Parameter uncertainty of chaotic systems

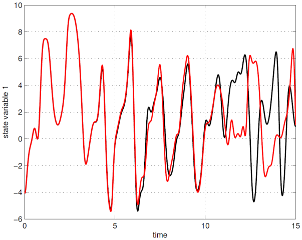
H. Haario, S. Springer, V. Shemyakin

Lappeenranta University of Technology, Finnish Meteorological Institute

DSABNS2018, Torino, 08.02.2018

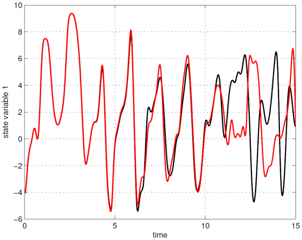
Index of the presentation

Chaotic Systems



After a predictable interval, any changes (of initial values, model parameters, solver settings) lead to unpredictable deviations. Options:

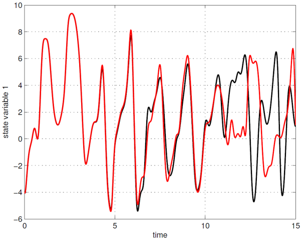
Chaotic Systems



After a predictable interval, any changes (of initial values, model parameters, solver settings) lead to unpredictable deviations. Options:

- **Avoid chaos**: deal with predictable time intervals only (Weather)
- **Face it**: deal with behaviour after predictability (Climate).

Chaotic Systems

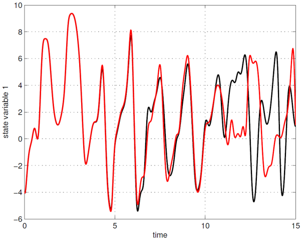


After a predictable interval, any changes (of initial values, model parameters, solver settings) lead to unpredictable deviations. Options:

- **Avoid chaos**: deal with predictable time intervals only (Weather)
- **Face it**: deal with behaviour after predictability (Climate).

- **How to create cost functions for parameters of a chaotic system?**

Chaotic Systems



After a predictable interval, any changes (of initial values, model parameters, solver settings) lead to unpredictable deviations. Options:

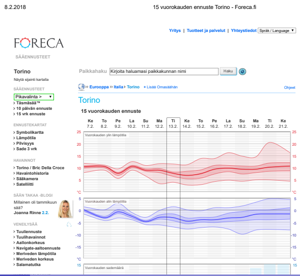
- **Avoid chaos**: deal with predictable time intervals only (Weather)
- **Face it**: deal with behaviour after predictability (Climate).

- **How to create cost functions for parameters of a chaotic system?**
- **Can we distinguish chaotic variability of a fixed system from systematic change between different systems?**

State estimation: NWP + UQ

Numerical Weather Prediction (NWP) and Uncertainty Quantification (UQ) by an Ensemble Prediction System (EPS) in a nutshell:

- NWP: update the initial values of the system by recent data (in the 'assimilation window'), then simulate a few days ahead. Use variants of 4DVAR.
- An ensemble of simulations run by perturbed initial values (50 members, by ECMWF, Reading, UK). Only for UQ, not for assimilation.



Parameter estimation, standard methods: filter likelihood

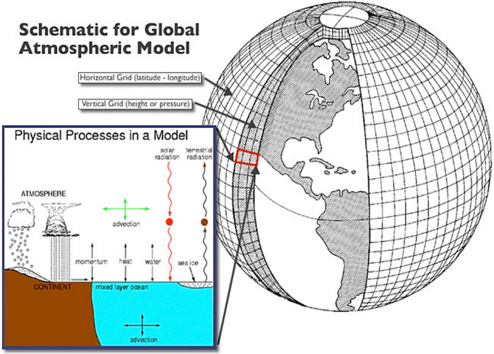
- Using Kalman filter, integrate out the state space, what remains gives a likelihood for parameters.
- Standard way for linear time series (DLM, Dynamical Linear Models) and SDE (stochastic differential equations) systems. Less standard for chaotic dynamics, but can be implemented with EKF.
- BUT:
  - Not applicable to large scale systems, due to memory and CPU issues.
  - Each filter algorithm has built-in 'tuning parameters' (model error covariance, linearization ...). The amount of bias introduced by them ?
  - Only for 'short' assimilation time integrations.

Ensemble-based heuristic algorithms available.

