found in incompressible inviscid fluids[5,19]. The smooth variation of the properties (such as size and shape[16], collective mode frequencies[24–26] and thermodynamics[27]) of dilute Bose gases as a function of γ, and the correspondence between our results and the behaviour of rotating liquid ^{4}He, strongly suggests that the symmetry-breaking phenomena we have described in the weak-coupling limit should occur over the entire range of γ in a qualitatively similar manner.

There are, however, notable quantitative differences. In the small-γ limit, we find that the core size and inter-vortex spacing are both comparable to the non-interacting ground-state width σ_r, and that the mean square radial (but not axial) dimension of the cloud grows linearly with l to accommodate more vortices. In the strong-coupling limit, however, the core size becomes comparable to the healing length $\xi \approx \sigma_r \gamma^{-1/5}$, which can be much smaller than the radial extent $R(0) \approx \sigma_r \gamma^{1/5}$ of the non-rotating cloud[16]. Under these conditions, the spacing between vortices is set by the condition[20] that the mean vorticity (that is, the vortex density) be equal to 2Ω. An extension of the Thomas–Fermi approach[16] to rapidly rotating gases in the large-γ limit then predicts that the radius of the cloud diverges as Ω approaches ω_r according to $R(\Omega) = R(0)\omega_{\rm eff}^{-3/5}$, where $\omega_{\rm eff} \equiv (\omega_r^2 - \Omega^2)^{1/2}$ is the effective trap frequency, taking centrifugal forces into account. With increasing Ω the condensate also flattens, and its axial:radial aspect ratio shrinks as $Z(\Omega)/R(\Omega) \approx \omega_{\rm eff}/\omega_r$, where $Z(\Omega)$ is the axial extent of the cloud. The angular momentum per particle diverges as $(\omega_{\rm eff}/\omega_r)^{-6/5}$. □

Methods

Using the Gross–Pitaevskii approach for trapped Bose gases[16], we consider variational condensates of the form

$$\Psi(\mathbf{r}) = \sum_{m>0} c_m \chi_m(\mathbf{r}) \qquad (1)$$

where the complex coefficients c_m are the probability amplitudes for finding a condensate atom in the low-energy angular-momentum eigenstates of the harmonic oscillator potential, $\chi_m(\mathbf{r}) = e^{im\phi} r^m e^{-[(r/\sigma_r)^2 + (z/\sigma_z)^2]/2} / (\pi^{3/2} m! \sigma_r^2 \sigma_z)^{1/2}$. Here $\sigma_i \equiv (\hbar/M\omega_i)^{1/2}$ for $i = r$ or z, M is the atomic mass, and ω_i is the oscillation frequency.

The angular momentum per particle in the state (1) is $l\hbar = \Sigma |c_m|^2 m\hbar$, and the kinetic plus trap potential energy per particle is:

$$E_{\rm ideal}[\Psi] = \sum |c_m|^2 m\hbar\omega_r = l\hbar\omega_r \qquad (2)$$

Thus the energy of a rotating non-interacting Bose–Einstein condensate depends only on its angular momentum l, and not on the detailed form of the superposition (1), indicating a large degeneracy[18].

In a real gas, interactions between the atoms break this degeneracy and select a particular linear combination to be the lowest energy state for each l. The Gross–Pitaevskii interaction energy per particle is:

$$E_{\rm int}[\Psi] \equiv \frac{2\pi\hbar^2 aN}{M} \int |\Psi(\mathbf{r})|^4 \, d^3\mathbf{r} \qquad (3)$$

We have numerically determined the complex amplitudes $\{c_m\}$ in equation (1) that minimize the total energy in the laboratory frame, $E_{\rm lab} = E_{\rm ideal} + E_{\rm int}$, subject to the constraint of fixed angular momentum per particle. Our calculations are exact in the small-γ limit, where the use of a single Gross–Pitaevskii condensate is equivalent to degenerate many-body perturbation theory at zero temperature (D.S.R., unpublished results). This result incorporates the effects of small symmetry-breaking perturbations.

The minimum value of $E_{\rm int}$ for given l can be written $\gamma\hbar\omega_r e_{\rm int}(l)$, where $e_{\rm int}(l)$ is dimensionless and depends only on the sign of γ for small γ. Then the angular velocity $\Omega(l) \equiv \partial E_{\rm lab}/\hbar\partial l = \omega_r(1 + \gamma \partial e_{\rm int}/\partial l)$. This function can be inverted to produce $l(\Omega)$, which in the weak-coupling limit depends only on $(\Omega - \omega_r)/\gamma$ (Fig. 2). We note that rotating gases expand (and hence become more dilute) with increasing l. Thus for positive γ the interaction energy decreases with increasing l, and we find that $\Omega(l) < \omega_r$. For negative γ, however, $\Omega(l) \geqslant \omega_r$ for $l \neq 0$, and centrifugal forces destabilize all rotating states.

