

# Mid-level Shape Priors for Supervised Contouring

T. Shepherd<sup>a</sup> and D. C. Alexander.<sup>a\*</sup>

<sup>a</sup>Centre for Medical Image Computing, University College London, Malet Place Engineering Building, WC1E 6BT

**Abstract.** This paper introduces shape priors that benefit 2-dimensional, interactive contouring, which is still a common practice in lesion segmentation due to problems of boundary ambiguity and variable shape. The new Statistical Shape Models (SSMs) capture global shape information without requiring correspondence. We use time series representations of 2-dimensional boundaries and adapt *Langevin* and *Gaussian Process* dynamical models to capture global shape properties based on boundary fluctuations. Experiments reveal the benefits of the shape priors in supervised segmentation frameworks.

## 1 Introduction

Medical region segmentation has high demands in terms of accuracy and variability, but is confounded by low image quality. Regions such as liver tumours and multiple sclerosis (MS) lesions pose an extra challenge due to their variable shape. In practice, automatic segmentation of these regions is limited and methods call for interactive post-editing. We address the problems of tumour and lesion segmentation with supervised contouring tools. We seek methods that balance the requirements of maximising user control and minimising the demand on the user without compromising accuracy. Our broad approach is to incorporate prior knowledge of image and shape, and to make efficient use of on-line supervision through interactions that complement this prior knowledge. In this paper, we develop new statistical shape models (SSMs) that form the basis of supervised segmentation frameworks. The SSMs exploit various techniques from the field of nonlinear dynamics, to add *mid-level* shape priors. We use the term 'mid-level' to mean the highest level of shape information available in the absence of correspondence.

Section 2 discusses previous work on supervised segmentation and shape modelling, highlighting the limitations for the applications above and motivating the use of nonlinear dynamics. Section 3 describes the chosen shape representation and introduces two dynamical models in general terms, namely *Langevin* and *Gaussian Process* models. Section 4 formally defines the SSMs, describing their deterministic component, training methods, discriminative and generative techniques and how these are built into segmentation frameworks. Experiments in section 5 reveal the benefits of the shape priors for the applications of liver tumour and MS lesion contouring.

## 2 Previous Work

A large family of SSMs represent an object boundary by a set of position vectors that correspond to boundary features [1, 2]. Following the seminal work by Cootes and Taylor [1], the 'point distribution model' (PDM) requires training data in the form of  $M$  region boundaries labelled with  $N$  'landmarks', or points  $\{x_i, y_i\}$ , that correspond among all examples. The mean shape  $\hat{\mathbf{Q}}$  and covariance matrix  $\Sigma$  are used in an active shape model (ASM). An ASM represents variations in the training set by the first  $n \ll M$  eigenvectors of  $\Sigma$  using principal component analysis (PCA). The ASM approximates all model shapes by a linear combination of eigenmodes  $\hat{\mathbf{Q}} + \epsilon \mathbf{a}$ , where  $\epsilon$  is the  $N \times n$  matrix of the first  $n$  eigenvectors and  $\mathbf{a}$  is a vector of  $n$  shape parameters.

Another family of SSMs represent a 2-dimensional region boundary as a vector of radial distances  $\mathbf{r} = \{r_0, \dots, r_{N-1}\}$ , measured from a fixed location  $\mathbf{x}_c = \{x_c, y_c\}$  inside the region [3–5]. We refer to this parametrisation as a *radial time series*. Radial time series were first used for shape modelling in [3]. The so-called Circular Autoregressive (CAR) model expresses a boundary point  $r_i$  as a weighted sum of  $M$  previous points  $\{r_{i-M}, \dots, r_i\}$  plus a noise term. Several authors have used this approach for biomedical shapes (eg [4]). A nonlinear variant of the CAR model, known as the Quadratic Volterra Type (QVT), has been shown to handle more complex shapes [5].

The main limitation of the point distribution model is the assumption of correspondence points or 'landmarks' in the training data. Apart from the difficulty of labelling landmarks in large training sets, regions such as tumours and lesions do not *have* correspondence points. A further limitation is that linear PCA struggles to model pathological variations. Although nonlinear variants exist [2], these still suffer from the assumption of correspondence. Radial time series representations do not assume correspondence points. However, the CAR and its nonlinear equivalent are not generative models and are used for shape classification rather than region segmentation.