Ollinaho, P., Bechtold, P., Leutbecher, M., Laine, M., Solonen, A., Haario, H., and Järvinen, H.: *Parameter variations in prediction skill optimization at ECMWF*, Nonlin. Processes Geophys., 20, 6,1001-1010, 2013.

Long time: summary statistics for climate models ?

In order to find 'typical' behaviour of a chaotic climate system, observations and simulations may be averaged in space and time to create 'summary statistics'.



Long time: summary statistics for climate models ?

- If the statistics of the summary expression is known, a likelihood is formulated which yields the posterior for the model parameters.

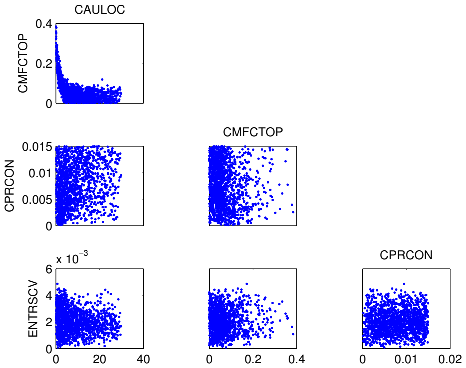
Long time: summary statistics for climate models ?

- If the statistics of the summary expression is known, a likelihood is formulated which yields the posterior for the model parameters.

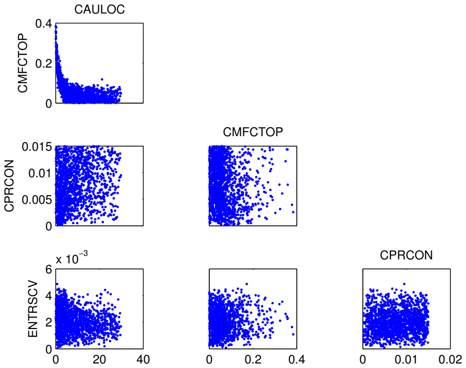
- Example: the approach was implemented for the ECHAM5 climate model, using likelihoods based on monthly global and zonal net radiation averages.
- MCMC was used to estimate four parameters related to cloud formation and precipitation. Technically possibly but...

  Järvinen, H., Räisänen, P., Laine, M., Tamminen, J., Ilin, A., Oja, E., Solonen, A., and Haario, H.: *Estimation of ECHAM5 climate model closure parameters with adaptive MCMC*, Atmos. Chem. Phys., Vol. 10, nro. 2, 9993-10002, 2010.

Example: climate model MCMC results



Example: climate model MCMC results



- Direct, naive summary statistics (projections) do not identify the parameters, i.e., characterise the simulated trajectories.

Parameter estimation beyond predictability? Example: 3D Lorenz parameters

$$\frac{dX}{dt} = \sigma(Y - X), \quad \frac{dY}{dt} = X(\rho - Z) - Y, \quad \frac{dZ}{dt} = XY - \beta Z. \quad (1)$$

Fit and sample model parameters when

- no good initial guess of parameters available
- no initial values of the state (X, Y, Z) is known
- observations only of (X, Y), noisy and sparse in time, beyond the time interval of predictability: none of the short-time methods available.

Example: 3D Lorenz, Data

Initial part of a long time series:

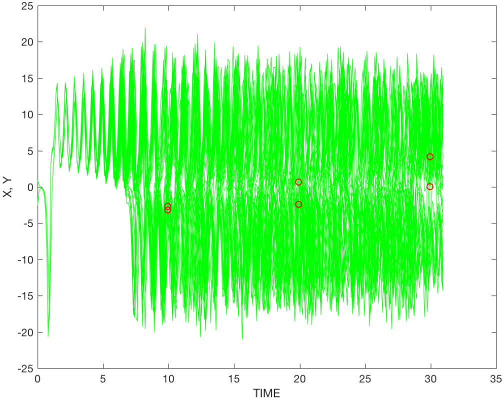


Figure: Observation samples (the red circles) for 3D Lorenz.

Optimizing by evolution algorithms

To find the MAP of the likelihood (coming below), use the Differential Evolution (DE) 'ensemble/population'

- Create ensembles of simulations as 'generations' of a population for a genetic optimisation algorithm.
- We employ the Differential Evolution algorithm
  - Mutation: add scaled differences of ensemble vectors to the present ones
  - Crossover: random survival of mutations
  - Selection: survival of improvements

Choose DE, as it works for a stochastic cost function, is robust against values leading outside the domain of attraction, and provides an initial proposal for adaptive MCMC.