Received 24 July; accepted 5 October 1998.

1. Leggett, A. J. in *Low Temperature Physics* (eds Hoch, M. J. R. & Lemmer, R. H.) 1–92 (Springer, Berlin, 1992).
2. Onsager, L. *Nuovo Cimento* **6,** 249–250 (1949).
3. Feynman, R. F. in *Progress in Law Temperature Physics* Vol. I (ed. Gorter, C. J.) 17–53 (North Holland, Amsterdam, 1955).
4. Vinen, W. F. Single quanta of circulation in helium II. *Proc. R. Soc. Lond. A* **260,** 218–236 (1961).
5. Yarmchuk, E. J., Gordon, M. J. V. & Packard, R. E. Observation of stationary vortex arrays in rotating superfluid helium. *Phys. Rev. Lett.* **43,** 214–217 (1979).
6. Ruutu, V. M. H. *et al.* Vortex formation in neutron-irradiated superfluid ^{3}He as an analogue for cosmological defect formation. *Nature* **382,** 334–336 (1996).
7. Anderson, M. H., Ensher, J. R., Matthews, M. R., Weiman, C. E. & Cornell, E. A. Observation of Bose-Einstein condensation in a dilute atomic vapor. *Science* **269,** 198–201 (1995).
8. Davis, K. B. *et al.* Bose-Einstein condensation in a gas of sodium atoms. *Phys. Rev. Lett.* **75,** 3969–3973 (1995).
9. Bradley, C. C., Sackett, C. A. & Hulet, R. G. Bose-Einstein condensation of litium: observation of limited condensate number. *Phys. Rev. Lett.* **78,** 985–989 (1997).
10. Inouye, S. *et al.* Observation of Feshbach resonances in a Bose-Einstein condensate. *Nature* **392,** 151–154 (1998).
11. Jin, D. S., Ensher, J. R., Matthews, M. R., Wieman, C. E. & Cornell, E. A. Collective excitations of a Bose-Einstein condensate in a dilute gas. *Phys. Rev. Lett.* **77,** 420–423 (1996).
12. Mewes, M. O. *et al.* Collective excitations of a Bose-Einstein condensate in a magnetic trap. *Phys. Rev. Lett.* **77,** 988–991 (1996).
13. Marzlin, K., Zhang, W. & Wright, E. M. Vortex coupler for atomic Bose-Einstein condensates. *Phys. Rev. Lett.* **79,** 4728–4731 (1997).
14. Dum, R. & Castin, Y. Creation of dark solitons and vortices in Bose-Einstein condensates. *Phys. Rev. Lett.* **80,** 2972–2975 (1998).
15. Landau, L. D. & Lifshitz, E. M. *Statistical Physics* 3rd edn (Pergamon, Oxford, 1980).
16. Baym, G. & Pethick, C. J. Ground-state properties of magnetically trapped Bose-condensed rubidium gas. *Phys. Rev. Lett.* **76,** 5–8 (1996).
17. Dalfovo, F. & Stringari, S. Bosons in anisotropic traps: ground state and vortices. *Phys. Rev. A* **53,** 2477–2485 (1996).
18. Wilkin, N. K., Gunn, J. M. F. & Smith, R. A. Do attractive bosons condense? *Phys. Rev. Lett.* **80,** 2265–2268 (1998).
19. Hess, G. B. Angular momentum of superfluid helium in a rotating cylinder. *Phys. Rev.* **161,** 189–193 (1967).
20. Hall, H. E. On the rotation of liquid helium II. *Adv. Phys.* **9,** 89–146 (1960).
21. Campbell, L. J. & Ziff, R. M. Vortex patterns and energies in a rotating superfluid. *Phys. Rev. B* **20,** 1886–1902 (1979).
22. Rokhsar, D. S. Vortex stability and persistent currents in trapped Bose gases. *Phys. Rev. Lett.* **79,** 2164–2167 (1997).
23. Rokhsar, D. S. Dilute Bose gas in a torus: vortices and persistent currents. Preprint cond-mat/9709212 at ⟨http://xxx.lanl.gov⟩ (1997).
24. Singh, K. G. & Rokhsar, D. S. Collective excitations of a confined Bose condensate. *Phys. Rev. Lett.* **77,** 1667–1670 (1996).
25. Edwards, M. *et al.* Collective excitations of atomic Bose-Einstein condensates. *Phys. Rev. Lett.* **77,** 1671–1674 (1996).
26. Stringari, S. Collective excitations of a trapped Bose-condensed gas. *Phys. Rev. Lett.* **77,** 2360–2363 (1996).
27. Mewes, M. O. *et al.* Bose-Einstein condensation in a tightly confining dc magnetic trap. *Phys. Rev. Lett.* **77,** 416–419 (1996).