---

\*{T.Shepherd,D.Alexander}@cs.ucl.ac.uk

We seek SSMs that are nonlinear, capture high-level shape information without assuming correspondence, and can be exploited in segmentation. Our approach is to combine the radial time series representation with nonlinear, generative models from the field of time series analysis, which we demonstrate for Langevin and Gaussian Process models introduced in the next section.

### 3 Background

A radial time series stores the radius of  $N$  successive points around a region boundary  $\mathbf{r} = \{r_0, \dots, r_{N-1}\}$ , measured from a fixed location  $\mathbf{x}_c = \{x_c, y_c\}$  inside the region. There are different types of radial time series, differing by what 'time' represents. Two examples are boundary arc-length  $\mathbf{s} = \{s_0, \dots, s_{N-1}\}$  and the angle between radial vectors  $\theta = \{\theta_0, \dots, \theta_{N-1}\}$ . The arc-length parametrisation  $\{\mathbf{r}, \mathbf{s}\}$  can represent any two-dimensional shape which we refer to as the *generalised* set. The polar parametrisation  $\{\mathbf{r}, \theta\}$  is limited to the *star-shaped* set where all boundary points of a shape are visible from  $\mathbf{x}_c$ .

#### 3.1 Langevin models

Langevin models describe the dynamics of a time dependent state vector  $\mathbf{v}(t)$  as a stochastic process. The models are characterised by a deterministic part  $A(\mathbf{v}(t))$  known as the *drift* function and a stochastic part  $B(\mathbf{v}(t))$  known as the *diffusion* function. The generalised Langevin equation is given by

$$d\mathbf{v}/dt = A(\mathbf{v}(t)) + B(\mathbf{v}(t))\omega(t), \quad (1)$$

where  $\omega(t)$  is uncorrelated, time dependent noise with zero mean. Langevin models assume a Markov property, encoded in the *transition density*  $\Pr(\mathbf{v}(t)) = \Pr(\mathbf{v}(t)|\mathbf{v}(t - \Delta t))$ , where  $\Delta t$  is a constant delay parameter. The transition density evolves according to a Fokker-Planck equation. The first and second conditional moments in the Fokker-Planck equation, denoted  $\mathbf{D}^{(1)}$  and  $\mathbf{D}^{(2)}$ , relate to the Langevin drift and diffusion functions by

$$A(\mathbf{v}(t), t) \sim \mathbf{D}^{(1)}(\mathbf{v}(t), t)/\Delta t \quad \text{and} \quad B(\mathbf{v}(t), t) \sim \mathbf{D}^{(2)}(\mathbf{v}(t), t)/\sqrt{\Delta t}. \quad (2)$$

Training a Langevin model involves finding parametric functions for the drift  $A(\mathbf{v}(t), \mathbf{a})$  and diffusion  $B(\mathbf{v}(t), \mathbf{b})$ , where  $\mathbf{a}$  and  $\mathbf{b}$  are parameter vectors, by choosing functional forms and estimating the parameters from training data.

Generative Langevin models approximate the time evolution of the state vector as the limit of a stochastic difference equation (SDE). A Langevin SDE has the form

$$d\mathbf{v}(t) = A(\mathbf{v}(t))dt + B(\mathbf{v}(t))\omega(t) \quad (3)$$

where  $dt$  is an integration time step, which can be solved numerically by stochastic integration using the explicit Euler-Maruyama scheme

$$\mathbf{v}(t + dt) = \mathbf{v}(t) + dt A(\mathbf{v}(t)) + \sqrt{dt} B(\mathbf{v}(t))\omega(t). \quad (4)$$

#### 3.2 Gaussian Process models

A Gaussian Process treats a time series as a  $N$ -dimensional random vector of outputs  $\mathbf{v} = \{v_0, \dots, v_i, \dots, v_{N-1}\}$  corresponding to inputs at discrete time points  $\mathbf{t} = \{t_0, \dots, t_i, \dots, t_{N-1}\}$ . The output at time  $t_i$  has an associated probability  $\Pr(v_i|t_i)$ , which follows a normal distribution and the vector of outputs has a multivariate normal distribution  $\mathbf{v} \sim \mathcal{N}_N(\boldsymbol{\mu}, \boldsymbol{\Sigma}(\mathbf{v}, \mathbf{v}))$ , where  $\boldsymbol{\mu}$  is a discrete mean function and  $\boldsymbol{\Sigma}(\mathbf{v}, \mathbf{v})$  is the  $N \times N$  matrix of covariances  $\varepsilon(v_i, v_j)$  between pairs of outputs. Each covariance is a function of the corresponding inputs  $\varepsilon(v_i, v_j) = f(t_i, t_j)$  where  $f$  is a kernel function. The mean function  $\boldsymbol{\mu}$  and the covariance function  $\varepsilon(v_i, v_j)$  completely define a Gaussian Process [6].