Example: 3D Lorenz, Convergence

Convercenge of the DE optimization:

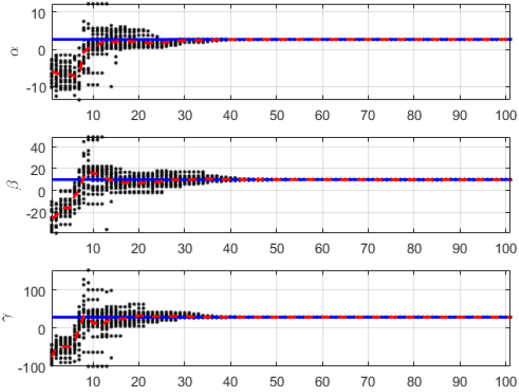
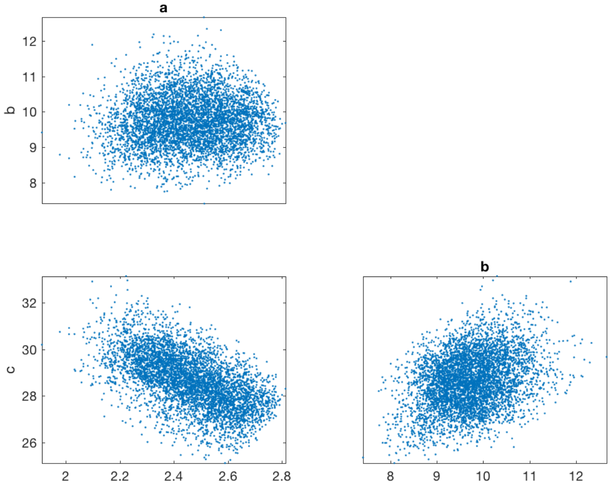


Figure: Blue: truth. Red: mean of ensemble values

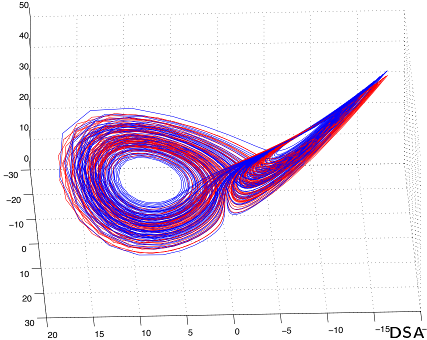
Example: 3D Lorenz, parameter posterior

The MCMC samples for the model parameters



Example: 3D Lorenz, verification

A trajectory simulated with model parameters slightly outside the sampled posterior (red) vs inside (blue).



Likelihood based on fractal concepts

Fractal dimensions of chaotic attractors, such as the Hausdorff dimension or box-counting, approximate the **internal properties** of the underlying attractor via numerically simulated trajectories. How to employ them to **define a distance between chaotic trajectories**?

We want to separate the **model variability due to initial values etc, but with fixed model parameters** from that **due to different model parameters**.

Key idea: interprete the **time-varying, chaotic** trajectories as samples from a **fixed** attractor.

Idea of Likelihood

Simulations of a model give samples from the underlying 'strange' attractor.

Create a training set of simulations – or one long enough time series – to characterize statistically the variability of the points, to define a likelihood for the 'internal' variability.

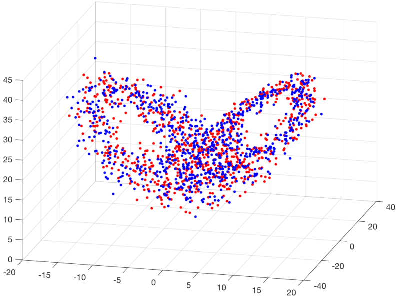


Figure: Two sets of samples, with given number (N=800) of points

Correlation dimension for fractal sets

Denote by $s_i, i = 1, 2, ..., N$ points of a trajectory vector $s \in R^n$, evaluated at time points t_i. For $R > 0$ set

$$C(R, N) = 1/N^2 \sum_{i,j} \#(||s_i - s_j|| < R)$$

and define then the correlation integral as the limit $C(R) = \lim_{N \to \infty} C(R, N)$. So we take the total number of points closer than R, normalize by the number of pairs N^2 and take the limit. Not that for each N we have $1/N \leq C(R, N) \leq 1$.