Acknowledgements. We thank J. C. Davis, A. L. Fetter, R. E. Packard and D. P. Arovas for comments on the manuscript, and the Institute for Theoretical Physics at Santa Barbara for its hospitality while this work was being completed.

Correspondence and requests for materials should be addressed to D.S.R. (e-mail: rokhsar@physics.berkeley.edu).

Cluster-weighted modelling for time-series analysis

N. Gershenfeld*, B. Schoner* & E. Metois*†

* *Physics and Media Group, MIT Media Laboratory, Cambridge, Massachusetts 02139, USA*

The need to characterize and forecast time series recurs throughout the sciences, but the complexity of the real world is poorly described by the traditional techniques of linear time-series analysis. Although newer methods can provide remarkable insights into particular domains, they still make restrictive assumptions about the data, the analyst, or the application[1]. Here we show that signals that are nonlinear, non-stationary, non-gaussian, and discontinuous can be described by expanding the probabilistic dependence of the future on the past around local models of their relationship. The predictors derived from

† Present address: ARIS Technologies, Cambridge, Massachusetts 02140, USA.

this general framework have the form of the global combinations of local functions that are used in statistics[2–4], machine learning[5–10] and studies of nonlinear dynamics[11,12]. Our method offers forecasts of errors in prediction and model estimation, provides a transparent architecture with meaningful parameters, and has straightforward implementations for offline and online applications. We demonstrate our approach by applying it to data obtained from a pseudo-random dynamical system, from a fluctuating laser, and from a bowed violin.

The past decade has seen the introduction of many new techniques for modelling signals produced by complex systems, but the potential power of these methods is matched by the ease with which they can produce misleading or incorrect results[1]. We show here that it is possible to obtain the desirable features of many of these particular algorithms in a probabilistic setting that is more broadly and reliably applicable.

Non-recursive approaches to time-series analysis relate a time-dependent input feature vector $\mathbf{x}$ to an output observable $\mathbf{y}$. $\mathbf{x}$ might be the time-lag vector for an embedding that retrieves unseen internal degrees of freedom from an accessible observable, or perhaps a set of wavelet coefficients for a system with multiple timescales; $\mathbf{y}$ could be the future value of a series. The simplest model of their relationship uses linear coefficients to combine possibly nonlinear basis functions, $\mathbf{y}_n = \Sigma_{m=1}^{M} \boldsymbol{\beta}_m \mathbf{f}_m(\mathbf{x}_n)$, for example a global polynomial model. If the coefficients are moved inside the nonlinearity, $\mathbf{y}_n = \Sigma_{m=1}^{M} \mathbf{f}_m(\mathbf{x}_n, \boldsymbol{\beta}_m)$, exponentially fewer terms are typically needed to approximate the relationship to a given error, but an iterative search is needed to find good values for the coefficients. Increasing the model size ensures that there are many good local minima for the search to find, while over-fitting can be prevented by regularization, maximizing the agreement of the model with prior beliefs as well as with the data[13].

These insights into functional approximation and high-dimensional search lie behind the success of approaches such as neural networks and radial basis functions. However, the decision to fit a function can itself be restrictive. Global regularizers that maximize smoothness are inappropriate for describing local features that might be sharp. Large models can be difficult to interpret, and answer only questions for which they are trained. Although the appealing alternative of probability density estimation is conventionally thought to require impractical amounts of data for routine use, it becomes not only feasible but also easier to apply if the density is expanded around local models.