Training a GP involves fitting a discrete function  $\boldsymbol{\mu}$  and a parametric kernel function  $\varepsilon(v_i, v_j, \mathbf{a})$  to training data, where  $\mathbf{a}$  is a vector of parameters. Ideally the functions and parameters are estimated together by Bayesian model selection [6]. It is however common to choose a sensible mean function and form of the covariance kernel, and the task reduces to estimating the parameters  $\mathbf{a}$ .

Generating a series from a Gaussian Process is equivalent to drawing a random vector of outputs  $\mathbf{v}'$  from the prior distribution. This involves solving

$$\mathbf{v}' = \boldsymbol{\mu} + \mathbf{A}\mathbf{z}, \quad (5)$$

where  $\mathbf{A}$  is the Cholesky decomposition of  $\boldsymbol{\Sigma}(\mathbf{v}, \mathbf{v})$  and  $\mathbf{z}$  is a  $N$ -dimensional vector of values drawn from  $\mathcal{N}(0, 1)$ .

## 4 SSMs in Supervised Contouring Frameworks

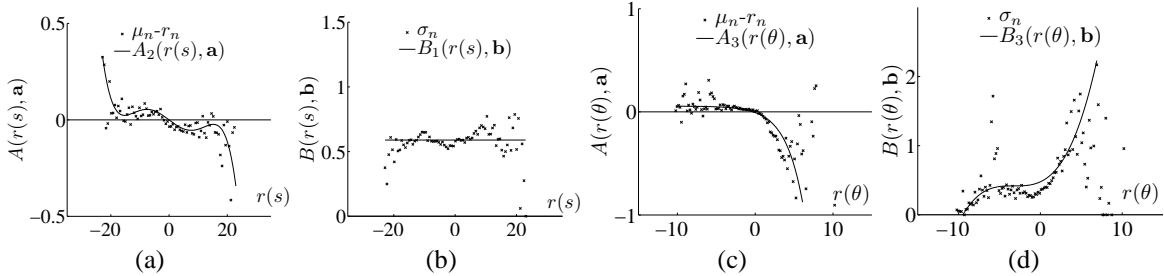
Components common to Langevin and GP SSMs are a scale parameter and an image observation model. Starting with a series derived from a training contour, we take the centre  $\bar{r}$  of the occupied state space and subtract this from the whole series, translating it into a *zero-mean field*, and separately store  $\bar{r}$  as the scale parameter. Next, we define an observation model in the time series framework, given here for the star-shaped case. Given an estimate of the region centre  $\mathbf{x}'_c$ , we sample the image gradient along the radial vector at  $\theta_i$ . We fit a Gaussian model  $g(r)$  to the radial profile of gradient magnitude, giving the mean  $\hat{g}_i$  and standard deviation  $\sigma_i^g$ . More details on constructing the observation model can be found in [7] Repeating for all  $\theta_i \in \theta$  gives the observation model  $\mathcal{D} = \{\hat{\mathbf{g}}, \sigma^g\}$ . Sections 4.1 and 4.2 present Langevin and GP SSMs using the general time series notation  $r(t)$ , and demonstrate for the polar parametrisation  $r(\theta)$ .

### 4.1 Langevin SSMs

We train a Langevin model by adapting the method in [8] to learn from  $M$  series  $\mathbf{r}^m$ ,  $m = 1 \dots M$ , created as above. We divide the range  $\{r_{\min}, \dots, r_{\max}\}$  of the whole training set into bins of equal width  $\Delta r$  centred on discrete values  $r_n$ . For the  $n$ th bin, we identify all points on all series for which  $|r(t) - r_n| < \Delta r/2$ . We then construct a histogram of radii at the displaced positions  $r(t + \Delta t)$  from all these points, allowing wrapping of the closed boundary, to obtain an approximation of the transition density  $\Pr(r(t + \Delta t)|r(t) \in r_n \pm \frac{\Delta r}{2})$ , which we model with a normal distribution by estimating the mean  $\mu_n$  and standard deviation  $\sigma_n$ . Repeating for all bins gives a discrete approximation of the drift and diffusion functions

$$\begin{aligned} A(r_n(t)) &= \mu_n - r_n & \text{and} \\ B(r_n(t)) &= \sigma_n & \text{where } r_n \in \{r_{\min}, \dots, r_n, \dots, r_{\max}\}. \end{aligned} \quad (6)$$