If ν is the dimension of the trajectory, we should have

$$C(R) \sim R^{\nu}$$

and the Correlation Dimension ν is defined as the limit

$$\nu = \lim_{R \to 0} \log C(R) / \log(R).$$

Distance via a generalized correlation sum

- Fix a radius R_0, large enough for each ball $B(s_i, R)$ to contain all the points $s_j, j \neq i$
- Select smaller radii by $R_k = b^{-k} R_0$, with $k = 1, 2, ..., M$. Select M and the base b (e.g., $M = 10$, $b = 2$).

The generalized correlation vector $y = (y_k), k = 1, ..., M$, between trajectories $s = s(\theta, x)$ and $\tilde{s} = s(\tilde{\theta}, \tilde{x})$ is given by

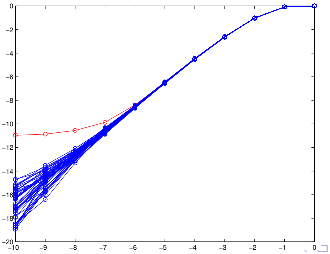
$$y_k = C(R_k, N, \theta, x, \tilde{\theta}, \tilde{x}) = 1/N^2 \sum_{i,j} \#(||s_i - \tilde{s}_j|| < R_k), \tag{2}$$

where $\theta, \tilde{\theta}$ denote the respective model parameters and $x, \tilde{x}$ the initial values. For $\tilde{\theta} = \theta, \tilde{x} = x$ the formula reduced to the original definition of the correlation sum.

Likelihood by Correlation Vector

First, characterize the 'within variability' of a fixed chaotic dynamical system

- Create an ensemble of point sets from the attractor (subsamples of a time series, or simulated values $s = s(\theta_0, x)$ if θ_0 known).
- Compute the distance matrix between (all) different trajectory pairs, to get the values y_k .
- The stochastic vector $(y_k), k = 1, ...M$ (the empirical CDF of distances) turns out to be **Gaussian** (by CLT, Donsker's theorem, etc)



Example: Likelihood for 3D Lorenz

- Test the Gaussianity of values $y_k = C(R_k, N, \theta_0, x, \theta_0, \tilde{x})$, by the usual χ^2 test: calculate the mean value μ_0 and covariance matrix Σ_0 of the training set.
- The statistics of the expression $(\mu_0 - y)\Sigma_0^{-1}(\mu_0 - y)$ should obey the χ^2_M distribution for a Gaussian y,

$$(\mu_0 - y)\Sigma_0^{-1}(\mu_0 - y) \sim \chi^2_M \tag{3}$$

Example: Likelihood for 3D Lorenz

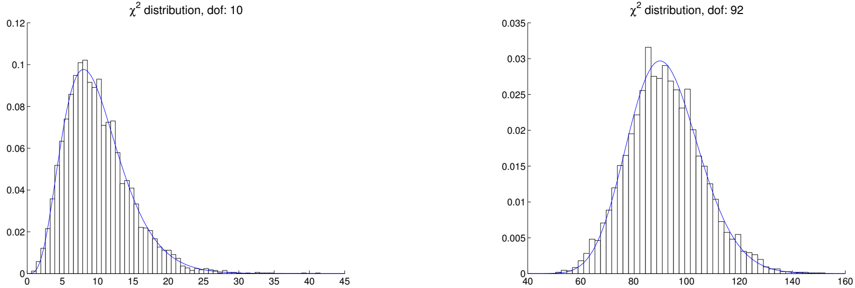


Figure: **Normality check** of the correlation integral vector by the χ^2 test for the Lorenz 63 system. Left: with 10 radius values used. Right: with 92 radius values

Likelihood by Correlation Vector

We treat the above vectors $y = C(R_k, N, \theta_0, x, \theta_0, \tilde{x}), k = 1, ..., M$ as 'measurements' of the variability of a chaotic trajectory with a given fixed model parameter. Construct the respective likelihood:

- Obtain $y = C(R, N, \theta_0, x, \theta_0, \tilde{x})$ from repeated simulations (or, get them from a long enough empirical time series).
- Create the empirical likelihood function: compute **mean and covariance.**

For any other parameter θ and trajectory $s(\theta)$ compute the distance matrix from the reference trajectory, and the respective $C(R_k, N, \theta, x, \tilde{\theta}, \tilde{x})$, to evaluate the likelihood for θ.

