



University  
of Glasgow

# COVARIANCE STRUCTURES FOR HIGH-DIMENSIONAL ENERGY FORECASTING

XXII Power Systems Computation Conference, Porto

---

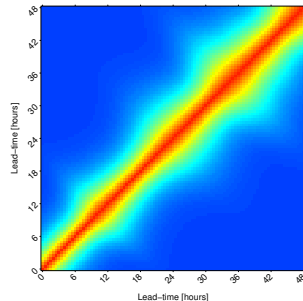
**Jethro Browell**, Ciaran Gilbert, Matteo Fasiolo

June 30, 2022

School of Mathematics and Statistics  
University of Glasgow  
jethro.browell@glasgow.ac.uk

1. Motivation: High-dimensional energy forecasting
2. Covariance functions and matrices
3. Generalised Additive Covariance Modelling
  - 3.1 Formulation
  - 3.2 Estimation
  - 3.3 Synthetic data example
  - 3.4 Wind power example
4. Summary

This work was supported by UKRI via EPSRC Innovation Fellowship *System-wide probabilistic energy forecasting* (EP/R023484/2).



## MOTIVATION

---

Decision-makers (TSOs, DSOs, traders) require forecasts of multiple quantities to operate efficiently and manage risk:

- How much demand will be met by wind and solar power tomorrow?
- What is the chance of power flows exceeding network capacity?
- Will I get a better price if I sell my power day-ahead or intraday?

Probabilistic forecasts quantify uncertainty by expressing predictions as probability density functions.

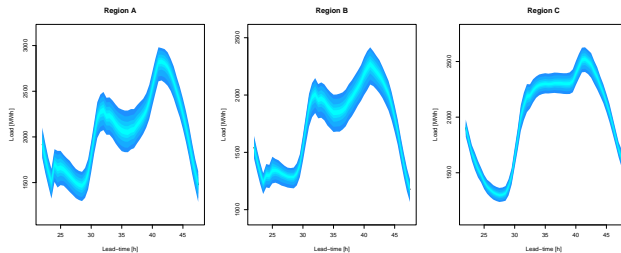


Figure 1: Fan plots of density forecasts for three locations and 48 time periods

⚠ These forecasts do not describe spatial or temporal dependency.

Probabilistic forecasts quantify uncertainty by expressing predictions as probability density functions.

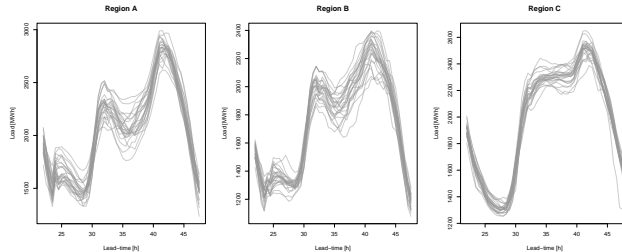


Figure 2: Space-time trajectories (or scenarios/samples) drawn from multivariate probabilistic forecast. These forecasts contain dependency information but are difficult to visualise.

This quickly becomes a **high-dimensional** problem!

Gaussian copulas provide a suitable framework for describing such high-dimensional predictive distributions:

- Margins of the copula are the familiar density forecasts
- Dependency structure specified by a covariance matrix,  $\Sigma$
- Scales well (from a modelling perspective) limited by estimation of the covariance matrix

Gaussian copulas provide a suitable framework for describing such high-dimensional predictive distributions:

- Margins of the copula are the familiar density forecasts
- Dependency structure specified by a covariance matrix,  $\Sigma$
- Scales well (from a modelling perspective) limited by estimation of the covariance matrix

The remainder of this talk is concerned with this **covariance matrix** and the possibility that it:

1. has a complex structure, and/or
2. varies over time, perhaps as a function of some covariate.



## COVARIANCE FUNCTIONS AND MATRICES

---

Consider a random process  $Z_t(\mathbf{s}, l)$  at location  $\mathbf{s}$ , forecast lead-time  $l$ , and forecast issue time  $t$ .

