

---

# Grouped Gaussian Processes for Solar Forecasting

---

**Astrid Dahl\***

School of Computer Science and Engineering  
University of New South Wales  
Sydney, Australia  
\*astridmdahl@gmail.com

**Edwin V. Bonilla**

Data61  
Australia

## Abstract

We propose a Gaussian process regression network model where latent functions are arbitrarily coupled *a priori*. Driven by the problem of developing forecast methods for distributed solar and other renewable power generation, we propose coupled priors that exploit spatial dependencies in a scalable structure. We estimate short term forecast models for solar power at multiple distributed sites and ground wind speed at proximate weather stations. Our approach maintains or improves point-forecast accuracy relative to competing benchmarks. At the same time our approach significantly reduces forecast variance.

## 1 Introduction

Local short term solar power forecasts are an important input for distributed grid control, energy market and home energy management [2, 11]. Output variation is driven by two main factors: variation due to sun angle and distance, and variation due to weather effects. Both induce spatially-related dependence between proximate sites. In the context of small scale residential sites, data on system configuration and local environment are often limited, motivating approaches that do not rely on rich history or feature sets as are typically required by current approaches [2, 5, 10, 11, 13]. Also inherent to the application is a need to minimize and model forecast uncertainty in a principled way [2]. Short term wind power forecasting is similarly critical to energy technologies [11]. Variability is driven by wind speed which, as for solar, is driven by interacting environmental factors giving rise to spatial dependencies *a priori*.

Gaussian process (GP) models are a flexible nonparametric Bayesian approach and have been extended to numerous multi-task problems including spatio-temporal contexts [3, 9]. Multi-task GP methods have been developed along several lines [see e.g. 1, for a review], with scalability a core challenge. For this reason, latent Gaussian processes in mixing-based multi-task methods are generally constrained to be statistically independent [8, 12]. In particular, the Gaussian process regression network (GPRN) model of Wilson et al. [12] assumes observations are a linear combination of several node and weight latent functions that are all drawn from independent univariate Gaussian process priors.

We build on the scalable generic inference method of [8] to extend the model of [12] and allow feature-driven, nonzero covariance between arbitrary subsets, or ‘groups’, of latent functions. The grouping structure is flexible and can be tailored to applications of interest. By adopting separable kernel specifications, we maintain scalability of the approach. We apply our model to forecast solar output of distributed residential sites and use spatial features to exploit spatial dependence between latent functions. We also test our method on a ground wind speed dataset. Our results show that, for solar models, exploiting spatial covariance maintains or improves point-forecast accuracy relative to benchmarks and at the same time reduces forecast variance. Further, wind forecast accuracy and uncertainty is improved on all measures.

## 2 Grouped Gaussian process model and inference

We consider data of the form  $\{\mathbf{X} \in \mathbb{R}^{N \times D}, \mathbf{Y} \in \mathbb{R}^{N \times P}\}$  where each  $\mathbf{x}_n$  in  $\mathbf{X}$  is a  $D$ -dimensional vector of input features and each  $\mathbf{y}_{(n)}$  in  $\mathbf{Y}$  is a  $P$ -dimensional vector of task outputs,  $n = 1, \dots, N$ . Let  $\mathbf{F}$  be a  $N \times Q$  matrix of latent GP functions  $\{f_j(\mathbf{x})\}_{j=1}^Q$  ( $\mathbf{f}_j$  is the  $N$ -dimensional vector for latent function  $j$ ) and let  $\mathbf{W} \in \mathbb{R}^{P \times Q_g \times N}$  and  $\mathbf{G} \in \mathbb{R}^{Q_g \times 1 \times N}$  be tensors formed from  $PQ_g$  and  $Q_g$  latent functions in  $\mathbf{F}$ , respectively, with  $Q_g(P+1) = Q$ . We can express the likelihood of Wilson et al. [12] as  $p(\mathbf{y}_{(n)} | \mathbf{f}_{(n)}, \phi) = \mathcal{N}(\mathbf{y}_{(n)}; \mathbf{W}_{(n)} \mathbf{g}_{(n)}, \Sigma_y)$ , where  $\phi = \Sigma_y$ ,  $\mathbf{f}_{(n)} = \{\mathbf{W}_{(n)}, \mathbf{g}_{(n)}\}$  and  $\Sigma_y$  is a diagonal matrix.  $P$ -dimensional outputs are constructed at  $\mathbf{x}_n$  as the product of a  $P \times Q_g$  matrix of weight functions,  $\mathbf{W}_{(n)}$ , and  $Q_g$ -dimensional vector of node functions  $\mathbf{g}_{(n)}$  (Figure 1 in the Appendix).