We will seek to find the joint density $p(\mathbf{y}, \mathbf{x})$ for the dependence of $\mathbf{y}$ on $\mathbf{x}$ from a set of experimental measurements $\{\mathbf{y}_n, \mathbf{x}_n\}_{n=1}^{N}$. This will be factored over clusters c_m, each containing the product of three terms:

$$p(\mathbf{y}, \mathbf{x}) = \sum_{m=1}^{M} p(\mathbf{y}, \mathbf{x}, c_m) = \sum_{m=1}^{M} p(\mathbf{y}|\mathbf{x}, c_m) p(\mathbf{x}|c_m) p(c_m) \quad (1)$$

Here $p(c_m)$ is the weight of a cluster, given by the fraction of the data that it explains; $p(\mathbf{x}|c_m)$ is the domain of influence in the input space of the cluster, and will be taken to be multivariate gaussian $N(\boldsymbol{\mu}_m, C_m)$ in terms of a mean $\boldsymbol{\mu}_m$ and covariance matrix C_m (or separable gaussians in a high-dimensional space for efficiency). The remaining item expresses the output behaviour of the cluster, and will taken to be gaussian with a functional dependence of the mean, $p(\mathbf{y}|\mathbf{x}, c_m) = N(\mathbf{f}(\mathbf{x}, \boldsymbol{\beta}_m), C_{y,m})$. The role of the function $\mathbf{f}(\mathbf{x}, \boldsymbol{\beta}_m)$ can be seen by calculating the expected value of $\mathbf{y}$ given $\mathbf{x}$;

$$\langle \mathbf{y}|\mathbf{x} \rangle = \int \mathbf{y} p(\mathbf{y}|\mathbf{x}) d\mathbf{y} = \int \mathbf{y} \frac{p(\mathbf{y}, \mathbf{x})}{p(\mathbf{x})} d\mathbf{y} = \frac{\Sigma_{m=1}^{M} \mathbf{f}(\mathbf{x}, \boldsymbol{\beta}_m) p(\mathbf{x}|c_m) p(c_m)}{\Sigma_{m=1}^{M} p(\mathbf{x}|c_m) p(c_m)} \quad (2)$$

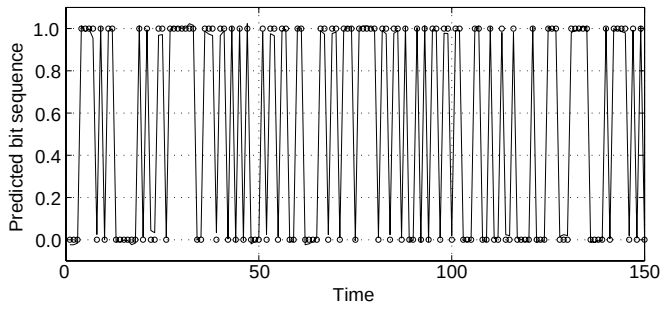

Figure 1 Forecasting a linear feedback shift register. Data from a 15-dimensional pseudorandom linear feedback shift register, and free-running predictions of a model trained on 10% of the repeat cycle (3,276 points, 20 clusters, 5,160 parameters, 10 expectation-maximization iterations). The model's initial condition was randomly perturbed by 20% from that of the digital shift register, showing that the correct sequence is a stable attractor.

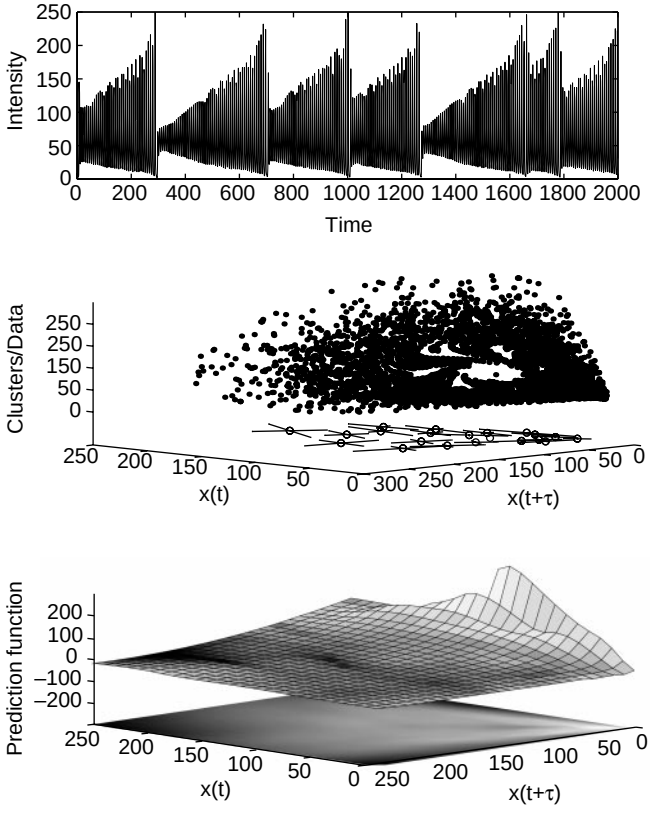

Figure 2 Modelling of a laser fluctuating near the gain threshold. Top, the time series. Middle, the data (dots) in a three-dimensional time-lag space and the resulting cluster means (circles) and covariances (lines). Bottom, the prediction surface derived from the resulting density, shaded by the conditional uncertainty in the prediction, plotted above the input density estimate. An animation of the convergence of this model is available at http://www.media.mit.edu/physics/publications/papers/cwm.