Finally, choose simple parametric drift and diffusion functions  $A(r(t), \mathbf{a})$  and  $B(r(t), \mathbf{b})$  and use a Levenberg-Marquardt routine to estimate parameters  $\mathbf{a}$  and  $\mathbf{b}$  that best fit the discrete estimations. We find that exponential, quadratic and polynomial functions work well in practice. We choose the function with the lowest  $\chi^2$  error to model a given region type with a given contour parametrisation. Figure 1 shows examples of trained Langevin models.



**Figure 1.** Drift/diffusion functions (a)/(b) for liver tumours with generalised parametrisation and (c)/(d) for MS lesions with star-shaped parametrisation.  $A_2$  is 5<sup>th</sup> order polynomial,  $B_1$  is constant,  $A_3$  is exponential and  $B_3$  is cubic.

We derive a 'score' for unseen shapes from the log of the joint conditional probabilities. The score is given by

$$S_{\text{LAN}} = \frac{1}{N} \left( \log \Pr(r(t_0)) + \left[ \sum_{i=1}^{N-1} \log \Pr(r(t_i + \Delta t)|r(t_i), \mathbf{a}) \right] + \log \Pr(\bar{r}) \right). \quad (7)$$

Next we modify the Euler-Maruyama scheme (4) for shape generation in polar coordinates

$$r(\theta_{i+1}) = r(\theta_i) + dt A(r(\theta_i)) + \sqrt{dt} B(r(\theta_i)) \omega(\theta_i) + \Delta x \cos(\theta_i) + \Delta y \sin(\theta_i), \quad (8)$$

where  $\{\Delta x, \Delta y\}$  allow uncertainty on the estimate of the region centre  $\mathbf{x}'_c$ , which we build into a further adaptation of the Euler-Maruyama scheme. We combine equation (8) with a boundary tracking algorithm in polar coordinates. The algorithm repeats the SDE solution several times at a given step  $\theta_{i+1}$ , which draws several samples from the transition density  $\Pr(r_{i+1}|r_i, \mathbf{a}, \mathbf{b})$ . This forms the prior distribution in a particle filtering algorithm that is otherwise similar to that in [9]. The algorithm uses factored sampling to compute posterior densities

$$\Pr(r_{i+1}|\mathbf{a}, \mathbf{b}, \mathcal{D}) \propto \Pr(r_i|\mathbf{a}, \mathbf{b}, \mathcal{D}) \times q(r_{i+1}|r_i, \mathbf{a}, \mathbf{b}) \times l(\mathcal{D}|r_{i+1}), \quad (9)$$

where  $q(r_{i+1}|r_i, \mathbf{a}, \mathbf{b}) = \mathcal{N}(r_i - A(r_{i+1}, \mathbf{a}), B(r_{i+1}, \mathbf{b}))$  is the shape prior arising from equation (8) and  $l(\mathcal{D}|r_{i+1})$  is a data likelihood term derived from the radial profile model above. For a given  $\{\Delta x, \Delta y\}$  we solve equation (9),

terminating when a series of  $N$  radii forms a closed loop with negligible discontinuity. We repeat for several  $\mathbf{x}'_c$  drawn from  $\mathcal{N}_2(\mathbf{x}'_c, \sigma_c)$  where the variance  $\sigma_c$  models the uncertainty on  $\mathbf{x}_c$ . We weight each hypothesis by equation (7) and estimate the maximum *a-posteriori* (MAP) solution by factored sampling.

