A generalized quadratic estimate for random field nonstationarity

Ethan Anderes† (UC Davis Statistics Department)

Joint with Joe Guinness (NCSU Statistics Department)

† Research supported by: NSF CAREER DMS-1252795

Introduction

- For most of the work I do in Cosmology I am trying to bring new statistical tools/methodology to the science.
- Today I am going to talk about the reverse:

Generalize a tool developed in Cosmology and make it available to the wider statistical community

Spatial statistics

The difficulty with non-stationarity

- I would propose that non-stationarity random fields is the bane of applied spatial statisticians.
- Non-stationarity is almost always present in the data.
- However, as statisticians we still don't have an agreed upon generic way to handle it.
- When compared to the theory of stationary random fields we know relatively little.

Comparison of stationarity vrs non-stationarity in 1-d

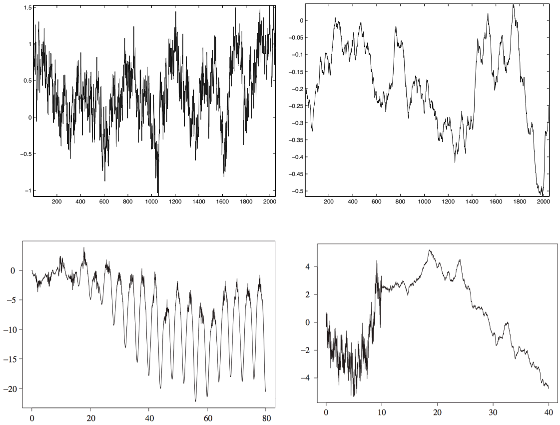
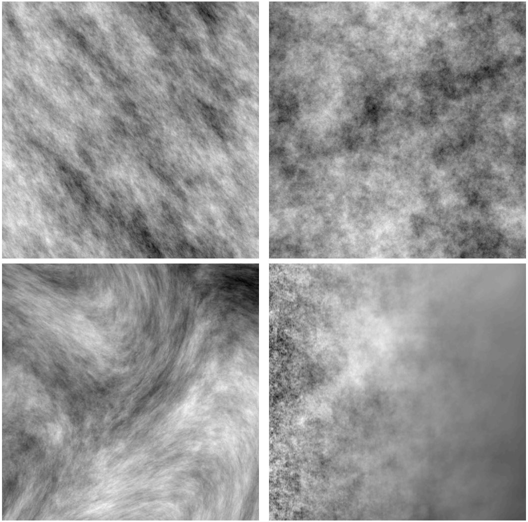


Figure: *Top:* Stationary processes. *Bottom:* Non-stationary processes (images from Sly, 2007).

Comparison of stationarity vrs non-stationarity in 2-d

Stationary
Translation invariant.

Isotropic
Translation and rotation invariant



Non-stationary

Non-stationary

There are many proposals in the statistics literature for dealing with non-stationarity.

Some examples

- *Spatially varying smoothing of white noise*. See Fuentes, Higdon, Benassi, Cohen, Istas, Ayache.
- *Basis expansions* (Wavelets, EOF, etc.). See Nychka, Cressie, Wikle.
- *Deformed stationary random fields*. See Sampson, Guttorp, Stein, EA.
- *Spatial varying spectral densities*. See Priestley, Dahlhaus.

Most require complicated, computationally difficult, estimates which do not lend themselves to uncertainty quantification and analysis.

Evolutionary spectra

The last example will come up later.

Spatially varying local spectral density

$$Z(x) = \int_{\mathbb{R}^d} e^{ix\cdot k} A(k,x)\sqrt{C_k}\frac{dW_k}{(2\pi)^{d/2}}$$

where

- C_k is a baseline spectral density.
- dW_k is a complex Gaussian white noise measure which satisfies $E|dW_k|^2 = dk$.
- $A(k,x)$ is a spatially varying modulation of $\sqrt{C_k}$.
- Think of $Z(x)$ as a locally stationary process with local spectral density given by $|A(k,x)|^2 C_k$.
- The goal is to estimate $A(k,x)$ given observations of $Z(x)$.

