Multivariate Interpolation of Wind Field Based on GPR

Mia Feng

January 24, 2018

Incompatible interpolation operator gives rise to representative error, which is a big challenge for improving the accuracy of numerical weather prediction. The multi-kernel interpolation method based on Gaussian Process Regression we proposed pave a new way to make use of multi variables to infer the weather process.

Reference

C. Rassemussen, "Gaussian processes for machine learning.'' 2004.

Outline

- GPR
- Multivariate Interpolation
- Issues

GPR and kernel tricks – Gaussian Process and GPR

Gaussian Process

$\mathcal{GP}$ is a collection of random variables, any finite number of which have a joint Gaussian distribution.

$$f(\mathbf{x}) \sim \mathcal{GP}\left(m(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')\right) \tag{1}$$

For any $\mathbf{X}$, the collection of $\mathbf{x}$ in $\mathcal{GP}$, it follows a jointly Gaussian distribution.

For $\mathcal{GP}$, knowing mean function and covariance function means knowing everything. GPR infers the predictive value based on the mean function and covariance function

GPR – Start from Bayesian linear regression

$$y = f(\mathbf{x}) + \varepsilon, \quad f(\mathbf{x}) = \mathbf{x}\mathbf{w}^{\mathrm{T}} \\ \varepsilon \sim \mathcal{N}\left(0, \sigma_n^2\right) \tag{2}$$

Put a prior distribution on $\mathbf{w}$: $\mathbf{w} \sim \mathcal{N}(0, \Sigma_p)$, the posterior of y_* at input vector $\mathbf{x}_*$ is

$$\begin{aligned} p(y_*|\mathbf{x}_*, \mathbf{X}, \mathbf{y}) &= \int p(y_*|\mathbf{x}_*, \mathbf{w})\, p(\mathbf{w}|\mathbf{X}, \mathbf{y})\, d\mathbf{w} \\ &= \mathcal{N}\left(\frac{1}{\sigma_n^2}\mathbf{x}_*{}^{\mathrm{T}} A^{-1}\mathbf{X}\mathbf{y}, \mathbf{x}_*^{\mathrm{T}} A^{-1}\mathbf{x}_*\right) \end{aligned} \tag{3}$$

which implies it is a GRP model, where $A = \sigma_n^{-2}\mathbf{X}\mathbf{X}^{\mathrm{T}} + \Sigma_p^{-1}$, with a control function $p(\mathbf{y}|\mathbf{X}, \mathbf{w})$.

$$\begin{aligned} p(\mathbf{y}|\mathbf{X}, \mathbf{w}) &= \prod_{i=1}^{n} p(y_i|\mathbf{x}_i, \mathbf{w}) \\ &= \mathcal{N}\left(\mathbf{X}^{\mathrm{T}}\mathbf{w}, \sigma_n^2\mathbf{I}\right) \end{aligned} \tag{4}$$

GPR – kernel tricks

★ To deal with the nonlinear problem, import mapping operator $\phi(\cdot)$

Idea

Change the BASIS SPACE

Why is it useful?
Change the measuring distance – the similarity of two input vectors $\mathbf{x}, \mathbf{x}'$, which implies the similarity of targets

kernel function

$$p\left(y_*|\mathbf{x}_*, \mathbf{X}, \mathbf{y}\right) \\ = \mathcal{N}\left(\frac{1}{\sigma_n^2}\phi\left(\mathbf{x}_*\right)^{\mathrm{T}} A^{-1}\phi\left(\mathbf{X}\right)\mathbf{y}, \phi\left(\mathbf{x}_*\right)^{\mathrm{T}} A^{-1}\mathbf{x}_*\right) \tag{5}$$

