples in a uniform image region. This follows from the property that the k-statistics are the minimum variance unbiased estimators for cumulants.

3. IMAGE DENOISING USING MS-VST

Isotropic structures are often presented in biological fluorescent images due to micrometric subcellular sources. Toward the goal of image denoising, we will combine the proposed VST with the IUWT. Indeed, since IUWT uses isotropic filter banks, this transform adapts very well the isotropic features in images. The left side of (7) gives the classical IUWT decomposition scheme, and by applying the VST on the (low-pass filtered) approximation coefficients at each scale, we obtain a MS-VST scheme shown on the right side:

$$\left\{ \begin{array}{ll} a_j &= \bar{h}^{\uparrow j-1} \star a_{j-1} \\ d_j &= a_{j-1} - a_j \end{array} \right. \Rightarrow \left\{ \begin{array}{ll} a_j &= \bar{h}^{\uparrow j-1} \star a_{j-1} \\ d_j &= T_{j-1}(a_{j-1}) - T_j(a_j) \end{array} \right. \quad (7)$$

Here h is a symmetric low-pass filter, a_j and d_j are respectively the approximation and the wavelet coefficients at scale j ($\leq J$), $h^{\uparrow k}[l] = h[l]$ if $l/2^k \in \mathbb{Z}$ and 0 otherwise, $\bar{h}[n] = h[-n]$ and "$\star$" denotes convolution. The filtering of a_{j-1} can be rewritten as a filtering of the original MPG data $\mathbf{x} \equiv a_0$, i.e., $a_j = h^{(j)} \star a_0$, where $h^{(j)} = \bar{h}^{\uparrow j-1} \star \cdots \star \bar{h}^{\uparrow 1} \star \bar{h}$ for $j \geq 1$ and $h^{(0)} = \delta$. T_j is the VST operator at scale j (cf. (2)):

$$T_j(a_j) = b^{(j)}\mathrm{sgn}(a_j + c^{(j)})|a_j + c^{(j)}|^{1/2}$$

The constants $b^{(j)}$ and $c^{(j)}$ are associated to $h^{(j)}$, and $c^{(j)}$ should be set to c^*. Theorem 1 shows that (7) transfers the asymptotic stabilized Gaussianity of the a_j's to the d_j's:

Theorem 1 (d_j under a high intensity assumption) *Setting $b^{(j)} := sgn(\tau_1^{(j)})/[\alpha|\tau_1^{(j)}|]^{1/2}$, we have:*

$$d_j \xrightarrow[\lambda\to+\infty]{\mathcal{D}} \mathcal{N}\left(0, \quad \frac{\tau_2^{(j-1)}}{4{\tau_1^{(j-1)}}^2} + \frac{\tau_2^{(j)}}{4{\tau_1^{(j)}}^2} - \frac{\langle h^{(j-1)}, h^{(j)}\rangle}{2\tau_1^{(j-1)}\tau_1^{(j)}}\right)$$

where $\tau_k^{(j)} := \sum_i \left(h^{(j)}[i]\right)^k$, and $\langle\cdot,\cdot\rangle$ denotes inner product. This result shows that the asymptotic variance of d_j depends only on the wavelet filter bank and the current scale, and thus can be pre-computed once h is chosen.

3.1. Detection of significant coefficients by FDR

Wavelet denoising can be achieved by zeroing the insignificant coefficients while preserving the significant ones. We detect the significant coefficients by testing binary hypothesis: $\forall\, d,\ H_0:\ d = 0$ vs. $H_1:\ d \neq 0$. The distribution of d under the null hypothesis H_0 is given in Theorem 1. Thus, a multiple hypothesis testing controlling the FDR can be carried out [3]. The control of FDR offers many advantages over the classical Bonferroni control of the Family-Wise Error Rate, i.e., the probability of erroneously rejecting even one of the true null hypothesis. For example, FDR usually has a greater detection power and can handle correlated data easily. The latter point is important since the IUWT is over-complete.

3.2. Sparsity-driven iterative reconstruction

After coefficient detection, we could invert the MS-VST (7) to get the final estimate: $\hat{a}_0 = T_0^{-1}[T_J(a_J) + \sum_{j=1}^{J} d_j]$, but this solution is far from optimal. Indeed, due to the non-linearity of the VST and the over-completeness of IUWT, the significant coefficients are not reproducible when IUWT is applied once more on this direct inverse, implying a loss of important structures in the estimation. A better way is to find a constrained sparsest solution, as sketched below (see [4] for details).