Evolutionary spectra

Spatially varying local spectral density

$$Z(x) = \int_{\mathbb{R}^d} e^{ix\cdot k} A(k,x)\sqrt{C_k}\frac{dW_k}{(2\pi)^{d/2}}$$

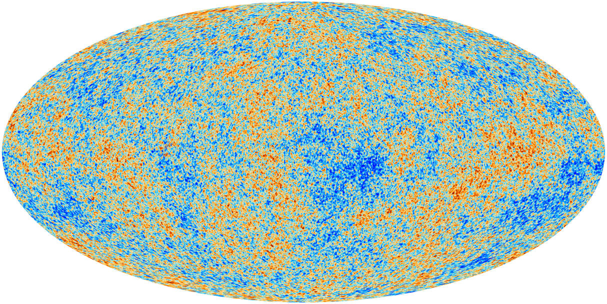
- Methods developed for this process usually involve a local spectral density estimate or local likelihood for estimating properties of $|A(k,x)|^2 C_k$.
- It is not clear, however, how one estimates the phase of $A(k,x)$.
  - Is it identifiable? Estimable?
  - What does one gain/lose by letting the phase modulation spatially vary?
  - What class of models satisfy $|A(k,x)|^2 = 1$?

The quadratic estimate of CMB lensing

A beautiful application for random field non-stationarity

The Cosmic Microwave Background (CMB)

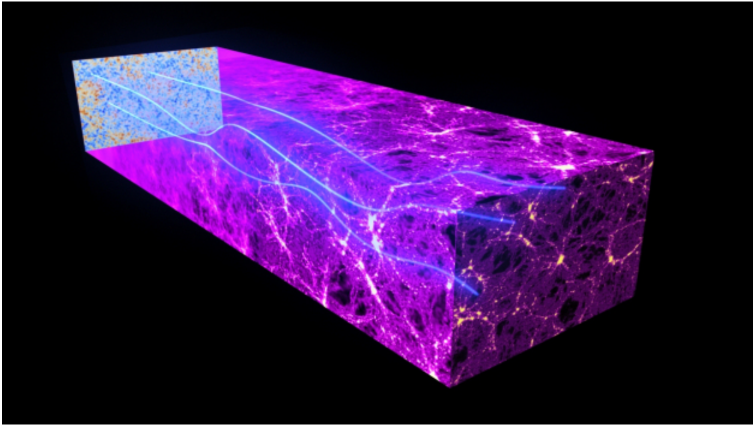
- The CMB is a old distant cosmological image. Here is a recent picture taken from the Planck satellite:



- Shows radiation fluctuations after the big bang
- Observations are gravitationally lensed from intervening matter.

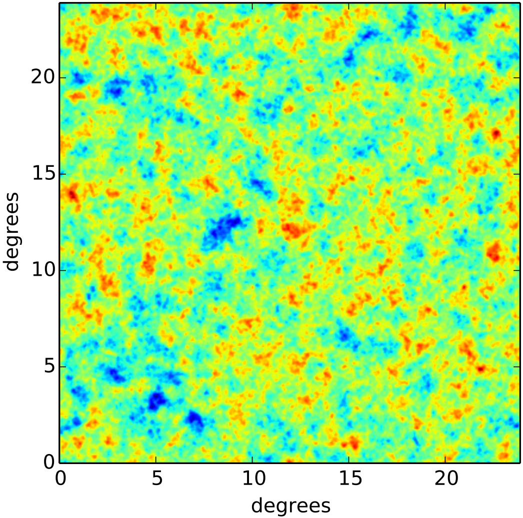
Gravitational lensing:

The bending or distortion of photon trajectories by the gravitational influence of intervening matter.