A covariance function,  $C_t$ , is *stationary* if the covariance

$$\text{cov}(Z_t(\mathbf{s}, l), Z_t(\mathbf{s} + \mathbf{h}, l + u)) = C_t(\mathbf{h}, u) \quad (1)$$

depends only on separation  $(\mathbf{h}, u)$ .

Consider a random process  $Z_t(\mathbf{s}, l)$  at location  $\mathbf{s}$ , forecast lead-time  $l$ , and forecast issue time  $t$ .

A covariance function,  $C_t$ , is *stationary* if the covariance

$$\text{cov}(Z_t(\mathbf{s}, l), Z_t(\mathbf{s} + \mathbf{h}, l + u)) = C_t(\mathbf{h}, u) \quad (1)$$

depends only on separation  $(\mathbf{h}, u)$ . Furthermore,  $C_t$  is *isotropic* if it is invariant to the direction of  $\mathbf{h}$  and  $u$

$$\text{cov}(Z_t(\mathbf{s}, l), Z_t(\mathbf{s} + \mathbf{h}, l + u)) = C_t(\|\mathbf{h}\|, |u|) \quad (2)$$

Table 1: Some parametric classes of isotropic covariance functions where  $C(\mathbf{h})$  takes the form  $C(\|\mathbf{h}\|; \boldsymbol{\xi})$ . The Whittle–Matérn covariance is defined in terms of the modified Bessel function of the second kind  $K_\nu$ .

Class	Function $C(r; \boldsymbol{\xi})$	Parameters $\boldsymbol{\xi}$
Powered Exponential	$\sigma^2 e^{-(\theta r)^\gamma}$	$0 < \gamma \leq 2; \theta > 0; \sigma \geq 0$
Whittle–Matérn	$\sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} (\theta r) K_\nu(\theta r)$	$\nu > 0; \theta > 0; \sigma \geq 0$
Cauchy	$\sigma^2 (1 + (\theta r)^\gamma)^{-\nu}$	$0 < \gamma \leq 2; \nu > 0; \theta > 0; \sigma \geq 0$
Spherical	$\sigma^2 \left( 1 - \frac{2}{\pi} \left( \frac{r}{\theta} \sqrt{1 - \left( \frac{r}{\theta} \right)^2} + \sin^{-1} \frac{r}{\theta} \right) \right)$	$c(r) = 0$ if $r > \theta; \sigma^2 \geq 0; \theta > 0$

Given the separation,  $(\|\mathbf{h}\|, |u|)$  between all pairs of variables  $i$  and  $j$ , given by  $R_{i,j}$ , the dynamic (time-dependent) covariance matrix  $\Sigma_t$  may be formed as

$$\Sigma_t = \begin{pmatrix} C_t(R_{1,1}) & C_t(R_{1,2}) & \dots & C_t(R_{1,p}) \\ C_t(R_{2,1}) & \ddots & & \vdots \\ \vdots & & & \\ C_t(R_{p,1}) & \dots & & C_t(R_{p,p}) \end{pmatrix} . \quad (3)$$

Therefore, we can specify a covariance matrix of arbitrary size by a covariance function, **with a small number of parameters**, and the known separation matrix,  $R$ .

However, what if the dependency structure we would like to model is

- Non-stationary, i.e.  $C_t(\cdot)$  depends on specific location  $\mathbf{s}$  or lead time  $l$ , or
- Dynamic, evolves over time or via a random process or via dependence on a time-varying covariate?

We can model these behaviours in a parsimonious fashion by allowing the parameters to covariance functions to be additive models of covariates (which may include  $\mathbf{s}$  and/or  $l$ ).

# GENERALISED ADDITIVE COVARIANCE MODELLING

---

Let  $C(r; \boldsymbol{\xi})$  be a covariance function parametrised by the  $m$ -dimensional parameter vector  $\boldsymbol{\xi}$ . For example,  $\boldsymbol{\xi} = \{\theta, \sigma, \gamma\}$  for the powered exponential in Table 1.