To group latent GP functions we re-express  $\mathbf{F}$  in terms of arbitrarily chosen submatrices  $\mathbf{F}_r \in \mathbb{R}^{N \times Q_r}$ ,  $r = 1, \dots, R$ , where  $R$  is the number of groups and  $Q_r$  is the number of functions in group  $r$  ('group size'),  $\sum_{r=1}^R Q_r = Q$ . The prior on  $\mathbf{F}$  can be expressed as  $p(\mathbf{F} | \theta) = \prod_{r=1}^R \mathcal{N}(\mathbf{F}_r; \mathbf{0}, \mathbf{K}_{ff}^r)$ , where  $\mathbf{K}_{ff}^r \in \mathbb{R}^{NQ_r \times NQ_r}$  is the covariance matrix generated by the group kernel function  $\kappa_r(f_j(\mathbf{x}), f_{j'}(\mathbf{x}'))$  evaluated for functions  $f_j$  and  $f_{j'}$  at locations  $\mathbf{x}$  and  $\mathbf{x}'$ .  $\kappa_r(f_j(\mathbf{x}), f_{j'}(\mathbf{x}')) = 0$  iff the functions  $f_j$  and  $f_{j'}$  do not belong to the same group  $r$ . We refer to this model as grouped Gaussian processes (GGP).

Our inference method extends the framework of Krauth et al. [8], which is a sparse variational method assuming conditionally independent latent functions, to the case where latent functions covary within groups. The prior on  $\mathbf{F}$  is augmented with inducing variables,  $\{\mathbf{u}_r\}_{r=1}^R$ , drawn from the same GP priors as  $\mathbf{F}_r$  at new inducing points  $\mathbf{Z}_r \in \mathbb{R}^{M \times D}$ . The prior above is thus replaced by

$$p(\mathbf{u} | \theta) = \prod_{r=1}^R \mathcal{N}(\mathbf{u}_r; \mathbf{0}, \mathbf{K}_{uu}^r), \quad \text{and} \quad p(\mathbf{F} | \mathbf{u}) = \prod_{r=1}^R \mathcal{N}(\mathbf{F}_r; \tilde{\boldsymbol{\mu}}_r, \tilde{\mathbf{K}}_r), \quad (1)$$

where  $\tilde{\boldsymbol{\mu}}_r = \mathbf{A}_r \mathbf{u}_r$ ,  $\tilde{\mathbf{K}}_r = \mathbf{K}_{ff}^r - \mathbf{A}_r \mathbf{K}_{uf}^r$  and  $\mathbf{A}_r = \mathbf{K}_{fu}^r (\mathbf{K}_{uu}^r)^{-1}$ .

We use separable kernels of the form  $\kappa_r(f_j(\mathbf{x}), f_{j'}(\mathbf{x}')) = \kappa_r(\mathbf{x}, \mathbf{x}') \kappa_r(\mathbf{h}_j, \mathbf{h}_{j'})$ , where  $\mathbf{h}$  are  $H$ -dimensioned feature vectors forming matrices  $\mathbf{H}_r \in \mathbb{R}^{Q_r \times H}$  that govern covariance across functions  $\mathbf{f}_j \in \mathbf{F}_r$ .  $\mathbf{K}_{uu}^r \in \mathbb{R}^{MQ_r \times MQ_r}$  is the covariance matrix induced by  $\kappa_r(f_j(\mathbf{x}), f_{j'}(\mathbf{x}'))$  evaluated over  $\{\mathbf{Z}_r, \mathbf{H}_r\}$ , yielding  $\mathbf{K}_{uu}^r = \mathbf{K}_{hh}^r \otimes \mathbf{K}_{zz}^r$  and importantly the decomposition  $(\mathbf{K}_{uu}^r)^{-1} = (\mathbf{K}_{hh}^r)^{-1} \otimes (\mathbf{K}_{zz}^r)^{-1}$ . We define  $\mathbf{K}_{fu}^r$  and  $\mathbf{K}_{uf}^r$  similarly. By adopting the Kronecker-structure and further embedding this in a sparse variational framework using  $M$  inducing points, we reduce the maximum dimension of required matrix inversions from  $NQ_r$  to  $\max(Q_r, M)$ ,  $M \ll N$ .