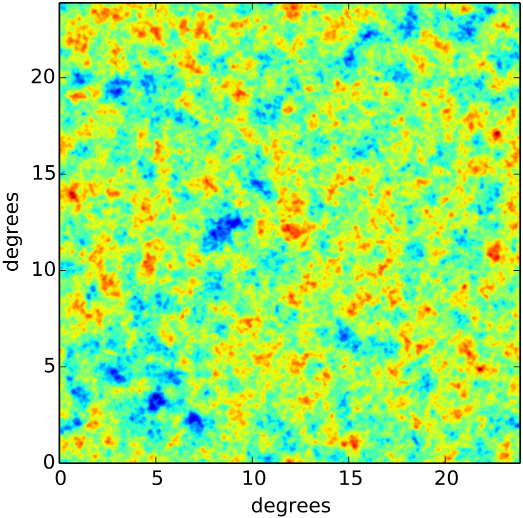


Picture Credit: ESA

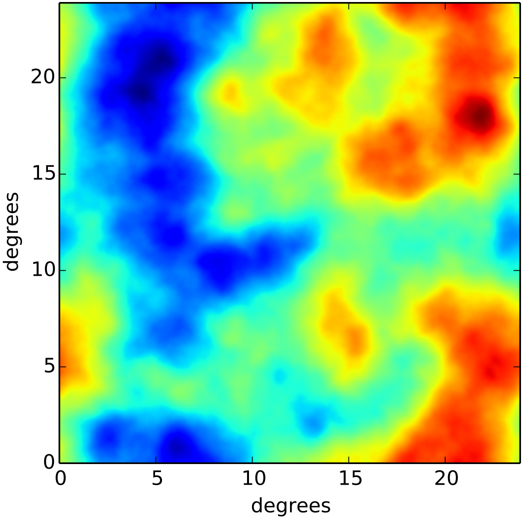
Simulated CMB: No lensing



Simulated CMB: lensing

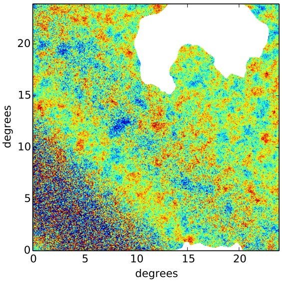


The lensing map is characterized by a lensing potential ϕ



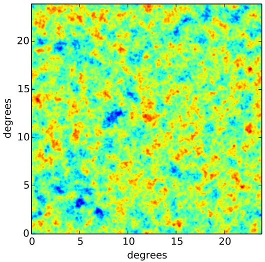
The primary goal of CMB lensing studies

The Data:

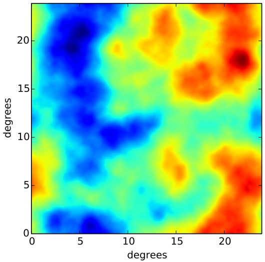


Lensed CMB
with nonstationary noise
with nonstationary beam
with masking

Estimate the unlensed CMB



...and the lensing potential



The quadratic estimator of lensing

- Developed by Hu and Okamoto (2001, 2002) for estimating the lensing potential $\phi(x)$ from the lensed CMB observations.
- I started studying the quadratic estimate thinking it was going to easy to beat with alternative methods.
- It turns out to have some amazing properties...
- ... some of which are not entirely understood.

Deriving the quadratic estimator

- The unlensed CMB is isotropic

  The Fourier transform of the unlensed CMB is uncorrelated

- The lensed CMB is non-isotropic

  The non-stationarity induces auto-correlated in the Fourier transform

- The quadratic estimate works by detecting this auto-correlation in the Fourier transform of the data.

Deriving the quadratic estimator

- Let $T(x)$ denote the unlensed CMB, $\phi(x)$ denote the lensing potential.
- The lensed CMB is modeled $\tilde{T}(x) = T(x + \nabla\phi(x))$.
- To derive the quadratic estimator $\hat{\phi}$ Taylor expand around x

$$\tilde{T}(x) \approx \underbrace{T(x)}_{\text{isotropic}} + \underbrace{\nabla T(x) \cdot \nabla\phi(x)}_{\text{not isotropic}}$$

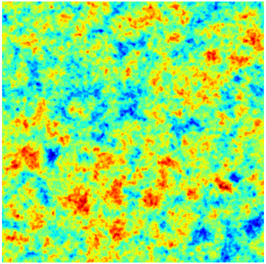
- Taking the Fourier transform

$$\tilde{T}_\ell \approx \underbrace{T_\ell}_{\substack{\text{spatially}\\ \text{uncorrelated}}} + \underbrace{i\ell\, T_\ell \star i\ell\, \phi_\ell}_{\substack{\text{spatially}\\ \text{correlated}}}$$

- The quadratic estimate uses observed auto-correlation in $\tilde{T}_\ell$, modeled by the above linear term, to estimate ϕ.

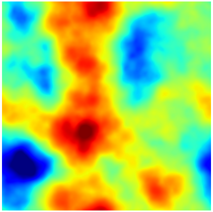
The quadratic estimate works amazingly well

Simulated lensed CMB

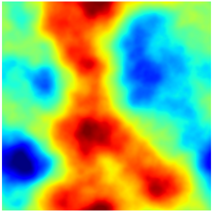


$17^o \times 17^o$ patch of the sky
1024×1024 pixels

Estimated potential



Simulation truth potential



The quadratic estimator is fast

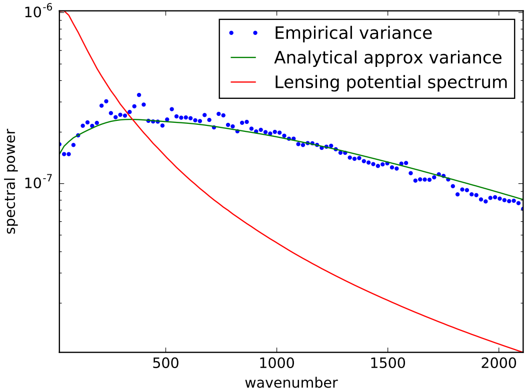
1. Here is how easy it is to implement (in my new favorite programing language: *Julia*)

```
function quad_est(Tk, CTTobs, CTT, Aℓ, g)
    Hf    =   1 ./ CTTobs
    Lf    = CTT ./ CTTobs
    Hx    = g.FFT \ (Hf .* Tk)
    ∂1Lx  = g.FFT \ (im .* g.k[1] .* Lf .* Tk)
    ∂2Lx  = g.FFT \ (im .* g.k[2] .* Lf .* Tk)
    estk  = - im .* g.k[1] .* (g.FFT * (2∂1Lx .* Hx))
    estk += - im .* g.k[2] .* (g.FFT * (2∂2Lx .* Hx))
    return Aℓ .* estk ./ 2 ./ π
end
```

