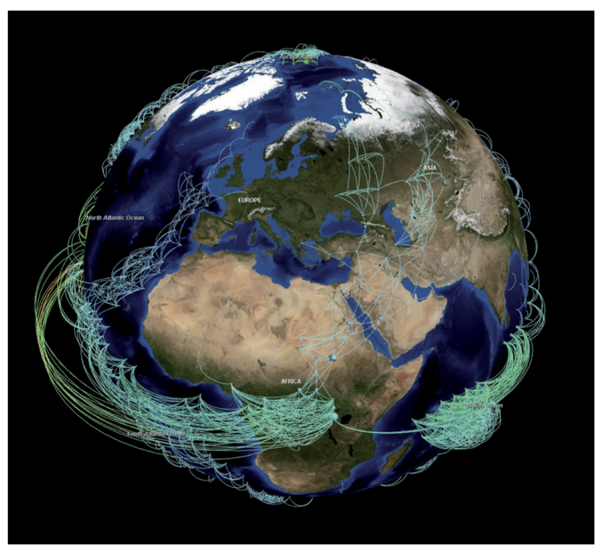
Pitfalls of Climate Network Construction - A Statistical Perspective

Moritz Haas[1], Bedartha Goswami[1], Ulrike von Luxburg[1]

[1]University of Tübingen, Germany.

1) Motivation

Construct networks to **detect complex structures** in the climate system such as teleconnections, clusters and regime transitions, based on time series such as temperature, precipitation or pressure.



Donges et al., Chaos, 2015.

Typical Climate Network Construction:

a. Given climatic variables on fixed locations $V = \{v_i \mid i \in [p]\}$ in some metric space such as the sphere.

b. Choose similarity measure between pairs of locations based on finite time-series $\{X_{it}\}_{i\in[p],t\in[n]}$,

c. Construct density-threshold network from pairwise similarity estimates,

d. Evaluate network characteristics.

Problem: Available data limited and noisy.

↳ Calculated similarity values are imprecise estimates.

↳ False and missing edges in the empirical networks.

Which of the findings in climate networks can be attributed to underlying structure, and which of them are random artifacts due to finite-sample noise?

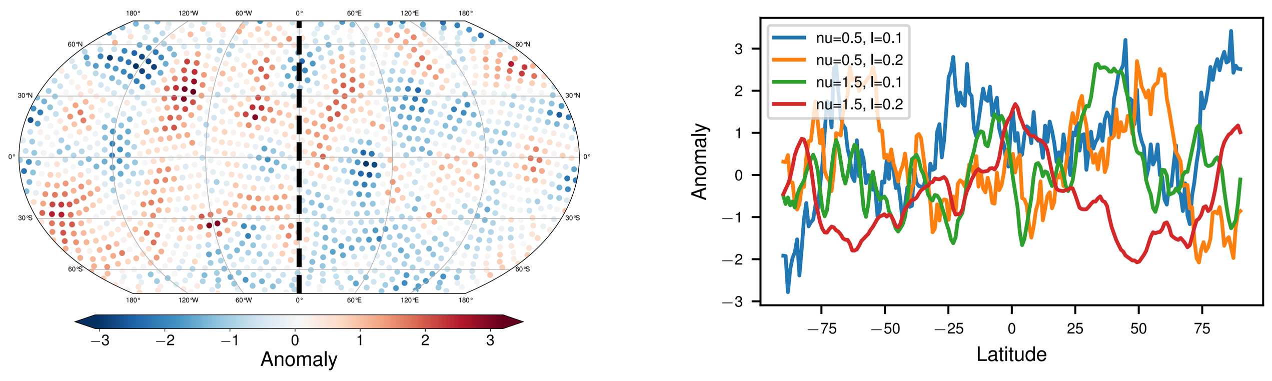
2) We Suggest a Diagnostic Tool: Isotropic Gaussian Random Fields on S^2

On a finite grid $\{v_i\}_{i=1,\dots,p} \subset S^2$, a 0-mean Matérn IGRF G fulfills

$$(G(v_1), \dots, G(v_p)) \sim N(0, \Sigma),$$

with covariance $\Sigma_{ij} = k_{\nu,\ell}(|v_i - v_j|)$.

Matérn covariance function $k_{\nu,\ell}$ flexible with smoothness and length scale parameters ν and ℓ.