The prediction is a sum of local models, with the gaussians providing the nonlinear interpolation rather than serving as a basis for functional approximation. In the limit of a single cluster this reduces to a global version of the local model; as the number of clusters is increased, the density can handle deviations from the assumptions of the local model. The choice of **f** (or choices—more than one kind of function can be included) provides a means to explicitly build past practice, such as linear systems theory, into a domain.

A smooth local model expresses the belief that nearby points behave similarly, but as the clusters may be widely separated this architecture can also represent discontinuities. Because clusters are used only where there are data to explain, it is possible to describe low-dimensional structure in a high-dimensional space, and interpolation and extrapolation are well-behaved as they are done by weighted combinations of local models. As an example, a time series was generated by a 15-dimensional binary map:

$$x_n = x_{n-1} + x_{n-15} \pmod 2 \tag{3}$$

This is a maximal length shift register, generating ideal pseudo-random bits up to the repeat period of $2^{15} - 1$ samples. 10% of this full sequence (3,276 points) was used to build a model (as described below) in a 15-dimensional lag space. The free-running output from the resulting model using 20 linear covariance clusters is shown in Fig. 1. Not only is the exact sequence correctly reproduced, but it is also an attractor for randomly perturbed starting values[14].

The parameters for this example were found in just 10 iterations of an iterative expectation-maximization procedure[15]. The expectation step calculates the joint probabilities of data points and clusters, $p(\mathbf{y}_n, \mathbf{x}_n, c_m)$, and inverts this expression to find the posterior probabilities of the clusters given the data:

$$\begin{aligned} p(c_m|\mathbf{y}_n, \mathbf{x}_n) &= \frac{p(\mathbf{y}_n, \mathbf{x}_n|c_m)p(c_m)}{p(\mathbf{y}_n, \mathbf{x}_n)} \\ &= \frac{p(\mathbf{y}_n, \mathbf{x}_n|c_m)p(c_m)}{\Sigma_{l=1}^{M} p(\mathbf{y}_n, \mathbf{x}_n|c_l)p(c_l)} \end{aligned} \tag{4}$$

The clusters will interact through the sum in the denominator to specialize in data they best explain.

The maximization step then finds the most likely parameters given the posteriors. For the cluster weights this is defined by:

$$\begin{aligned} p(c_m) &= \int p(c_m|\mathbf{y}, \mathbf{x})p(\mathbf{y}, \mathbf{x})\mathrm{d}\mathbf{y}\mathrm{d}\mathbf{x} \\ &\approx \frac{1}{N}\sum_{n=1}^{N} p(c_m|\mathbf{y}_n, \mathbf{x}_n) \end{aligned} \tag{5}$$

The second step follows from the assumption that the experimental data were drawn from the joint density. Given $p(c_m)$, the cluster-weighted expectation of any function $\varphi(\mathbf{x})$ is defined to be:

$$\begin{aligned} \langle\varphi(\mathbf{x})\rangle_m &\equiv \int \varphi(\mathbf{x})p(\mathbf{x}|c_m)\mathrm{d}\mathbf{x} \\ &= \int \varphi(\mathbf{x})p(\mathbf{y}, \mathbf{x}|c_m)\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y} \\ &= \int \varphi(\mathbf{x})\frac{p(c_m|\mathbf{y}, \mathbf{x})}{p(c_m)}p(\mathbf{y}, \mathbf{x})\mathrm{d}\mathbf{x}\mathrm{d}\mathbf{y} \\ &\approx \frac{1}{Np(c_m)}\sum_{n=1}^{N}\varphi(\mathbf{x}_n)p(c_m|\mathbf{y}_n, \mathbf{x}_n) \\ &= \frac{\Sigma_{n=1}^{N}\varphi(\mathbf{x}_n)p(c_m|\mathbf{y}_n, \mathbf{x}_n)}{\Sigma_{n=1}^{N}p(c_m|\mathbf{y}_n, \mathbf{x}_n)} \end{aligned} \tag{6}$$