HH, Leonid Kalachev, Janne Hakkarainen *Generalized Correlation integral vectors: A distance concept for chaotic dynamical systems*. Chaos, 25, 2015.

Inference as a pseudo-marginal MCMC algorithm

Due to chaoticity and randomised x the likelihood is non-deterministic. But sampling from can be interpreted as sampling from the joint distribution of the initial values and model parameters.

Denote the likelihood function of y, evaluated for an arbitrary θ by $T_{\theta_0}(\theta, x)$. The target distribution for θ is given as

$$\pi(\theta) = \int T_{\theta_0}(\theta, x)\lambda(x)\mathrm{d}x,$$

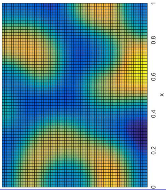
where $\lambda(x)$ is the distribution of the initial values x. In our situation, $T_{\theta_0}(\theta, x)$ is unknown, but an empirical approximation can be created as above.

The method is a bivariate Markov chain: $(\theta_n, T_n)_{n\geq 0}$, where T_n are auxiliary variables that are non-negative, unbiased estimators of the underlying intractable target density $\pi(\theta_n)$. So we use a **pseudo-marginal algorithm** targeting π.

Other examples

Similar results for

- 3D: Rössler equation, Chua3/5/7, ChenLee,Genesio,YuWang,Bouali,TSUCS1/2,Wang,Aizawa,... together with 'stochastic physics' and randomized parameters.
- Lorenz95 (dimension 42 or 210)
- Kuramoto-Shivashinsky (1D chaotic PDE, GPU implementation)
- Shallow water (high dimensional, GPU implementation)
- FitzHugh-Nagumo pattern formation (non-chaotic PDE, patterns by random initial values):



Example: the Aizawa attractor

The Aizawa equations:

$$\begin{cases} dX &= (Z-e)*X - d*Y; \\ dY &= d*X + (Z-e)*Y; \\ dZ &= c + b*Z - Z^3/3 - (X^2+Y^2)*(1+a*Z) + f*Z*X^3. \end{cases}$$

$$\begin{cases} X_0 &= 0.1; \\ Y_0 &= 0.1; \\ Z_0 &= 0.1. \end{cases}$$

Where:

- The model state space is $(X, Y, Z) \in \mathbf{R}^3$;
- The model initial state is $(X_0, Y_0, Z_0) \in \mathbf{R}^3$;
- The model parameters are $(a = 0.25, b = 0.95, c = 0.6, d = 3.5, e = 0.7, f = 0.6)$.

The posterior distribution

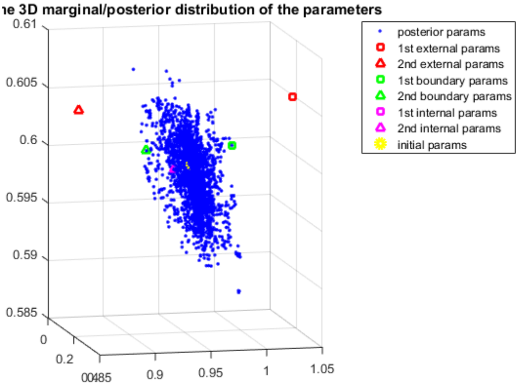


Figure: The posterior distribution of the Aizawa attractor, created with an Adaptive Monte Carlo (AM) method

Different Aizawa attractors

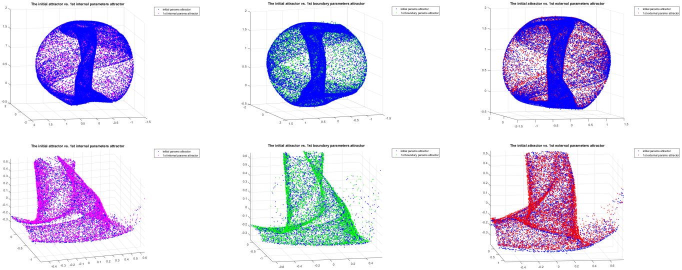


Figure: Visual comparison of Aizawa attractors, using the marked parameters of the posterior distribution