Introduce time dependence via VAR(1)-process:

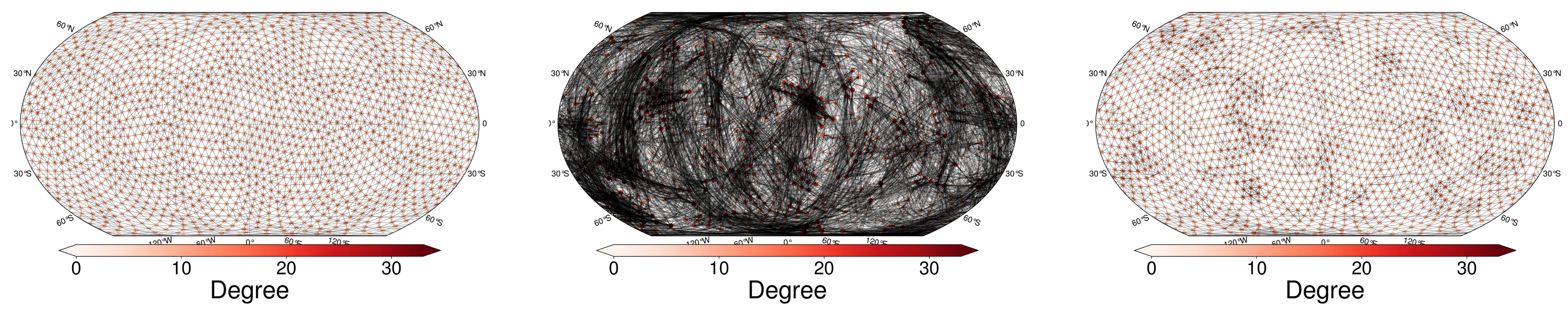
$$X_t = AX_{t-1} + \varepsilon_t,$$

where A diagonal autocorrelation matrix and innovations $\varepsilon_t \sim N(0, \Sigma_\varepsilon)$ i.i.d. to generate $X_t \sim N(0, \Sigma)$ for all t.

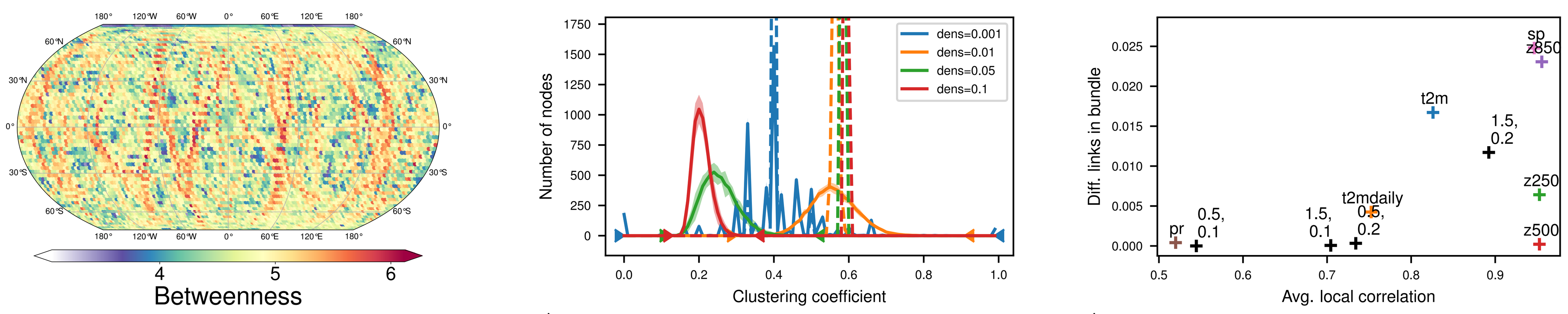
3) We Discover Spurious Behaviour in Empirical Networks

Localized correlation structure leads to errors propagating locally, inducing **spurious link-bundles** and **spuriously dense/sparse regions**.

Ground truth network | Empirical Pearson Correlation Network | Empirical Spearman Correlation Network



Random **global** betweenness 'backbones' | Distorted distributions of network measures | Bundling behaviour increases with local correlations



↳ Significance questionable.

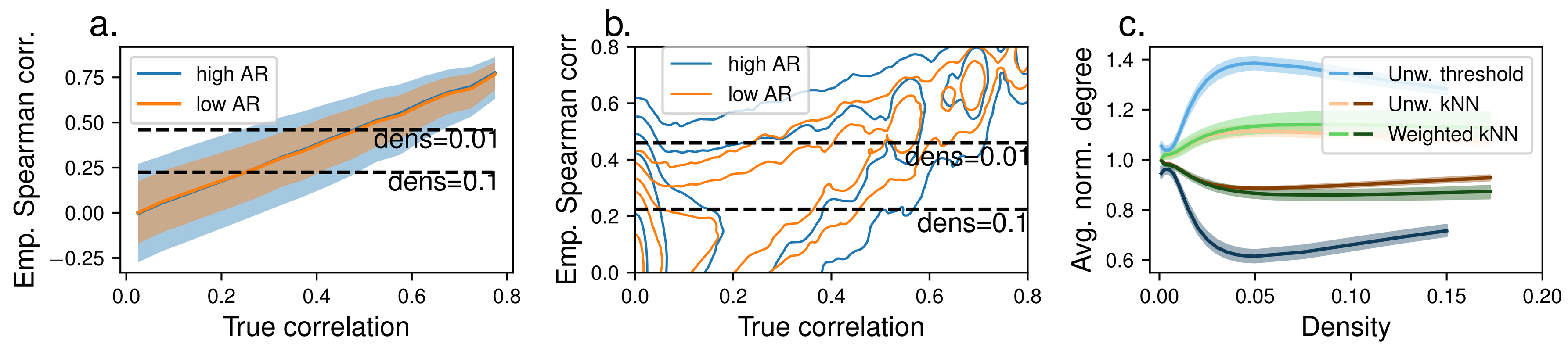
↳ Not detected by multiple empirical networks.

