An overview of statistical methods

How does it fit all together?

* Background Review: introduction of fundamental statistical measure (e.g. PDF, moments of a PDF, JPDF and moments of JPDF, random variables and function of random variable, central limit theorem.)

⇓

the goal is to provide us basic statistical descriptions of the system that we are intrested in.

Why are PDF usefull? ~~Bayesian estimation~~ →

* ~~combining model and observations~~

empirical

least squares typically used to fit model

inverse modeling

adjoint modeling = a systematic approach to explore model sensitivity

statistical + errors

covariance modeling = based on linear relationships

multiscale covariance

- combining models and observations

1) "EMPIRICAL" (e.g. system of linear equations) with parameters: $y = ax$ or $y = a\cos(\omega t)$ ← you choose the model

↓

least square typically used to fit model to observations and computer aposteriori errors.

2) "DYNAMICAL" (eg. 1) an atmospheric general circulation model ← you choose the dynamics
2) eq. describing the ~~[illegible]~~ evolution of seismic waves)

with parameters: initial and boundary conditions, convective parametrization, coefficients of the stress tensor, ... in general anything that you think is uncertain in the model that needs to be "constrained" by the available information.

↓

again least square tecnique or typically used to fit model to observations —

↓

fitting the model requires varying the parameters and understanding how the model responds to these changes "in a linear way" → 1) concept of model sensitivity
2) variational data assimilation
3) inverse modeling / adjoint modeling

3) "STATISTICAL MODEL" : ~~no~~ the core of the model is based on statistics computed from the data.

⇓

most simple statistic is based on linear relationships and involves a covariance or multiple covariances. → regressions or multivariate regressions

⇓

The linear relationship is typicaly used in interpolation → space → extrapolation
↘ time → forecasting

How to build a covariance?

Models build on covariance often lead to the so called "statistical optimal estimators"

CASE 1: ⟷ CASE 2:

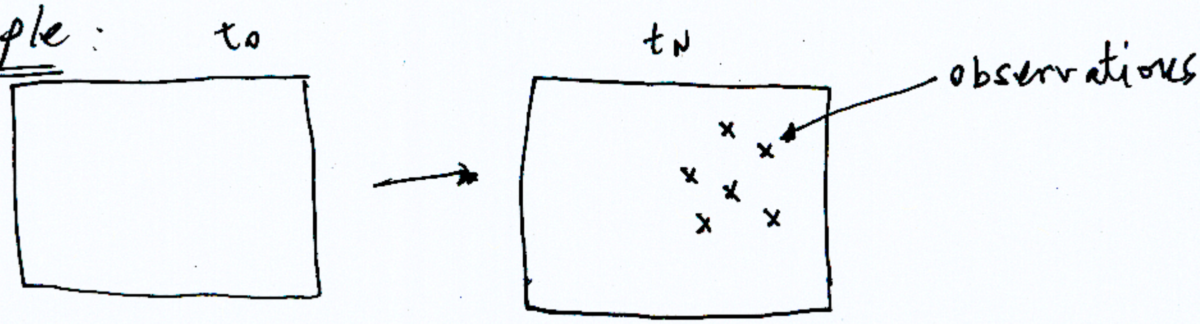
you can compute the covariance from the data

you ~~have~~ have to assume the shape of the covariance using other information or assumptions

"EMPIRICAL" and "DYNAMICAL" models provide such covariance information!

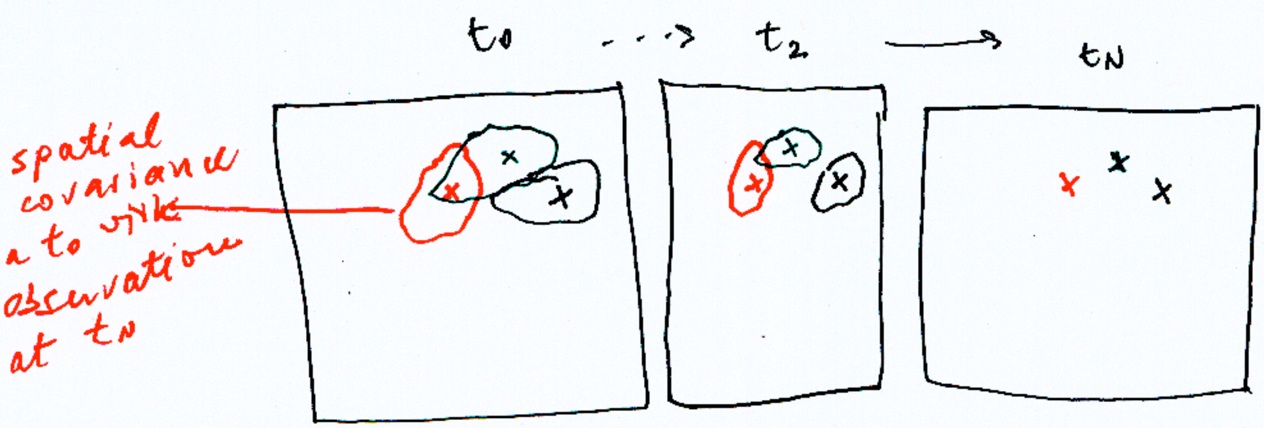
NOTE: when using "EMPIRICAL" and "DYNAMICAL" models statistics are still needed to constraint the unobserved degrees of freedom.

example:

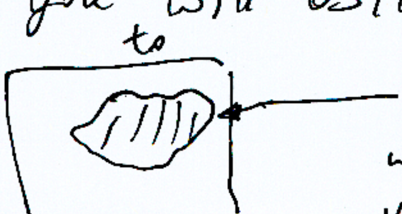


1) Assume you know the dynamics that go from $t_0 \to t_N$
2) you have a set of TEMPERATURE OBSERVATIONS at t_N and would like to know the temperature field at t_0.