## 4.2 Gaussian Process SSMs

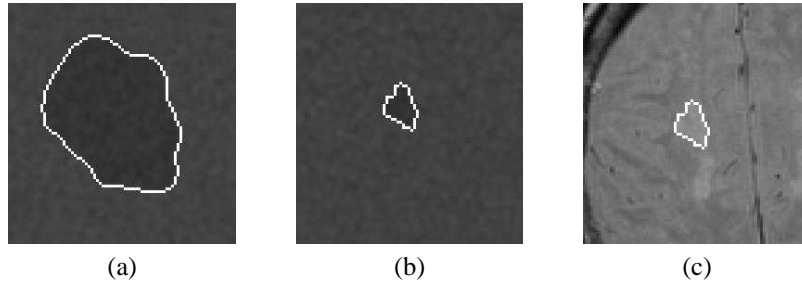
We present a general GP model using a periodic kernel function  $\varepsilon(t_i - t_j, \mathbf{a}) = \exp[-a \sin^2((t_j - t_i))]$ , with a single free parameter  $a$  related to the length scale of correlation. We use a constant mean function for rotation invariance. To train the GP model we estimate kernel parameters using Markov Chain Monte Carlo (MCMC) methods. Starting with  $M$  training series, we seek the parameter  $a$  that maximises the joint probability density function  $\Pr(\mathbf{r}^{1:M} | a)$ . For convenience we take the sum of the log probabilities  $\sum_{m=1}^M S_{\text{GP}}^m$  where

$$S_{\text{GP}}^m = \frac{N}{2} \log(2\pi) - \frac{1}{2} \log(|\Sigma(\mathbf{a})|) - \frac{1}{2} ((\mathbf{r}^m - \mu)^T \Sigma^{-1}(\mathbf{a})(\mathbf{r}^m - \mu)) + \log \Pr(\bar{r}). \quad (10)$$

We constrain the generative model by treating the radial profiles in  $\mathcal{D}$  as *noisy observations* and using these to condition the prior following the description in [6]. This yields the posterior covariance matrix  $\Sigma_{\text{post}}$  replacing  $\Sigma$  in equation (5). Finally the MAP solution is given by the posterior mean  $\mu_{\text{post}}$  which we calculate analytically. As in the Langevin case we repeat for several centre point estimates  $\mathbf{x}'_c$ , scoring each solution by (10) and selecting the final contour by factored sampling. Gaussian Process SSMs incorporate a novel method of supervision that works directly with the model. The user marks a point on the region boundary missed by the contour. The polar coordinates of such a point define a low-noise observation that further conditions the prior model.

## 5 Experiments

We perform user trials to test the segmentation frameworks and evaluate the SSMs for the chosen applications of liver tumour and MS lesion contouring. We start with 241 star-shaped training contours from manual liver tumour segmentations [10] and 1307 from MS lesions. We remove three test contours from each set and train SSMs on the remaining contours. We use the test contours to create synthetic images by assigning different greylevel histograms inside and outside a contour, then smoothing the result with a  $3 \times 3$  pixel averaging kernel (figure 2). In the case of MS lesions we also use the PD weighted MR images from where the training data were extracted.



**Figure 2.** Synthetic liver tumour (a) and MS lesion (b) and MRI MS lesion (c). White lines are ground truth boundaries.

To isolate the role of the shape priors we compare each framework with a second version, where the learned shape information is replaced with a *normal* prior. In the Langevin case we replace drift and diffusion functions with stationary transition densities  $\mathcal{N}(0, 1)$ . In the GP case we replace the covariance matrix with the  $N \times N$  identity matrix.

We asked 10 volunteers to delineate the 9 regions in a randomised sequence, using both tools with learned and normal priors. Segmentation comprises two stages of initialisation by choosing a centre point  $\mathbf{x}_c$ , which the user can change until satisfied, and post-editing. In the Langevin framework, post-editing involves 'dragging' contour points onto the region boundary. In the GP framework, users can mark boundary points as low-noise observations (section 4.2) and the updated model is displayed in real-time. We evaluate the SSMs in terms of user demand, accuracy and variability, and perform paired-samples T-tests to identify significant differences between a tool used with learned and normal prior.

Table 1 reveals the benefit of the shape priors to the useability and accuracy of the interactive tools segmenting liver tumours (LT) and multiple sclerosis lesions (MS). We measure useability by the number of interactions  $N_{\text{int}}$ , where one interaction is a contour point 'dragged' for the Langevin SSM or a boundary point selected for the GP SSM. We also measure the difference between an initial contour and the same contour accepted after post editing, by the Hausdorff distance  $d_H$ . We take the mean  $N_{\text{int}}$  and  $d_H$  over 10 users and compare a tool using learned and normal priors. In all

cases the learned shape prior leads to improved useability, with significant improvement for synthetic liver tumours and all region types with the GP SSM. We also counted the number of times a user repeats an initialisation before accepting or editing a contour. The learned shape priors lead to a significant reduction in this number for all region types for both Langevin and GP SSMs. We measure accuracy by the similarity between a segmentation and the ground truth, using the overlap measure of the Dice similarity coefficient (DSC) and the the boundary-based similarity measure of the mean minimum distance between contours (MMD). In most cases the learned shape prior leads to improved useability. Both tools show significant improvement when segmenting MRI MS lesions and the GP tool shows significant improvement for synthetic liver tumours.