The elements of  $\boldsymbol{\xi}$  are modelled via

$$g_j(\xi_j) = \mathbf{A}_{j,t} \boldsymbol{\beta}_j + \sum_i f_{j,i}(\mathbf{x}_t^{S_{j,i}}), \quad \text{for } j = 1, \dots, m, \quad (4)$$

the details of which may be found in the paper, but this is essentially a Generalised Additive Model for each element of  $\boldsymbol{\xi}$ , the parameters of which,  $\boldsymbol{\beta}$ , are to be estimated.

*Aside: any learner may be considered here, though we proceed with GAMs for their flexibility and interoperability.*



Where the covariance matrix is static (no time dependence), the Weighted Least Squares loss of [Cressie, 1985, Cressie, 2015] may be used directly,

$$L_{\text{WLS}}^S(\boldsymbol{\beta}) = \sum_{i \neq j} \left( \frac{\hat{C}(R_{i,j}) - \hat{C}(R_{i,j}; \boldsymbol{\xi})}{1 - \hat{C}_{\text{cor}}(R_{i,j}; \boldsymbol{\xi})} \right)^2 ,$$

to estimate a parametric covariance function, giving greater weight to elements with larger correlation.

However, this must be adapted in the dynamic case as follows,

$$L_{\text{WLS}}^D(\beta) = \frac{1}{T} \sum_{t=1}^T \sum_{i \neq j} \left( \frac{[\mathbf{z}_t \otimes \mathbf{z}_t - \hat{\Sigma}(\boldsymbol{\xi}_t)]_{i,j}}{1 - [\hat{\Sigma}_{\text{cor}}(\boldsymbol{\xi}_t)]_{i,j}} \right)^2 ,$$

in order to take into account the potentially different covariance at different times.

In the present GAM setting, we also penalise the "wiggleness" of the splines,

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} L_{\text{WLS}}^{S/D}(\beta) + \sum_{k=1}^K \lambda_k J_k(\beta)$$

where the second term contains penalties of the form  $J_k(\beta) = \beta^\top \mathbf{S}_k \beta$ , with  $\mathbf{S}_k$  being a positive semi-definite matrix, while  $\lambda_k$  are positive tuning parameters to be chosen via cross-validation.

Here, we use an off-the-shelf optimisation to estimate  $\beta$ , specifically the BGFS algorithm.

## EXAMPLES

---

The only difference between models is the covariance structure, all margins/density forecasts are the same.

We use standard scoring rules for multivariate probabilistic forecasting:

- Multivariate Energy Score
- Log (or Ignorance) Score
- Variogram Score (with  $p = 0.5$  and  $p = 1$ )
- Skill scores and significance tests

The only difference between models is the covariance structure, all margins/density forecasts are the same.

We use standard scoring rules for multivariate probabilistic forecasting:

- Multivariate Energy Score
- Log (or Ignorance) Score
- Variogram Score (with  $p = 0.5$  and  $p = 1$ )
- Skill scores and significance tests

In the synthetic data experiments, where the true distribution is known, we can also calculate the Kullback–Leibler divergence (or relative entropy), which is not subject to sample variation like the forecast evaluation metrics above.

Consider the following exponential covariance function which depends on time through covariate  $x_t$

$$\begin{aligned}C_t(r) &= e^{-\theta(x_t)r} \quad , \\ \theta(x_t) &= \sin(2\pi x_t) + 2, \quad X \sim \mathcal{U}(0, 1) \quad .\end{aligned}$$

Consider the following exponential covariance function which depends on time through covariate  $x_t$

$$C_t(r) = e^{-\theta(x_t)r} ,$$
$$\theta(x_t) = \sin(2\pi x_t) + 2, \quad X \sim \mathcal{U}(0, 1) .$$

Experiment:

1. Simulate from the true structure
2. Estimate a GAC model with a linear structure (GAC-Linear) and a cubic regression spline (GAC-CR) for  $\theta(x_t)$  based on simulated data
3. Compare predictions from the true model, (static) empirical covariance, and the two GAC models on new simulated data