Latent functions in the general framework need not map to a particular feature set. We solve this by setting  $Q_g = P$  and grouping latent functions  $\mathbf{f}_j$  according to rows of  $\mathbf{W}_{(n)}$ , and defining spatial features  $\mathbf{h}_j = (\textit{latitude}_j, \textit{longitude}_j)$ . Node functions in  $\mathbf{G}$  are assumed to be independent i.e.  $\langle \mathbf{g}_j, \mathbf{g}_{j'} \rangle = 0$  for  $i \neq i'$  with features for  $\mathbf{g}_j$  relating to task  $j$ . Thus, in addition to depending on input features  $\mathbf{x}_{(n)}$ , weights over node functions for each task are also spatially smoothed via  $\mathbf{K}_{hh}^r$ . This structure allows both task-specific parameterization and regularization over weights in the prior.

The (analytically intractable) joint posterior distribution of the latent functions and inducing variables under the prior and likelihood models above is approximated via variational inference [6]. Specifically,  $p(\mathbf{F}, \mathbf{u} | \mathbf{Y}) \approx q(\mathbf{F}, \mathbf{u} | \boldsymbol{\lambda}) \stackrel{\text{def}}{=} p(\mathbf{F} | \mathbf{u}) q(\mathbf{u} | \boldsymbol{\lambda})$ . The variational posterior  $q(\mathbf{u} | \boldsymbol{\lambda})$  is defined generally as a mixture of  $K$  Gaussians with mixture proportions  $\pi_k$  i.e.  $q(\mathbf{u} | \boldsymbol{\lambda}) = \sum_{k=1}^K \pi_k \prod_{r=1}^R \mathcal{N}(\mathbf{u}_r; \mathbf{m}_{kr}, \mathbf{S}_{kr})$ . We then estimate the model by maximizing the so-called evidence lower bound. Under our structure, (Gaussian) posterior mixture covariances for  $\mathbf{f}_{r(n)}$ ,  $\Sigma_{kr(n)} \in Q_r \times Q_r$ , can be shown to decompose as:

$$\begin{aligned} \Sigma_{kr(n)} &= \tilde{\mathbf{K}}_{r(n)} + \mathbf{A}_{r(n)} \mathbf{S}_{kr} \mathbf{A}'_{r(n)}, \quad \text{where} \quad \mathbf{A}_{r(n)} = \mathbf{I}_{Q_r} \otimes \kappa_r(\mathbf{x}_{(n)}, \mathbf{Z}_r) (\mathbf{K}_{zz}^r)^{-1}, \quad \text{and} \\ \tilde{\mathbf{K}}_{r(n)} &= \mathbf{K}_{hh}^r \times [\kappa_r(\mathbf{x}_{(n)}, \mathbf{x}_{(n)}) - \kappa_r(\mathbf{x}_{(n)}, \mathbf{Z}_r) \mathbf{K}_{zz}^r \kappa_r(\mathbf{Z}_r, \mathbf{x}_{(n)})]. \end{aligned} \quad (2)$$

Predictive posterior distributions over  $q_k(\mathbf{f}_* | \boldsymbol{\lambda}_k)$  are defined similarly. Hence, a nice property of our model is that cross-function covariance within a group can be driven by spatial features (or other covariates) from the prior, where  $\mathbf{S}_{kr}$  is diagonal, or more flexible in form where  $\mathbf{S}_{kr}$  is non-diagonal.