2. Runs in about 0.4 seconds on a 1024×1024 image.
3. In contrast, a local likelihood approach takes hours to compute (see E.A, Knox, van Engelen) .

Accurate approximation of sampling variability

- There is also an easily computable analytic approximation to the sampling variability of the estimate.



- In contrast, a local likelihood approach needs a huge amount of simulation to quantify estimation uncertainty.

An example of some puzzling behavior

- Recall that the quadratic estimate is derived with a first order Taylor expansion of the lensing effect

$$T(x + \nabla\phi(x)) \approx T(x) + \nabla T(x) \cdot \nabla\phi(x)$$

called linear order lensing.

- Abstractly write the full expansion as

$$\widetilde{T} = \delta^0 \widetilde{T} + \delta^1 \widetilde{T} + \delta^2 \widetilde{T} + \cdots$$

where $\delta^n \widetilde{T} = \mathcal{O}(\phi^n)$. So that $\delta^1 \widetilde{T} \equiv \nabla T(x) \cdot \nabla\phi(x)$.

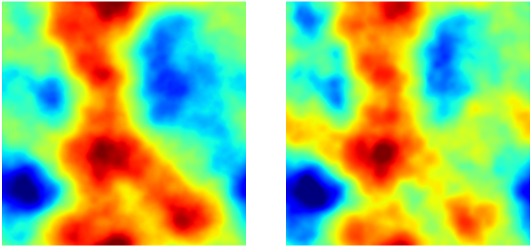
An example of some puzzling behavior

- Even though the quadratic estimate is derived from the linear model...

$$\text{data} = \delta^0 \widetilde{T} + \delta^1 \widetilde{T} + N$$

- ...it works spectacularly on "all order lensing" data

$$\text{data} = \delta^0 \widetilde{T} + \delta^1 \widetilde{T} + \delta^2 \widetilde{T} + \cdots + N.$$



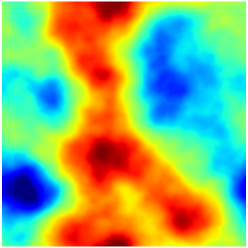
- This is what you might expect when the higher order terms $\delta^n \widetilde{T}$, for $n \geqslant 2$ are small.

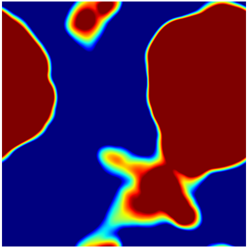
An example of some puzzling behavior

- However, if one simulates from the model used to derive the estimate ....

$$\text{data} = \delta^0\,\widetilde{T} + \delta^1\,\widetilde{T} + \cancel{\delta^2\,\widetilde{T} + \cdots} + N.$$

...the estimator bonks with incredibly huge bias.





- Weird.

- The quadratic estimator is unlike any estimator of non-stationarity I've studied.
- It would be an extremely powerful estimator if one could generalize it to other problems.
- To do this, however, we need to understand what makes the quadratic work and why it has such small bias on all order lensing.

A connection with phase modulation

- Note the Fourier decomposition of the unlensed CMB temperature (in the flat sky approximation)

$$T(x) = \int_{\mathbb{R}^d} e^{ix\cdot k} \sqrt{C_k^{TT}} \frac{dW_k}{(2\pi)^{d/2}}$$

so the lensed CMB can be written

$$\begin{aligned}\widetilde{T}(x) &= \int_{\mathbb{R}^d} e^{i(x+\nabla\phi(x))\cdot k} \sqrt{C_k^{TT}} \frac{dW_k}{(2\pi)^{d/2}} \\ &= \int_{\mathbb{R}^d} e^{ix\cdot k} \underbrace{e^{i\nabla\phi(x)\cdot k}}_{A(k,x)} \sqrt{C_k^{TT}} \frac{dW_k}{(2\pi)^{d/2}}\end{aligned}$$

- Lensing non-stationarity is a spatially varying spectral phase modulation.
- Could there be a larger theory here?