The apparently formal introduction of **y** in the second line lets the estimation proceed based on both the cluster location in the input space and how well its model performs in the output space. For online applications that require updating the estimates based on a new measurement without reanalysing the prior data, the clusters can be used to approximate the sum over previous points:

$$\begin{aligned} \langle\varphi(\mathbf{x})\rangle_m^{(N+1)} &= \frac{1}{(N+1)p(c_m)}\sum_{n=1}^{N+1}\varphi(\mathbf{x}_n)p(c_m|\mathbf{y}_n, \mathbf{x}_n) \\ &\approx \frac{Np(c_m)\langle\varphi(\mathbf{x})\rangle_m^{(N)} + \varphi(\mathbf{x}_{N+1})p(c_m|\mathbf{y}_{N+1}, \mathbf{x}_{N+1})}{Np(c_m) + p(c_m|\mathbf{y}_{N+1}, \mathbf{x}_{N+1})} \end{aligned} \tag{7}$$

The cluster-weighted expectation is used to find the cluster means and covariance matrices

$$\begin{aligned} \boldsymbol{\mu}_m &= \langle\mathbf{x}\rangle_m \\ [C_m]_{ij} &= \langle(x_i - \mu_i)(x_j - \mu_j)\rangle_m \end{aligned} \tag{8}$$

Taking the local models to have linear coefficients and maximizing the likelihood of the data results in a cluster-weighted regression for the model parameters[13];

$$\boldsymbol{\beta}_m = B_m^{-1}\cdot A_m \tag{9}$$

with $[B_m]_{i,j} = \langle f_i(\mathbf{x})f_j(\mathbf{x})\rangle_m$ and $[A_m]_{i,j} = \langle y_i f_j(\mathbf{x})\rangle_m$. Finally, the output covariance matrices are found from:

$$C_{y,m} = \langle[\mathbf{y} - \mathbf{f}(\mathbf{x}, \boldsymbol{\beta}_m)]\cdot[\mathbf{y} - \mathbf{f}(\mathbf{x}, \boldsymbol{\beta}_m)]^{\mathrm{T}}\rangle_m \tag{10}$$

Here superscript T indicates the matrix transpose. Iterating the expectation and maximization steps leads to the most likely arrangement of clusters that can be reached from the initial conditions. As there are many nearly equivalent solutions, the final likelihood does not depend sensitively on the initialization. For this reason, simple out-of-sample cross-validation[16] instead of full bayesian Monte-Carlo Markov chain integration[17] is used to determine the single hyper-parameter, M, the number of clusters.

Once the density has been found, other quantities of interest can be derived. The error in the forecast is given by the expected output

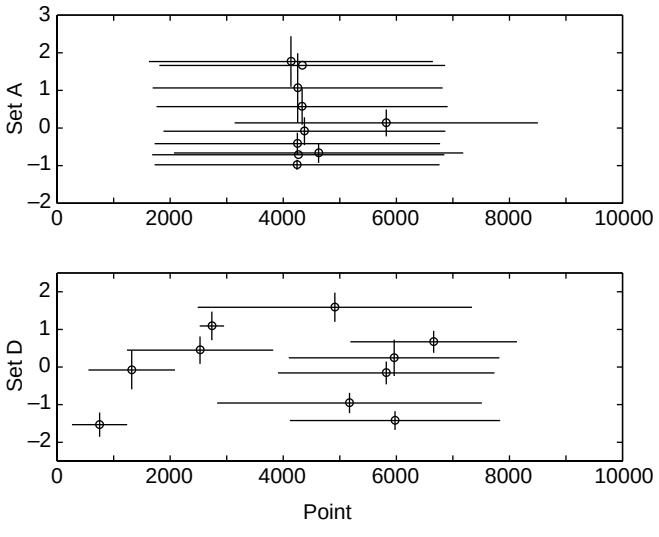

Figure 3 Clustering stationary and non-stationary data. Top, two-dimensional projection of cluster means and variance for stationary data, a fluctuating laser (set A from ref. 1), using time as one input degree of freedom. Bottom, clustering for non-stationary data, a particle in a drifting multiple-well potential (set D from ref. 1). For stationary data, the clusters expand in time to encompass all of the data; for non-stationary data, they shrink down to the local scale that maximizes the overall likelihood.

covariance matrix:

$$\langle C_{y,m}|\mathbf{x}\rangle = \frac{\Sigma_{m=1}^{M}[C_{y,m} + \mathbf{f}(\mathbf{x},\boldsymbol{\beta}_m)\cdot\mathbf{f}(\mathbf{x},\boldsymbol{\beta}_m)^{\mathrm{T}}]p(\mathbf{x}|c_m)p(c_m)}{\Sigma_{m=1}^{M}p(\mathbf{x}|c_m)p(c_m)} - \langle\mathbf{y}|\mathbf{x}\rangle^2 \tag{11}$$

This gives the width of the output distribution, without assuming gaussianity because multiple clusters can overlap. During training, the agreement between the predicted and measured error also provides a self-consistency check on the validity of the model. The availability of the input density estimate

$$p(\mathbf{x}) = \sum_{m=1}^{M} p(\mathbf{x}|c_m)p(c_m) \tag{12}$$

provides complementary information, showing where the prediction uncertainty is large because few data were available to estimate the model. In the limit of gaussian forecasting errors, the log of the output uncertainty is equal to the source entropy, and hence the sum of positive Lyapunov exponents. The radial correlation integral of the density can be calculated analytically as well, giving the correlation dimension of the density[13].