↳ Significance cannot be based on link-bundles.

Anisotropic marginals → **Anisotropic estimation variance on edges** → **Biased networks**

Anisotropic autocorrelation → **Anisotropic estimation variance on edges**

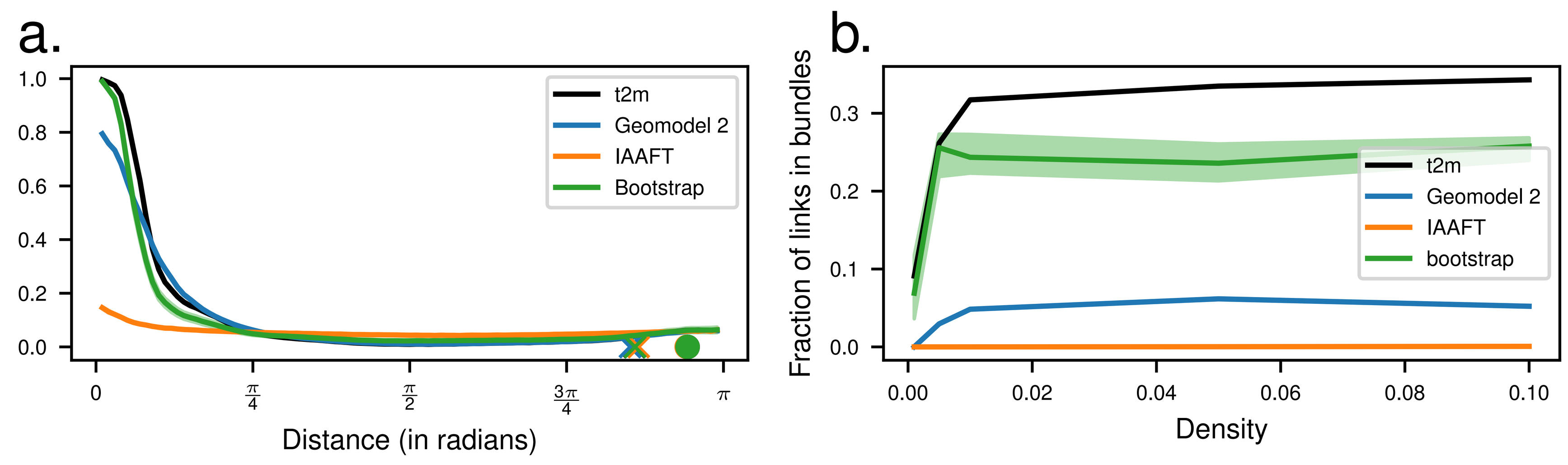
Anisotropic ... → **Anisotropic estimation variance on edges**



↳ Strong degree bias in real climate networks induced by anisotropic autocorrelation patterns.

4) Assessing Significance from Network Ensembles: Spatially Coherent Bootstrap on the Time Series

SOTA resampling procedures produce unrealistic network distributions:



SOTA procedures either preserve marginal time-series spectra or perform edge resampling.

↯ Not addressed: How robust is my network estimate? How large is the intrinsic variability?

↳ Data-based resampling that respects spatial dependencies:

- Multiple block-bootstrapping/subsampling estimates per edge to quantify edge-wise estimation variance,
- multiple equally valuable networks allow identifying recurring patterns.

Open question:

Which resampling procedure results in best finite-sample performance, consistently estimates anisotropic estimation variance and prevents biases from high dimension?

5) Conclusions

- Extreme **distortion of network measures** (degree, clustering coefficient, betweenness, ...).
- Localized correlation structure induces **spurious link-bundles** and **spuriously dense/sparse regions**.
- **Trade-off** for network density selection: Sparse networks have a lower fraction of false links, but single false links have higher impact.
- **Anisotropic estimation variance** biases empirical networks.

Simple Improvements:

- **Suitable estimators** can prevent many false links.
- **kNN graphs** can reduce biases and uncover different structures.

6) Future work

- **Resampling for networks** from spatio-temporal data,
- **Reassess** significance of several results in climate network literature,
- similar complex networks in **neuroscience**,
- In which ways are **PGMs and causal networks** distorted?

In Journal of Climate