A generalized quadratic estimator

Taylor expansion analysis

- Lets try and figure out why the quadratic estimate has such low bias.
- Start with the all order lensing expansion.

$$\widetilde{T} = \delta^0 \widetilde{T} + \delta^1 \widetilde{T} + \delta^2 \widetilde{T} + \cdots .$$

- Since quadratic estimate $\hat{\phi}$ is linear in the quadratic form $\widetilde{T}(x)\widetilde{T}(y)$, written abstractly as $(\widetilde{T})^2$, so it is natural to expand $(\widetilde{T})^2$ and regroup by order of ϕ

$$(\widetilde{T})^2 = \underbrace{(\delta^0 \widetilde{T})^2}_{I} + \underbrace{2(\delta^0 \widetilde{T})(\delta^1 \widetilde{T})}_{II} + \underbrace{(\delta^1 \widetilde{T})^2 + 2(\delta^0 \widetilde{T})(\delta^2 \widetilde{T})}_{III} + \cdots$$

- Now we can analyze the quadratic estimate applied to each term individually.

$$(\widetilde{T})^2 = \underbrace{(\delta^0 \widetilde{T})^2}_{I} + \underbrace{2(\delta^0 \widetilde{T})(\delta^1 \widetilde{T})}_{II} + \underbrace{(\delta^1 \widetilde{T})^2 + 2(\delta^0 \widetilde{T})(\delta^2 \widetilde{T})}_{III} + \cdots$$

- Term I controls the estimation variance in $\hat{\phi}$. It is often called the 'shape noise' or 'Gaussian noise'.
- The estimator $\hat{\phi}$ is designed around term II. Applying $\hat{\phi}$ to II will yield a low-noise zero-bias estimate.
- The remaining terms $III, IV, \ldots$ result in estimation bias.

$$(\widetilde{T})^2 = \underbrace{(\delta^0\,\widetilde{T})^2}_{I} + \underbrace{2(\delta^0\,\widetilde{T})(\delta^1\,\widetilde{T})}_{II} + \underbrace{(\delta^1\,\widetilde{T})^2 + 2(\delta^0\,\widetilde{T})(\delta^2\,\widetilde{T})}_{III} + \cdots$$

- The success of the quadratic estimate is driven by the fact that $III + IV + \cdots$ is small.
- What's interesting is that it is **not** because the higher order terms $\delta^2\,\widetilde{T}$, $\delta^3\,\widetilde{T} + \cdots$ are small.
- Rather, the terms within $III, IV, \ldots$ are highly negatively correlated and combine to make small bias

$$(\widetilde{T})^2 = \underbrace{(\delta^0\widetilde{T})^2}_{I} + \underbrace{2(\delta^0\widetilde{T})(\delta^1\widetilde{T})}_{II} + \underbrace{(\delta^1\widetilde{T})^2 + \cancel{2(\delta^0\widetilde{T})(\delta^2\widetilde{T})}}_{III} + \cdots$$

- This is why linear lensing (or truncating to any order) breaks the quadratic estimate: it prohibits the cancellation within each term $III, IV, \ldots$
- The question remains:

  What is it about lensing that enables this cancellation and does it exist anywhere else?

- To understand which non-stationary models exhibit this cancellation one needs to realize that map level Taylor expansions are misleading when there exists distributional invariance.
- For example let $v \in \mathbb{R}^d$ be large enough as to make a zero$^{\text{th}}$ order Taylor expansion inaccurate

$$T(x+v) \not\approx T(x)$$

- However, if T is stationary this Taylor expansion is exact in a distributional sense

$$T(x+v) \overset{\mathcal{D}}{=} T(x)$$

- The right way to analyze the Taylor expansion is in the distributional sense.
- Taking expected value of $(\widetilde{T})^2$ expansion gives

$$
\begin{aligned}
&E(\widetilde{T}(x)\widetilde{T}(y)) \\
&\quad = \underbrace{C^{(0)}(x-y)}_{E(I)} + \underbrace{C^{(1)}(x-y)\cdot(\boldsymbol{\theta}(x)-\boldsymbol{\theta}(y))}_{E(II)} + \underbrace{\mathcal{O}((\boldsymbol{\theta}(x)-\boldsymbol{\theta}(y))^2)}_{E(III)} + \cdots
\end{aligned}
$$

where $\boldsymbol{\theta}(x) := \nabla\phi(x)$.

Each term in the above expansion is function of $x-y$ and $\boldsymbol{\theta}(x)-\boldsymbol{\theta}(y)$ only. In E.A.& Guinness [arxiv:1603.03496] we argue this is the essential feature which makes the quadratic estimate have such low bias and which can be generalized.

Generalizing to other non-stationary random fields

Consider a general non-stationary random field $Z(x)$ on $\mathbb{R}^d$ where the non-stationarity is characterized by a vector field $\boldsymbol{\theta}(x)\colon \mathbb{R}^d \to \mathbb{R}^d$.