Kuramoto-Shivashinsky

Find the parameters that produce the 'same' attractor approximation for the system

$$u_t = -uu_x - \eta u_{xx} - \gamma u_{xxxx}. \tag{4}$$

An example result:

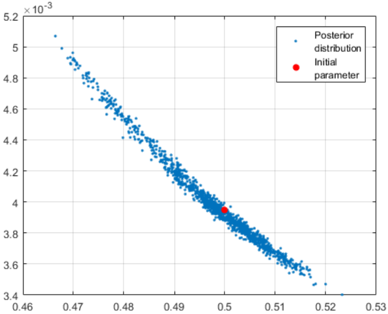


Figure: The posterior of the KS parameters

Kuramoto-Shivashinsky

Two same, two different (!) solutions, with parameters inside/outside the sampled posterior

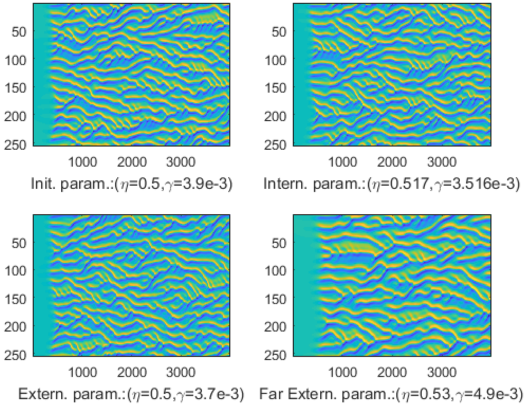


Figure: Top/bottom row: simulations with parameters inside/outside posterior

Summary

- 'Naive' linear summary statistics did not work
- The distance statistics likelihood can distinguish quite small changes

Our summary statistics loses information, too: the order of trajectory points, and dynamics. Ongoing topics:

- Add dynamics: employ both the state and the time derivative of it.
- Stochastic differential equations (SDE); also with 'stochastic physics' and 'stochastic parametrizations' as used in NWP models.
- High dimensions: parallel simulations, parallel distance calculations.

References, thanks to collaborators:

- Järvinen H, Laine M, Solonen A, Haario H. *Ensemble prediction and parameter estimation system: the concept*. Q. J. R. Meteorol. Soc. Vol. 138, nro. 663, pp. 281-288, 2012.
- Laine M, Solonen A, Haario H, Järvinen H. *Ensemble prediction and parameter estimation system: the method*. Q. J. R. Meteorol. Soc. Vol. 138, nro. 663, pp. 289-297, 2012.
- Ollinaho P, Laine M, Solonen A,Haario H, Järvinen H.:*NWP model forecast skill optimization via closure parameter variations*. Q. J. R. Meteorol. Soc., 139, 675, pp. 1520-1532, 2013.
- Ollinaho, P., Bechtold, P., Leutbecher, M., Laine, M., Solonen, A., Haario, H., and Järvinen, H.: *Parameter variations in prediction skill optimization at ECMWF*, Nonlin. Processes Geophys., 20, 6,1001-1010, 2013.
- Ollinaho, P., Järvinen, H., Bauer, P., Laine, M., Bechtold, P., Susiluoto, J., and Haario, H.: *Optimization of NWP model closure parameters using total energy norm of forecast error as a target*, Geosci. Model Dev., 7, 1889-1900, doi:10.5194/gmd-7-1889-2014, 2014.
- Vladimir Shemyakin, Heikki Haario: *Optimizing Ensemble Prediction System parameters with Differential Evolution* Submitted.

- Järvinen, H., Räisänen, P., Laine, M., Tamminen, J., Ilin, A., Oja, E., Solonen, A., and Haario, H.: *Estimation of ECHAM5 climate model closure parameters with adaptive MCMC*, Atmos. Chem. Phys., Vol. 10, nro. 2, 9993-10002, 2010.
- J. Hakkarainen, A. Ilin, S. Solonen, M. Laine, H. Haario, J. Tamminen, E. Oja, H. Järvinen: *On Closure Parameter Estimation in Chaotic Systems*. Nonlin. Processes Geophys., 19,127–143, 2012.
- **Heikki Haario, Leonid Kalachev, Janne Hakkarainen** *Generalized Correlation integral vectors: A distance concept for chaotic dynamical systems*. **Chaos, 25, 2015.**