We first define the multi-resolution support [1] $\mathcal{M} := \{(j,l) \mid d_j[l] \text{ is significant}\}$, which is determined by the set of the detected significant coefficients. The estimation is then formulated as a constrained convex optimization problem in terms of wavelet coefficients:

$$\min_{\mathbf{d}\in\mathcal{C}} J(\mathbf{d}) := \|\mathbf{d}\|_1 \text{ where } \mathcal{C} := \mathcal{S}_1 \cap \mathcal{S}_2$$
$$\mathcal{S}_1 := \{\mathbf{d} | \mathbf{d} = \mathcal{W}\mathbf{x} \text{ in } \mathcal{M}\} \text{ and } \mathcal{S}_2 := \{\mathbf{d} | \mathcal{R}\mathbf{d} \geq \mu\} \tag{8}$$

where $\mathcal{W}$ is the wavelet analysis operator, and $\mathcal{R}$ its synthesis operator. Clearly by doing so, we minimize a sparsity-promoting ℓ^1 objective function [6] within the feasible set $\mathcal{C} := \mathcal{S}_1 \cap \mathcal{S}_2$. The set $\mathcal{S}_1$ requires that the elements of $\mathbf{d}$ preserve the significant coefficients; the set $\mathcal{S}_2$ assures a model-consistent estimate since $\mathbb{E}[X_i] = \alpha\lambda_i + \mu \geq \mu$.

Gradient descent method such as the hybrid steepest descent (HSD) iterations [7] can be used to solve (8):

$$\mathbf{d}^{(k+1)} := T_{\mathcal{C}}\mathbf{d}^{(k)} - \beta_{k+1}\operatorname{sgn}\left(T_{\mathcal{C}}\mathbf{d}^{(k)}\right) \tag{9}$$

where the step length β_k satisfies: (i) $\lim_{k\to\infty} \beta_k = 0$, (ii) $\sum_{k\geq 1} \beta_k = +\infty$, (iii) $\sum_{k\geq 1} |\beta_k - \beta_{k+1}| < +\infty$. The operator $T_{\mathcal{C}}$ is defined as $T_{\mathcal{C}} := P_{\mathcal{S}_1} \circ Q_{\mathcal{S}_2}$, and

$$P_{\mathcal{S}_1}\mathbf{d} := \begin{cases} \mathcal{W}\mathbf{x} & \text{in } \mathcal{M} \\ \mathbf{d} & \text{otherwise} \end{cases} ; \quad Q_{\mathcal{S}_2}\mathbf{d} := \mathcal{W}P_\mu\mathcal{R}\mathbf{d} \tag{10}$$

where P_μ is the projector onto the set $\{\mathbf{x} | x_i \geq \mu\}$. It is worth noting that compared with the direct reconstruction, every iteration of (9) involves a projection onto the set $\mathcal{S}_1$ that restores all the significant coefficients. Therefore, important structures are better preserved by the iteratively reconstructed solution.

3.3. Results

We first test our denoising approach on a simulated 18×10 isotropic-source grid (pixel size $= 100$ nm) shown in Fig. 1. From the leftmost to the rightmost column, the source radii increase from 50 nm to 350 nm. The image is then convolved by a 2D Gaussian function with a standard deviation $\sigma_g = 116$ nm, which approximates the point spread function of a typical fluorescence microscope [8]. Fig. 1(a) shows the sources with amplitudes $\lambda_A \in [0.05, 50]$. After adding a MPG noise, we obtain Fig. 1(b). Fig. 1(c) and (d) respectively show the denoising examples using the GAT and the MS-VST. More faint sources are restored by the MS-VST approach, showing its higher sensitivity. In terms of the mean ℓ^1-loss per bin, i.e., $\bar{\varepsilon} := \mathbb{E}[\frac{1}{n}\|\hat{a}_0 - \mathbb{E}[a_0]\|_{\ell^1}]$ where n is the number of pixels, the MS-VST denoising is more accurate ($\bar{\varepsilon} = 1.75$) than the GAT ($\bar{\varepsilon} = 1.94$), where $\bar{\varepsilon}$ is computed based on 100 replications.