Consider the following exponential covariance function which depends on time through covariate  $x_t$

$$C_t(r) = e^{-\theta(x_t)r} ,$$

$$\theta(x_t) = \sin(2\pi x_t) + 2, \quad X \sim \mathcal{U}(0, 1) .$$

Experiment:

1. Simulate from the true structure
2. Estimate a GAC model with a linear structure (GAC-Linear) and a cubic regression spline (GAC-CR) for  $\theta(x_t)$  based on simulated data
3. Compare predictions from the true model, (static) empirical covariance, and the two GAC models on new simulated data

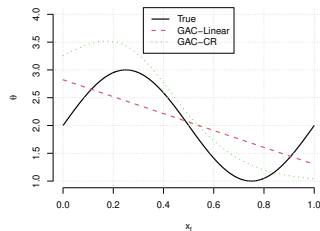


Figure 3: True function  $\theta(x_t)$  with GAC estimates based on synthetic data (5000 samples for six-dimensional data).

Table 2: Results of simulation experiment. Underline indicates that the corresponding skill score relative to the GAC-CR model are not significantly different from zero.

Name	Energy	Log	VS-0.5	VS-1	KL
Static Empirical	1.607	6.993	3.777	11.950	0.292
GAC-Linear	<u>1.606</u>	<u>6.930</u>	3.736	11.810	0.147
GAC-CR	<b>1.605</b>	<b>6.921</b>	<b>3.724</b>	<b>11.770</b>	<b>0.108</b>
True	<u>1.605</u>	6.870	3.697	11.670	0.000

The empirical covariance (no dynamics) performs relatively poorly, whereas the **flexible GAC model yields significant improvement**.

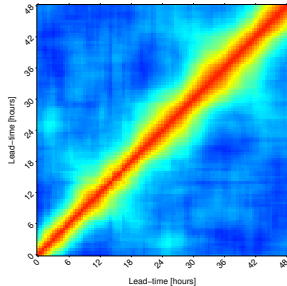


Figure 4: Empirical temporal dependency structure of wind power forecasts

The temporal dependency structure of wind power forecast is non-stationary and complex.

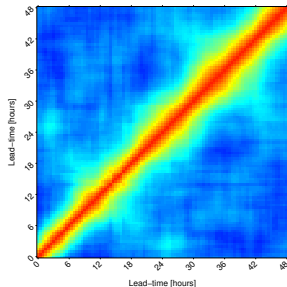


Figure 4: Empirical temporal dependency structure of wind power forecasts

The temporal dependency structure of wind power forecast is non-stationary and complex.

We model this with an Exponential correlation function and cubic splines where we allow  $\theta$  to be a smooth function of the distance  $d$  along the diagonal

$$\theta = \hat{\theta}_{cr}(d) = \beta_0 + f_{cr}(d) \quad .$$

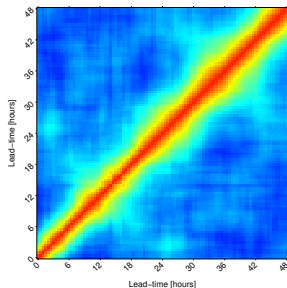


Figure 4: Empirical temporal dependency structure of wind power forecasts

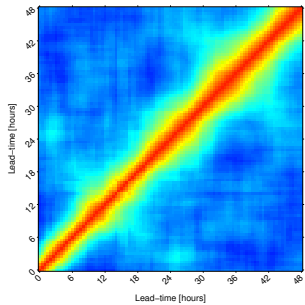
The temporal dependency structure of wind power forecast is non-stationary and complex.