Definition (E.A.& Guinness)

$Z(x)$ is said to have **local invariant non-stationarity** if it's covariance function has the form

$$\mathrm{cov}(Z(x), Z(y)) = K\big(x-y, \boldsymbol{\theta}(x)-\boldsymbol{\theta}(y)\big)$$

with $K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$.

- The name *local invariance* expresses the fact that the local behavior of $Z(x)$ is invariant to the magnitude of $\boldsymbol{\theta}(x)$.

Local invariant non-stationarity

To derive a quadratic estimate for these random fields, expand the second argument of K

$$K\big(x - y, \boldsymbol{\theta}(x) - \boldsymbol{\theta}(y)\big) = C^{(0)}(x-y) + C^{(1)}(x-y) \cdot (\boldsymbol{\theta}(x) - \boldsymbol{\theta}(y)) + \cdots$$

Generalized quadratic estimate (E.A.& Guinness)

A generalized quadratic estimate can always be constructed from the above expansion when $\boldsymbol{\theta}(x)$ is characterized by a potential $\boldsymbol{\theta}_k = \boldsymbol{\xi}_k \phi_k$ and has the form

$$\hat{\phi}_{\boldsymbol{\ell}} = A_{\boldsymbol{\ell}} \sum_{p=1}^{d} \xi^*_{p,\boldsymbol{\ell}} \int e^{-i\boldsymbol{x}\cdot\boldsymbol{\ell}} \mathscr{A}(\boldsymbol{x}) \mathscr{B}_p(\boldsymbol{x}) \frac{d\boldsymbol{x}}{(2\pi)^{d/2}}$$

where $\mathscr{A}_{\boldsymbol{\ell}} := Z^{obs}_{\boldsymbol{\ell}} / C^{ZZobs}_{\boldsymbol{\ell}}$ and $\mathscr{B}_{p,\boldsymbol{\ell}} := i2\,\mathrm{imag}(C^{(1)}_{p,\boldsymbol{\ell}}) Z^{obs}_{\boldsymbol{\ell}} / C^{ZZobs}_{\boldsymbol{\ell}}$

Generalized quadratic estimate

$$\hat{\phi}_{\boldsymbol{\ell}} = A_{\boldsymbol{\ell}} \sum_{p=1}^{d} \xi^{*}_{p,\boldsymbol{\ell}} \int e^{-i\boldsymbol{x}\cdot\boldsymbol{\ell}} \mathscr{A}(\boldsymbol{x}) \mathscr{B}_p(\boldsymbol{x}) \frac{d\boldsymbol{x}}{(2\pi)^{d/2}}$$

- Inherits the same attractive features of the quadratic estimate of Hu and Okamoto
  - Estimates non-stationarity via the empirical correlation of Fourier frequencies.
  - Fast and easy to implement.
  - Fast and accurate analytic approximations to sampling variability and bias.
  - Exhibits surprisingly low bias
- Low bias is not guaranteed, but seems to persist when $\boldsymbol{\theta}(x)$ has small to moderate fluctuations and/or smoothness.

Example: spatially varying spectral phase modulation

$$Z(x) = \int_{\mathbb{R}^d} e^{ix\cdot k} e^{i\boldsymbol{\theta}(x)\cdot\boldsymbol{\eta}_k} \sqrt{C_k} \frac{dW_k}{(2\pi)^{d/2}}$$

- $\boldsymbol{\eta}_k : \mathbb{R}^d \to \mathbb{R}^d$ satisfies $\boldsymbol{\eta}_{-k} = -\boldsymbol{\eta}_k$.
- Evolutionary spectral model with $A(k,x) = e^{i\boldsymbol{\theta}(x)\cdot\eta_k}$ so that $|A(k,x)|^2 = 1$.
- Behaves like a generalized warping process...local magnification/demagnification can alter the differentiability of the process.
- $Z(x)$ has local invariant non-stationarity $\Longrightarrow$ there exists a generalized quadratic estimate for $\boldsymbol{\theta}(x)$ assuming $\boldsymbol{\eta}_k$ and C_k are given.