	Measure	Prior	Langevin SSM			Gaussian Process SSM		
			synth. LT	synth. MS	MRI MS	synth. LT	synth. MS	MRI MS
use.	$N_{\text{int}}$	normal	16.08±6.45	8.68±2.71	11.58±5.62	9.88±3.08	3.66±2.34	4.23±1.81
		learned	10.52±5.90*	8.18±3.90	10.25±3.93	4.93±1.96*	2.43±1.55*	2.73±1.30*
	$d_H$	normal	6.20±1.27	3.01±0.43	2.81±0.48	4.26±0.91	1.62±0.50	1.98±0.35
		learned	4.41±1.47*	2.76±0.73	2.77±0.70	3.69±0.74*	1.54±0.58	1.88±0.49
acc.	DSC	normal	0.96±0.01	0.87±0.02	0.80±0.02	0.93±0.01	0.85±0.01	0.76±0.04
		learned	0.96±0.01	0.88±0.02	0.80±0.02	0.95±0.01*	0.86±0.02	0.78±0.04*
	MMD	normal	0.77±0.13	0.63±0.15	1.23±0.04	1.37±0.13	0.73±0.08	1.47±0.05
		learned	0.79±0.13	0.63±0.11	1.15±0.05*	1.09±0.14*	0.70±0.11	1.38±0.07

**Table 1.** Differences in useability (use.) and accuracy (acc.) between segmentation with and without learned shape priors, where \* indicates significant improvement for the learned priors.

We also measure the inter- and intra-operator variability by the similarity between the contours created for the same region by one user on two occasions, and by two separate users. The learned shape priors do not reduce variability, which reflects the level of user control in post editing.

## 6 Conclusions and Future Work

We have designed mid-level, global shape priors for applications lacking high-level shape similarity or correspondence points, and shown them to improve the useability and accuracy of supervised contouring tools. In terms of segmentation variability, the benefits of the shape prior are overridden by the high level of user control in each framework means. Future work will extend the GP model to use novel kernels specific to different applications, and generalise the models for the case of arc-length parametrisation  $\{\mathbf{r}, \mathbf{s}\}$ . We also expect the discriminative models to benefit other tasks in medical image analysis. Equations (7) and (10) could be used to regularise general ACMs, image registration and reconstruction algorithms, or could be used in shape classification and object detection.

## References

1. T. F. Cootes & C. Taylor. “Active shape models – smart snakes.” In *Proceedings, BMVA Conference on British Machine Vision*, pp. 266–276. 1992.
2. D. Cremers, T. Kohlberger & C. Schnörr. “Shape statistics in kernel space for variational image segmentation.” *Pattern Recognition* **36**, pp. 1929–1943, 2003.
3. R. L. Kashyap & R. Chellappa. “Stochastic models for closed boundary analysis: Representation and reconstruction.” *IEEE Transactions on Pattern Analysis and Machine Intelligence* **27**, pp. 627–637, 1981.
4. A. H. Mir, M. Hanmandlu & S. N. Tandon. “Description of shapes in CT images.” *IEEE Engineering in Medicine and Biology* **18**, pp. 79–84, 1999.
5. B. Kartikeyan & A. Sarkar. “Shape description by time series.” *IEEE Transactions on Pattern Analysis and Machine Intelligence* **11**, pp. 977–984, 1989.
6. C. E. Rasmussen & C. K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.
7. T. Shepherd. *Dynamical Models and Machine Learning for Supervised Segmentation*. Ph.D. thesis, University College London, 2009.
8. R. Friedrich & J. Peinke. “Description of a turbulent cascade by a Fokker-Planck equation.” *Physical Review Letters* **78**, 1997.
9. P. Pérez, A. Blake & M. Gangnet. “Jetstream: Probabilistic contour extraction with particles.” In *Proceedings, IEEE International Conference on Computer Vision*, pp. 524–531. 2001.
10. T. Heimann & B. van Ginneken. “3D liver tumor segmentation challenge 2008, part of 3D segmentation in the clinic: A Grand Challenge II, at 11<sup>th</sup> international conference on Medical Image Computing and Computer Assisted Intervention.”, 2008.