We model this with an Exponential correlation function and cubic splines where we allow  $\theta$  to be a smooth function of the distance  $d$  along the diagonal

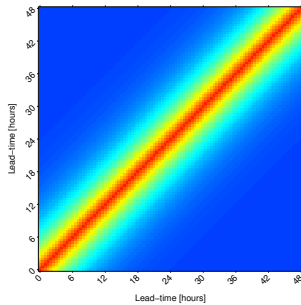
$$\theta = \hat{\theta}_{cr}(d) = \beta_0 + f_{cr}(d) \quad .$$

We loose the guarantee that  $\hat{\Sigma}$  will be positive-definite in the non-stationary case. Here we find the nearest PD matrix [Higham, 2002] if necessary, but other strategies may be considered.

# REAL DATA: TEMPORAL STRUCTURE IN WIND POWER FORECASTS

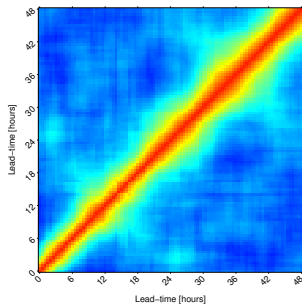


(a) Empirical

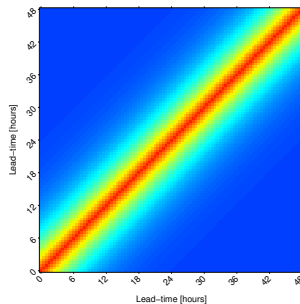


(b) Constant  $\hat{\theta}$  (Stationary)

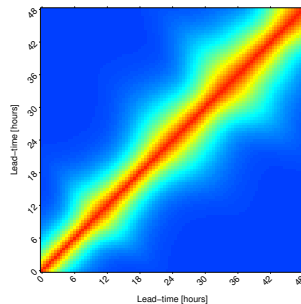
# REAL DATA: TEMPORAL STRUCTURE IN WIND POWER FORECASTS



(a) Empirical



(b) Constant  $\hat{\theta}$  (Stationary)



(c) GAC  $\hat{\theta}_{cr}(d)$

Figure 5: Temporal dependency structure of wind power forecasts from 0 to 48 hours-ahead. Forecasts have a visible non-stationary structure. The width of the diagonal ridge indicates how long forecast errors are likely to persist for in time.

Table 3: Results different temporal dependency models for wind power forecasting. Underline indicates that the corresponding skill score relative to the GAC model are not significantly different from zero.

Name	Energy	Log	VS-0.5	VS-1
Empirical	<u>7.139</u>	Inf	<u>1409</u>	<u>5444</u>
Constant	<u>7.142</u>	19.86	<u>1409</u>	<u>5439</u>
GAC	7.137	<b>15.46</b>	1406	5433



## SUMMARY

---

- We present an approach to model **dynamic** and **non-stationary** covariance structures by generalising parametric covariance functions
- Doing so may **substantially improve** the quality of multi-variate probabilistic forecasts (and other covariance-based models!)
- This approach de-couples the complexity of covariance model from dimension of process being modelled !!!
- There is much still to be done to understand and improve model selection and estimation...

Full details and an additional synthetic data example are in the paper. Data and code available at [10.5281/zenodo.5541782](https://doi.org/10.5281/zenodo.5541782)

*See also the similar approach proposed in [Muschinski et al., 2022] alongside a decomposition-based method.*

Paper, slides, code and more linked from [www.jethrobrowell.com](http://www.jethrobrowell.com)

### References:



Cressie, N. (1985).

Fitting variogram models by weighted least squares.

*Journal of the International Association for Mathematical Geology* 1985 17:5, 17(5):563–586.



Cressie, N. A. (2015).

Statistics for spatial data revised edition.

*Statistics for Spatial Data*, pages 1–900.



Higham, N. J. (2002).

Computing the nearest correlation matrix: a problem from finance.

*IMA Journal of Numerical Analysis*, 22(3):329–343.



Muschinski, T., Lang, M. N., Mayr, G. J., Messner, J. W., Zeileis, A., and Simon, T. (2022).

Predicting power ramps from joint distributions of future wind speeds.

*Wind Energy Science Discussions*, 2022:1–20.

Source of this theme and the demo presentation from [github.com/matze/mtheme](https://github.com/matze/mtheme), licensed under CC BY-SA 4.0.

Q & A