In 1-d

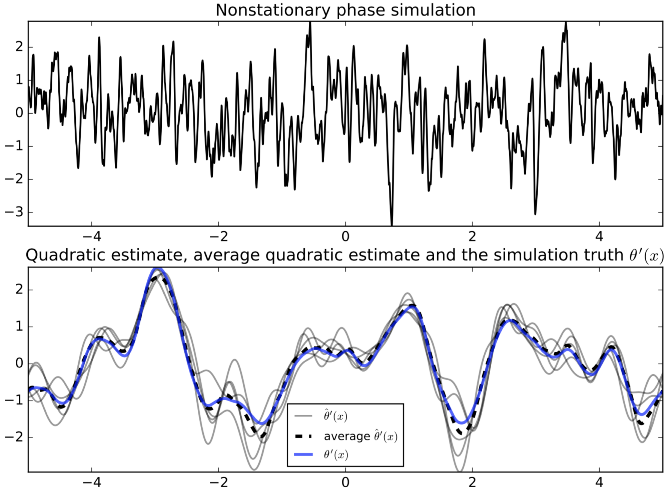


Figure: η_k is constructed so that $Z(x)$ approximates a spatial variation in the Matérn smoothness parameter ν.

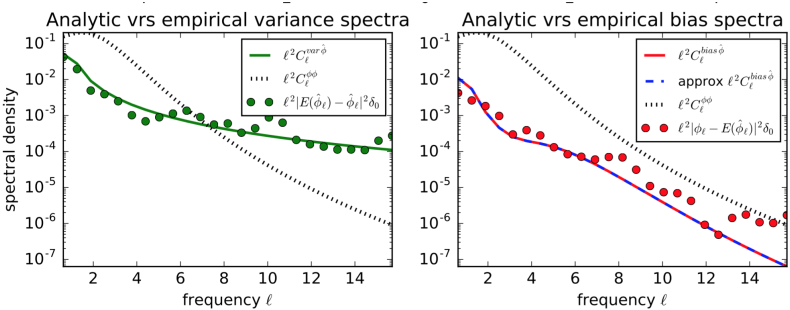


Figure: The derivations in arxiv:1603.03496 yield accurate analytic approximations to variance and bias segmented by spatial frequency.

In 2-d

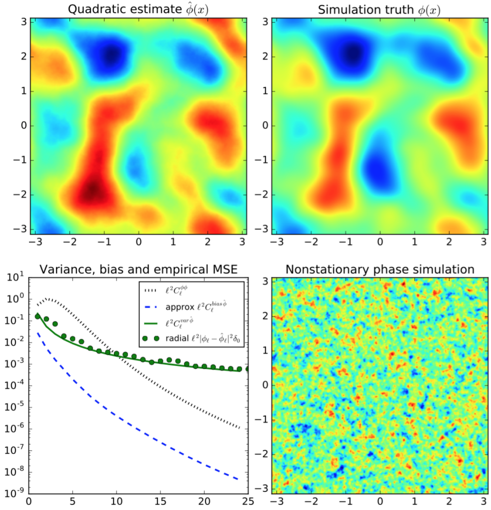
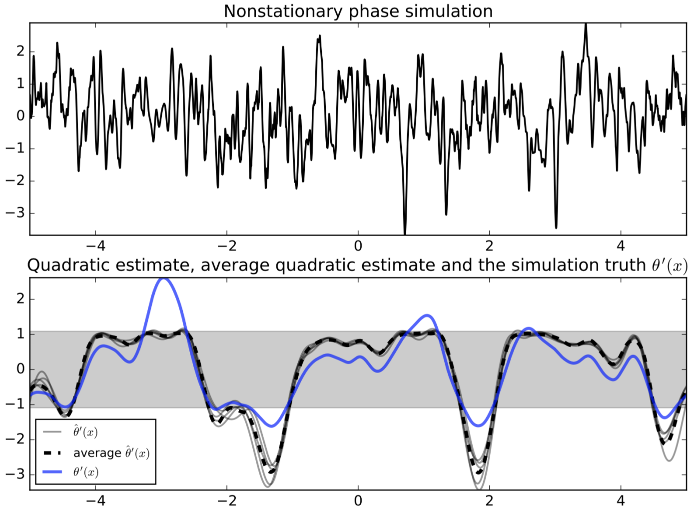
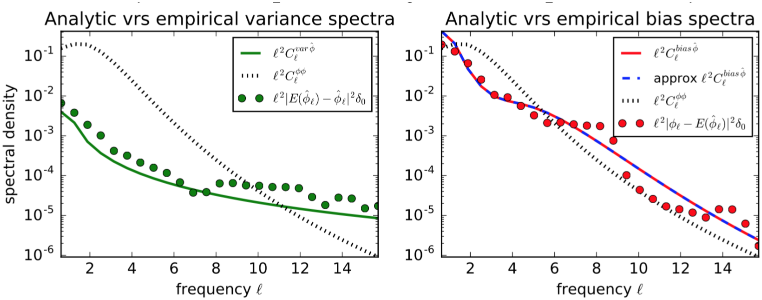


Figure: $\phi(x)$ is a curl potential in this example. $Z(x)$ models a locally varying change in Matérn smoothness and local scale (inversely related).