3) Assume that using the model "dynamics" you can retrive the optimal covariance between the observations and the temperature field at t_0



Do the same for each point observation. At the end you will obtain a map



where is the area at t_0 for which observations at t_N provide us information for.

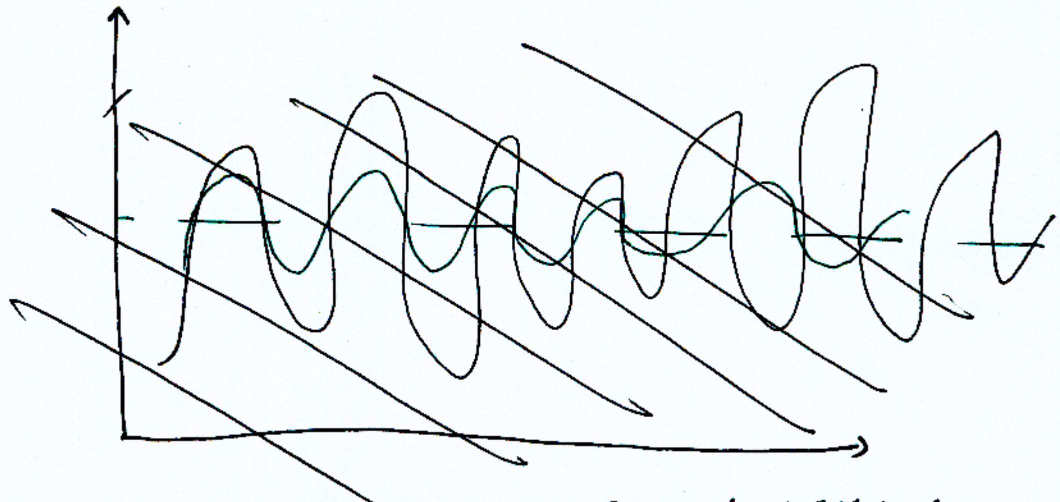
This means that the observations at t_N provide $\emptyset$ information (= carry $\emptyset$ relationship) with the TEMPERATURE field at t_0 outside this area.

All the TEMP. points at t_0 are the degrees of freedom → this implies that we have a large number of unobserved degree of freedom for the system (= observations do not provide sufficient information to constrain them).

So what do you do?

Let us look at "EMPIRICAL MODELS" again:

FIGURE 1



consider a time series of TEMPERATURE and assume you want to remove the seasonal cycle. One can do this by fitting the following function.

$$y = a_0 + b \sin(\omega t) + c \cos(\omega t)$$

Once the fit is done the variance of the seasonal cycle is

$$\frac{b^2 + c^2}{2} = \text{variance of seasonal cycle}$$

In principal one could do the same exercize for each frequency and study the variance explained by each of those components $\longrightarrow$ SPECTRA

$\Downarrow$

spectral analysis.

SPECTRAL analysis is an example of decomposing a signal on a different basis, in this case the sin / cos functions → FOURIER SERIES.

⇓

SIGNAL DECOMPOSITION

other types of decompositions exist depending on the choice of "basis set" or "basis function" (eg. sherical harmonics, ~~...~~ etc...)
~~...~~ ... → wavelet analysis.

Let us now consider a "statistical model"

$$y = \underset{\text{param.}}{\alpha} x + \underset{\text{error}}{n} \qquad \text{model}$$

take the mean $\langle\ \rangle$

$$\langle y \rangle = \alpha \langle x \rangle + \langle n \rangle$$

define $y' = y - \langle y \rangle$

$$y' = \alpha x' + n'$$

now compute the covariance between $\langle x' y' \rangle$

$$\langle x' y' \rangle = \alpha \langle x' x' \rangle + \underbrace{\cancel{\langle n' h' \rangle}}_{= \emptyset} \implies \alpha = \frac{\langle x' y' \rangle}{\langle x' x' \rangle}$$

If you define y = as the model output

x = as the model input

$\underline{\underline{C}}_{oo}$ = covariance of the outputs × outputs

$\underline{\underline{C}}_{oi}$ = covariance of the outputs × inputs

vector of output variable

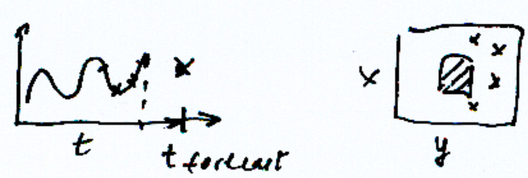
$$\underline{o} = \underbrace{\underline{\underline{C}}_{oi}\,\underline{\underline{C}}_{ii}^{-1}}_{\text{statistical model based on the covariance}}\,\underline{i}$$

vector of input data

↳ FUNDAMENTAL RESULT which is at the root of most statistical
① interpolation
② extrapolation
③ hindcast = time interpolation
④ forecast

The flavor of the various methods in ①, ②, ③, ④ depends on how one specifies the "covariance"

example:



One can decompose the covariance in eigenvalues and eigenmodes → EOFs

↓

another example of signal decomposition, in which the basis set is made up of ~~modes~~ orthogonal modes explaining a certain fraction of the variance.