$$p\left(y_*|\mathbf{x}_*, \mathbf{X}, \mathbf{y}\right) = \mathcal{N}\left(\phi_*^{\mathrm{T}}\Sigma_p\Phi\left(K+\sigma_n^2 I\right)^{-1}\mathbf{y}, \\ \phi_*^{\mathrm{T}}\Sigma_p\phi_* - \phi_*^{\mathrm{T}}\Sigma_p\Phi\left(K+\sigma_n^2 I\right)^{-1}\Phi^{\mathrm{T}}\Sigma_p\phi_*\right) \tag{6}$$

where $K = \phi\left(\mathbf{X}\right)^{\mathrm{T}}\Sigma_p\phi\left(\mathbf{X}\right), \Phi = \phi\left(\mathbf{X}\right)$
Suppose $k\left(\mathbf{x}, \mathbf{x}'\right) = \phi\left(\mathbf{x}\right)\Sigma_p\phi\left(\mathbf{x}'\right)$

$$p\left(y_*|\mathbf{x}_*, \mathbf{X}, \mathbf{y}\right) = \mathcal{N}\left(\mathbf{K}\left(\mathbf{x}_*, \mathbf{X}\right)\left(K+\sigma_n^2 I\right)^{-1}\mathbf{y}, \\ \mathbf{K}\left(\mathbf{x}_*, \mathbf{x}_*\right) - \mathbf{K}\left(\mathbf{x}_*, \mathbf{X}\right)\left(K+\sigma_n^2 I\right)^{-1}\mathbf{K}\left(\mathbf{x}_*, \mathbf{X}\right)\right) \tag{7}$$

kernel function

$k(\cdot,\cdot)$ is called kernel function , it is a covariance function, which defines the similarity

The predictive value of y_* at input vector x_*,

$$y_* = K(\mathbf{x}_*, \mathbf{X}) \left[K(\mathbf{X}, \mathbf{X}) + \sigma_n^2 I\right]^{-1} \mathbf{y} \tag{8}$$

with a control function what solves the unknown parameter θ , which are σ_n^2, Σ_p here.

$$\log p(\mathbf{y}|\mathbf{X}, \theta) = -\frac{1}{2}\mathbf{y}^{\mathrm{T}} K^{-1}\mathbf{y} - \frac{1}{2}\log|K| - \frac{n}{2}\log 2\pi \tag{9}$$

popular kernel function

kenel	formula	advantages
SE	$\exp\left(-\frac{r^2}{2l^2}\right)$	smooth
Matérn	$\left(1+\frac{\sqrt{3}r}{l}\right)\exp\left(-\frac{\sqrt{3}r}{l}\right)$	rough
Gabor	$\exp\left(-\frac{1}{2}\mathbf{t}^{\mathrm{T}}\Lambda^{-2}\mathbf{t}\cos\left(2\pi\mathbf{t}_p^{\mathrm{T}}\mathbf{l}\right)\right)$	edge extraction
Periodic	$u(x)=\left(\cos\left(x\right),\sin\left(x\right)\right)$	periodical space

Generating kernel from old kernel

$$\begin{aligned} k\left(\mathbf{x},\mathbf{x}'\right) &= k_1\left(\mathbf{x},\mathbf{x}'\right)+k_2\left(\mathbf{x},\mathbf{x}'\right) \\ k\left(\mathbf{x},\mathbf{x}'\right) &= k_1\left(\mathbf{x},\mathbf{x}'\right)*k_2\left(\mathbf{x},\mathbf{x}'\right) \\ k\left(\mathbf{x},\mathbf{x}'\right) &= \alpha k_1\left(\mathbf{x},\mathbf{x}'\right) \end{aligned} \tag{10}$$

Multi-kernel

$$k_v(\cdot,\cdot) = k_m(\cdot,\cdot) + k_p(\cdot,\cdot) + k_g(\cdot,\cdot) + k_\varepsilon(\cdot,\cdot) \qquad (11)$$

where $k_m, k_p, k_g, k_\varepsilon$ denotes Matérn of $t = \frac{3}{2}$, periodical Matérn, gabor and noise kernel, respectively.