Figure 2 shows these derived quantities for a familiar experimental data set, the intensity of an NH_3 laser fluctuating near the gain threshold[18]. The conditional uncertainty associated with the divergence near the reinjection of the oscillator is readily apparent. An animation of the rapid convergence of this model is available (see Fig. 2 legend), showing prediction errors comparable to those obtained by competing techniques in ref. 1.

Because the clusters grow or shrink to maximize the overall model likelihood, they can be used to determine the local relevance of input degrees of freedom. Figure 3 shows the variance of clusters for data sets that are stationary (the laser) and non-stationary (the trajectory of a particle in a drifting multiple-well potential), using time as an input degree of freedom. For the stationary case they expand to encompass the whole data set; for the non-stationary case they shrink to an appropriate local scale without needing to define a global factor for weighting the past.

As a final example showing the modelling of a complex driven nonlinear system, Fig. 4 plots time-series data recorded from a violin, with the output audio as well as the relevant player inputs (bow and finger positions) used to perform an input–output embedding[19,20]. A spectral representation was used that tracks the frequency and amplitude of harmonics, in order to ignore the perceptually less-relevant phase information in modelling the underlying process. The resulting model can reproduce the response of the violin both in-sample and with new player input (see Fig. 4 legend), providing a computationally efficient alternative to first-principles physical modelling and sound sampling.

Cluster-weighted modelling is both old and new. It is a simple special case of the general theory of probabilistic networks[21,22], but one that can handle most of the limitations of practical data sets without unduly constraining either the data or the user. This architecture is appropriate for stationary systems described by a single global model, or non-stationary systems described by unrelated local models. For the more complex case of a multi-stationary or long-memory process, it is necessary to introduce internal degrees of freedom into the model. Without *a priori* insight into the model architecture, this presents a difficult problem of architecture selection; an open question is whether it is possible to solve this problem through the kind of probabilistic factoring and sampling that we have shown here to be so convenient for determining model allocation.

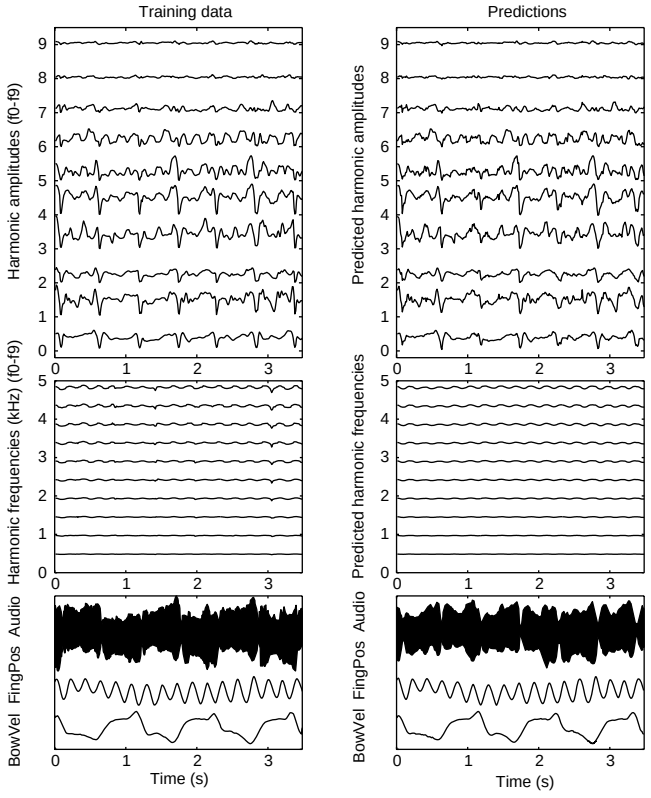

Figure 4 Modelling of a bowed violin. Left, measurements on a violin, showing the bow velocity, player's finger position, resulting audio time series and harmonic structure. Right, the audio re-synthesized from the sensor data by a model trained in the joint input–output space. Audio samples are available at http://www.media.mit.edu/physics/publications/papers/cwm

Received 10 July; accepted 12 October 1998.