Fig. 2(a) and (b) show two optical slices of a 3D confocal image of a drosophila melanogaster ovary. The part of nurse cells consist of many nucleus with Green-Fluorescent-Protein-marked Staufen genes. The slices of the denoised image are shown in Fig. 2(c) and (d). We can see clearly that

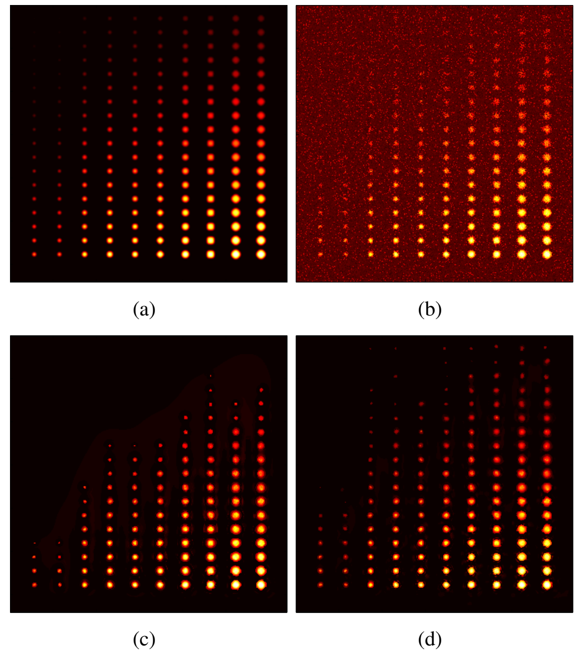


Fig. 1. Simulated source denoising. $h =$ 2D B_3-Spline filter, FDR $= 0.01$, $J = 5$ and 10 iterations. (a) simulated sources (amplitudes $\lambda_A \in [0.05, 50]$; background $= 0.05$); (b) MPG noisy image ($\alpha = 20$, $\mu = 10$, and $\sigma = 1$); (c) GAT-denoised image, $\bar{\varepsilon} = 1.94$; (d) MS-VST-denoised image ($\eta = 0.5$), $\bar{\varepsilon} = 1.75$

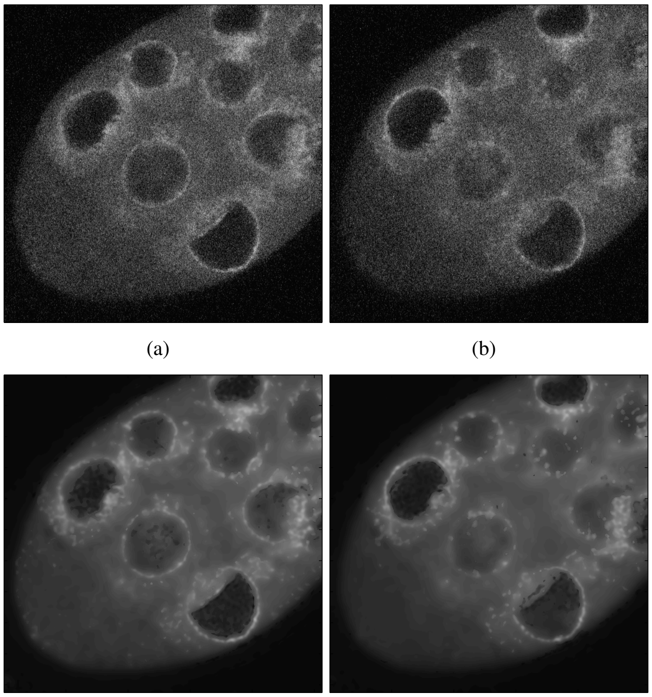
the cytoplasm (homogeneous areas) is well smoothed and the gene signals are restored from the noise.

4. SPOT DETECTION USING MS-VST

The MS-VST also allows us to construct a fluorescent-spot detector. Indeed, since wavelets are band-pass filters, background information is mostly encoded in the approximation band. Now, suppose that we have obtained $\mathcal{M}$ by the same detection procedure as in the denoising case. Then, if we take the wavelet transform of a_0, zero both the insignificant coefficients (by referring to $\mathcal{M}$) and the approximation band, and reconstruct the image, the background will be largely suppressed from the final estimate and, consequently, only detail (spot) structures are retained. Finally, we binarize the result by thresholding the negative pixels to zero, and then extract all connected components as putative bright spots. With this approach, if the FDR of the wavelet coefficient detection is upper bounded by γ, the probability of erroneously detecting spots in a spot-free homogeneous MPG noise ($\lambda_i = \lambda$) is also upper bounded by γ.