The information that they try to capture:

kenel	target
k_m	depicting the roughness of weather process
k_p	periodical information of vortex in typhoon region or wind belt
k_g	extracting the vortex texture, the edge of wind belt
k_ε	noise

Multivariate Multi-kernel

Limitation

- 1D-GPR equals the piecewise spline
- The partial information of spacial distribution

Tips: importing more information

- space information(longitude and latitude) – key feature
- the principle component of wind direction, pressure and temperature – secondary feature

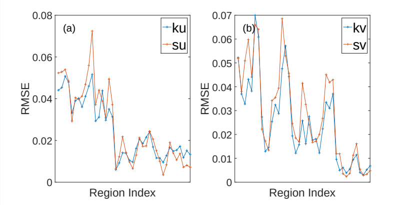
Multivariate Multi-kernel

$$\begin{aligned} k_{vms}(\cdot,\cdot) &= k_v(\cdot,\cdot) + k_v^{'}(\cdot,\cdot) \\ k_{vmp}(\cdot,\cdot) &= k_v(\cdot,\cdot) * k_v^{'}(\cdot,\cdot) \end{aligned} \tag{12}$$

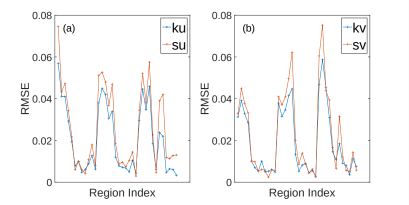
where k_{vms} denotes the correction kernel for the weather in normal condition, k_{vmp} denotes the correction kernel for the weather in extreme condition.
k_v is trained by key feature, k_v' is trained by secondary feature.

- Should I Explain it?

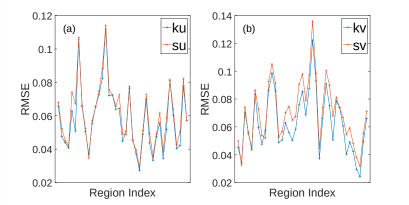
Wind field interpolation – Normal weather condition