1. Weigend, A. S. & Gershenfeld, N. A. (eds) *Time Series Prediction: Forecasting the Future and Understanding the Past* (Santa Fe Inst. Studies in the Sciences of Complexity, Addison-Wesley, Reading, MA, 1993).
2. Cleveland, W. S. & Devlin, S. J. Regression analysis by local fitting. *J. Am. Stat. Assoc.* **83,** 596–610 (1988).
3. Wand, M. P. & Jones, M. C. *Kernel Smoothing* (Chapman & Hall, London, 1995).
4. Fan, J. & Gijbels, I. *Local Polynomial Modelling and Its Applications* (Chapman & Hall, London, 1996).
5. Fan, J. & Gijbels, I. Data-driven bandwidth selection in local polynomial fitting: variable bandwidth and spatial adaptation. *J. R. Stat. Soc. B* **57,** 371–394 (1995).
6. Jordan, M. I. & Jacobs, R. A. Hierarchical mixtures of experts and the EM algorithm. *Neural Comput.* **6,** 181–214 (1994).
7. Weigend, A. S., Mangeas, M. & Srivastava, A. N. Nonlinear gated experts for time series: discovering regimes and avoiding overfitting. *Int. J. Neural Syst.* **6,** 373–399 (1995).
8. Ghahramani, Z. & Jordan, M. I. in *Advances In Neural Information Processing Systems* Vol. 6 120–127 (MIT Press, Cambridge, MA, 1995).
9. Lei, X., Jordan, M. I. & Hinton, G. E. in *Advances in Neural Information Processing Systems* Vol. 7 633–640 (MIT Press, Cambridge, MA, 1995).
10. Waterhouse, S., MacKay, D. & Robinson, T. in *Advances in Neural Information Processing Systems* Vol. 8 351–357 (MIT Press, Cambridge, MA, 1996).
11. Farmer, J. D. & Sidorowich, J. J. Predicting chaotic time series. *Phys. Rev. Lett.* **59,** 845–848 (1987).
12. Sauer, T. in *Time Series Prediction: Forecasting the Future and Understanding the Past* (eds Weigend, A. S. & Gershenfeld, N. A.) 175–193 (Santa Fe Inst. Studies in the Sciences of Complexity, Addison-Wesley, Reading, MA, 1983).
13. Gershenfeld, N. A. *The Nature of Mathematical Modeling* (Cambridge Univ. Press, 1999).
14. Gershenfeld, N. A. & Grinstein, G. Entrainment and communication with dissipative pseudorandom dynamics. *Phys. Rev. Lett.* **74,** 5024–5027 (1995).
15. Dempster, A. P., Laird, N. M. & Rubin, D. B. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. B* **39,** 1–38 (1977).
16. Wahba, G. & Wold, S. A. Completely automatic French curve: fitting spline functions by cross validation. *Commun. Stat.* **4,** 1–17 (1975).
17. Richardson, S. & Green, P. J. On Bayesian analysis of mixtures with an unknown number of components. *J. R. Stat. Soc. B* **59,** 731–792 (1997).
18. Hüebner, U., Weiss, C.-O., Abraham, N. & Tang, D. in *Time Series Prediction: Forecasting the Future and Understanding the Past* (eds Weigend, A. S. & Gershenfeld, N. A.) 73–105 (Santa Fe Inst. Studies in the Sciences of Complexity, Addison-Wesley, Reading, MA, 1993).
19. Casdagli, M. in *Nonlinear Modeling and Forecasting* (eds Casdagli, M. & Eubank, S.) 265–281 (Santa Fe Inst. Studies in the Sciences of Complexity, Addison-Wesley, Redwood City, CA, 1992).
20. Stark, J., Broomhead, D. S., Davies, M. E. & Huke, J. Takens embedding theorems for forces and stochastic systems. *Nonlinear Anal.* **30,** 5303–5314 (1997).
21. Buntine, W. L. Operations for learning with graphical models. *J. Artif. Intelligence Res.* **2,** 159–225 (1994).
22. Smyth, P., Heckerman, D. & Jordan, M. I. Probabilistic independence networks for hidden Markov probability models. *Neural Comp.* **9,** 227–269 (1997).

Acknowledgements. We thank C. Douglas, C. Cooper, R. Shioda and E. Boyden for help with the collection and analysis of the violin data. This work was supported by the Things That Think consortium of MIT Media Laboratory.

Correspondence and requests for materials should be addressed to N.G.