These generalized quadratic estimates can still have bias when $\nabla\boldsymbol{\theta}(x)$ is large ....



... but the bias is accurately quantified* under Gaussian random field models for ϕ



*The local spectral densities of these models have an interesting connection with L_2 Wasserstein geodesics that allow one to predict (the shaded region on the previous slide) the size of $\nabla\boldsymbol{\theta}(x)$ which results in large bias.

The bias reduction due to local invariance

- Consider the following two stochastic processes on $t \in [\pi, -\pi)$

$$Z(t) := \int e^{itk} e^{i\phi(t)k} \sqrt{C_k} \frac{dB_k}{\sqrt{2\pi}}, \qquad \widetilde{Z}(t) := \int e^{itk} e^{\phi(t)|k|} \sqrt{C_k} \frac{dB_k}{\sqrt{2\pi}}$$

- $Z(t)$ has local invariant non-stationarity. $\widetilde{Z}(t)$ does not.
- Yet they have very similar expansions

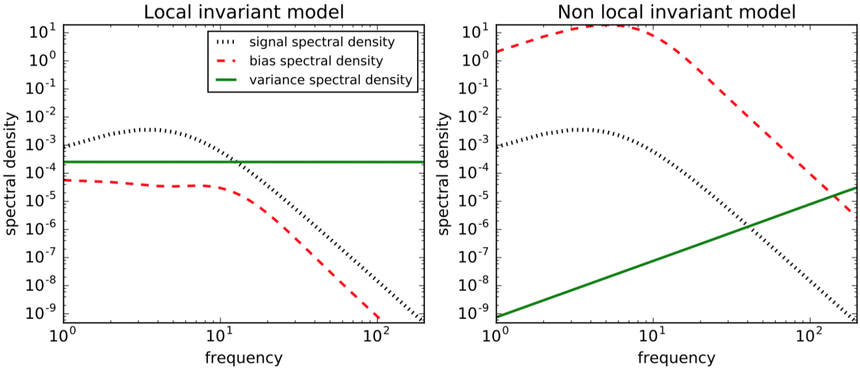
$$\begin{aligned}
\mathrm{cov}(Z(t), Z(s)) &= C^{(0)}(t-s) + \big(\phi(t) - \phi(s)\big)\; C^{(1)}(t-s) \\
&\qquad + \big(\phi(t) - \phi(s)\big)^2 C^{(2)}(t-s) + \mathcal{O}(\phi^3) \\
\mathrm{cov}(\widetilde{Z}(t), \widetilde{Z}(s)) &= C^{(0)}(t-s) + \big(\phi(t) + \phi(s)\big)\; \widetilde{C}^{(1)}(t-s) \\
&\qquad + \big(\phi(t) + \phi(s)\big)^2 \widetilde{C}^{(2)}(t-s) + \mathcal{O}(\phi^3)
\end{aligned}$$

where

$$\widetilde{C}_k^{(1)} := |C_k^{(1)}|, \qquad \widetilde{C}_k^{(2)} := -C_k^{(2)}$$

The bias reduction due to local invariance

- Using the previous covariance expansions one can derive separate quadratic estimates of $\phi(t)$ under $Z(t)$ and $\widetilde{Z}(t)$.
- Even though the two expansions are only different by sign changes, the behavior of the estimates are completely different.



Concluding Remarks

Overview

- Defined an extended class of non-stationary random fields, and a corresponding generalized quadratic estimate, which share the same attractive statistical properties of the original quadratic estimate from Hu and Okamoto.
- We identified a particular form of non-stationarity, *local invariance*, which encourages a delicate cancellation of estimation bias.
- This generalized quadratic estimate is particularly adept at detecting small departures from stationarity and allows fast, accurate quantification of mean square sampling properties.

Future work

- Extensions to general $\boldsymbol{\theta}(x)\colon \mathbb{R}^d \to \mathbb{R}^m$ in the random phase model.
- Non-separable exponents $e^{i\boldsymbol{\theta}(x)\cdot\boldsymbol{\eta}_k} \to e^{i\theta(x,k)}$
- In the evolutionary spectra case, one can always write

$$A(k,x) = e^{\gamma(x,k)} e^{i\theta(x,k)}$$

  where $\gamma(x,k), \theta(x,k) \in \mathbb{R}$.
- It may be possible to build a general theory of inference for $\gamma(x,k)$ and $\theta(x,k)$:
  - The variance of local frequencies $\longrightarrow$ estimating $\gamma(x,k)$
  - Correlation of frequencies $\longrightarrow$ estimating $\theta(x,k)$.
  - Scaling up this estimate for detecting spatially varying cosmological parameters.

Thanks!