space series

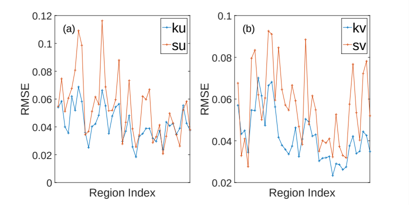


space series



time series

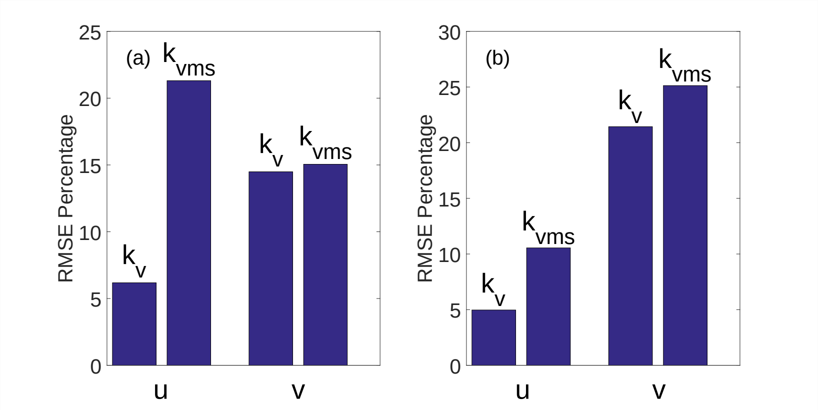
k_v



time series

k_{vms}

Wind field interpolation – Normal weather condition

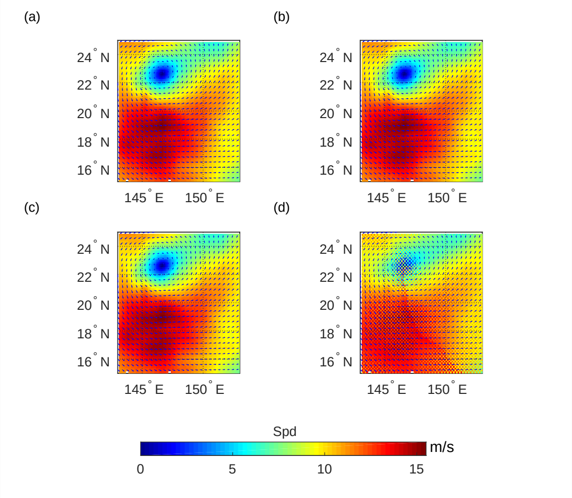


space series

time series

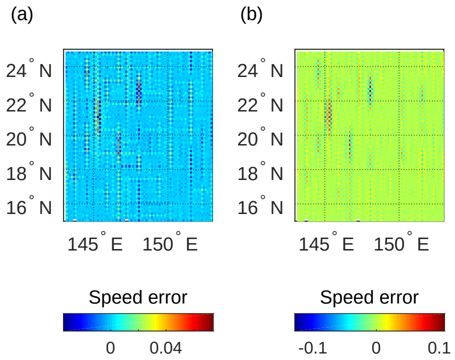
- Experiment?

Wind field interpolation – Extreme weather condition

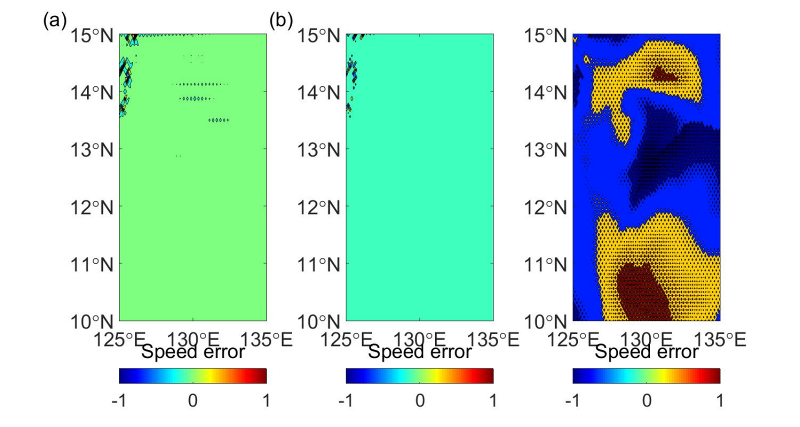


- (a) reference field
- (b) k_{vmp}
- (c) spline
- (d) BP

Wind field interpolation – Extreme weather condition



Wind field interpolation – Extreme weather condition



GPR and 3DVAR

loss function

$$-\frac{1}{2}\mathbf{y}^{\mathrm{T}}K^{-1}\mathbf{y}-\frac{1}{2}\log|K|-\frac{n}{2}\log 2\pi$$

$$-\frac{1}{2}\left(x_a-x_b\right)^{\mathrm{T}}\mathbf{B}^{-1}\left(x_a-x_b\right)-\frac{1}{2}\left(y-H\left(x_a\right)\right)^{\mathrm{T}}\mathbf{R}^{-1}\left(y-H\left(x_a\right)\right)$$

Dose the matrix **B** describes the similarity?

GPR and 3DVAR

assumption

$$y \sim \mathcal{GP}$$

$$prior : x \sim \mathcal{N}\left(x_b, B\right)$$

x_a follows a single Gaussian distribution, Whether or not? What if we suppose x_a comes from a Gaussian mixture distribution?

kernel functions

step

1. Mining: the backgroud of your dataset
2. Draw: visualizing your data
3. Choice: appropriate kernel – stationary? periodical? linear? smooth?
4. Try

Autoregression model – may be nonsense

AR,MA,ARMA etc.
Taking the series itself as the only explaining variable, the idea of them is extremely simple. They are popular in stationary series analysis.

Some about M.L.

Any model can be powerful, even the simplest one, as long as you make a good decision, that is, choose a model that fits your data.

Neither M.L. nor D.L. are the magician, you are, you are the one who teach them how to do and what to do.