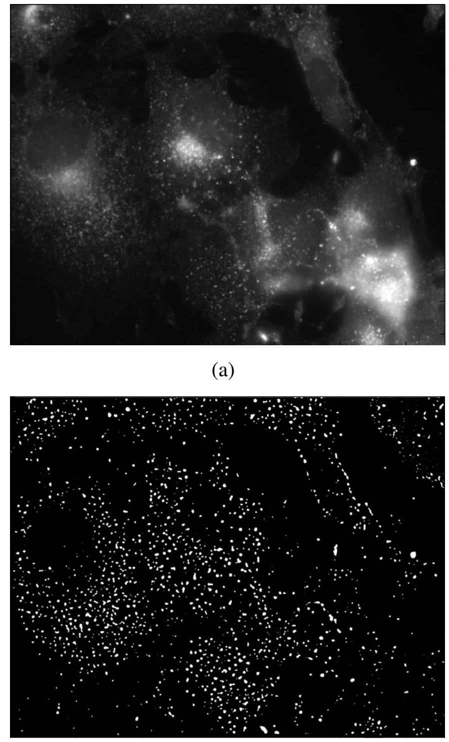
4.1. Results

Fig. 3 shows the detection of endocytic vesicles of COS-7 cells in a wide-field microscopy image. Although the original



(a) (b) (c) (d)

Fig. 2. Denoising of a 3D confocal image of a drosophila melanogaster ovary. $h =$ 3D B_3-Spline filter, FDR $= 0.01$, $\eta = 0.5$, $J = 4$, and 10 iterations. Observed image: (a) $z = 22\mu m$; (b) $z = 26\mu m$; MS-VST-denoised image: (c) $z = 22\mu m$; (d) $z = 26\mu m$.



(a) (b)

Fig. 3. Endocytic-vesicle detection in a wide-field microscopy image of COS-7 cells. (a) original image; (b) identified spots ($h =$ 2D B_3-Spline filter, $\eta = 0.5$, $J = 3$ and FDR $= 10^{-6}$).

image exhibits a highly nonuniform background (Fig. 3(a)), the detection (Fig. 3(b)) is very effective as most spots are well extracted while the background is canceled.

5. CONCLUSION

We have designed a VST to stabilize and Gaussianize a low-pass filtered MPG process. The VST is then combined with the IUWT yielding the MS-VST. We have shown the MS-VST approach to be very effective in fluorescent image denoising and spot detection. Our future work will apply the MS-VST in deconvolution and super-resolution detection.

ACKNOWLEDGMENT

The authors would like to thank M. M. Mhlanga (Institut Pasteur) for providing the image of drosophila melanogaster ovary, and A. Hémar (CNRS UMR 5091) for providing the COS-7 cell image.

6. REFERENCES

[1] J.-L. Starck, F. Murtagh, and A. Bijaoui, *Image Processing and Data Analysis*, Cambridge University Press, 1998.

[2] B. Zhang, M. J. Fadili, and J.-L. Starck, "Multi-scale Variance Stabilizing Transform for Multi-dimensional Poisson Count Image Denoising," in *ICASSP*, 2006, vol. 2, pp. 81–84.

[3] Y. Benjamini and D. Yekutieli, "The control of the false discovery rate in multiple testing under dependency," *Ann. Statist.*, vol. 29, no. 4, pp. 1165–1188, 2001.

[4] B. Zhang, M. J. Fadili, and J.-L. Starck, "Wavelets, Ridgelets and Curvelets for Poisson Noise Removal," *IEEE Transactions on Image Processing*, 2006, submitted.

[5] C. Rose and M. D. Smith, *Mathematical Statistics with Mathematica*, chapter 7.2C: k-Statistics: Unbiased Estimators of Cumulants, pp. 256–259, Springer-Verlag, 2002.

[6] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization," *PNAS*, vol. 100, no. 5, pp. 2197–2202, 2003.

[7] I. Yamada, "The hybrid steepest descent method for the variational inequality problem over the intersection of fixed point sets of nonexpansive mappings," in *Inherently Parallel Algorithm in Feasibility and Optimization and their Applications*, pp. 473–504. Elsevier, 2001.

[8] B. Zhang, J. Zerubia, and J.-C. Olivo-Marin, "Gaussian approximations of fluorescence microscope PSF models," *Applied Optics*, vol. 46, no. 10, pp. 1819–1829, 2007.