### 3 Experiments

We evaluate GGP on forecasting problems for residential solar power and ground wind speed datasets. For solar we forecast power fifteen minutes ahead using five minute average power for three datasets: ten Adelaide sites (ADEL-AUTM) and twelve Sydney sites (SYD-AUTM), both over 60 days during Autumn 2016; and ten Adelaide sites (ADEL-SUMM) over 60 days in Spring-Summer 2016. We train all models on 36 days of data, and test forecast accuracy for 24 subsequent days (days are defined as 7 am to 7pm). For WIND we forecast ground wind speed thirty minutes ahead at six weather stations in Victoria, Australia. Data are half-hourly wind speed readings collected over an eight month period. The WIND data present an interesting challenge, with frequent missing and noisy observations. In all we have 5000 (4000) training points and 3636 (1024) test points per site for solar (per station for wind). Datasets have varying spatial dispersions ranging from 15 by 20 to 30 by 40 kilometer areas.

Kernels and features for  $\kappa_r(\mathbf{x}, \mathbf{x}')$  and  $\kappa_r(\mathbf{h}_j, \mathbf{h}_{j'})$  are selected in line with previous studies of distributed solar forecasting [4, 5, 10]. We define  $\kappa_{g_j}(\mathbf{x}_t, \mathbf{x}_s) = \kappa_{g_j}(\mathbf{l}_t, \mathbf{l}_s)$  as a radial basis function kernel ( $\kappa_{RBF}$ ) applied to a vector of recent lagged power at site  $j$ , i.e. for site  $j$  at time  $t$ ,  $\mathbf{l}_{j,t} = (y_{j,t}, y_{j,t-1}, y_{j,t-2})$ . For row-group (task)  $r$ , we define a multiplicative kernel structure with  $\kappa_r(\mathbf{x}, \mathbf{x}') = \kappa_{Per.}(t, s)\kappa_{RBF}(\mathbf{l}_{rt}, \mathbf{l}_{rs})$ , where  $\kappa_{Per.}(t, s)$  is a periodic kernel on a time index  $t$  capturing diurnal cyclical trends in output. For  $\kappa_r(\mathbf{h}_j, \mathbf{h}_{j'})$  we use a compact RBF-Epanechnikov structure, i.e.,  $\kappa_r(\mathbf{h}_j, \mathbf{h}_{j'}) = \kappa_{RBF}(\mathbf{h}_j, \mathbf{h}_{j'})\kappa_{Ep.}(\mathbf{h}_j, \mathbf{h}_{j'})$ ,  $j, j' = 1 \dots P$ , allowing cross-site weights to reduce to zero at a distance optimized for each task. For WIND we use the same kernel and feature definitions as for solar but a different grouping structure for GGP. We allow functions on the diagonal of  $\mathbf{W}$  to be independent and group off-diagonal functions within each row.

We estimate several benchmarks: (1) independent GP forecast models for each site (IGP), (2) pooled multi-task models with task-specific (spatial) features (MTG), (3) multi-task linear coregional models (LCM) and (4) GPRN with independent latent functions [12]. Both IGP and MTG models have univariate Gaussian likelihood functions, while the LCM is comprised of  $P$  node functions mapped to outputs via a  $P \times Q_g$  matrix of deterministic weights, i.e.  $p(\mathbf{y}_{(n)}|\mathbf{f}_{(n)}, \phi) = \mathcal{N}(\mathbf{y}_{(n)}; \mathbf{W}_{(n)}\mathbf{g}_{(n)}, \Sigma_y)$  where  $\mathbf{W}_{(n)ij} = w_{ij} \quad \forall n = 1, \dots, N$ . We maintain consistent kernel specification across models and present results for diagonal and full Gaussian posterior specifications (for GGP we use a Kronecker construction of the full posterior for each group).

To compare model performance under equivalent settings, we standardize settings by reference to a consistent target computational complexity per iteration, which in our variational framework is dominated by operations with cubic complexity on the number of inducing points  $M$ . We therefore set  $M = 200$  per group for GGP and adjust  $M$  for each benchmark accordingly. We optimize the ELBO iteratively using ADAM [7] until its relative change over successive epochs is less than  $10^{-5}$  up to a maximum of 200 epochs. All data except time index features are normalized. Reported forecast accuracy measures are root mean squared error (RMSE) and negative log predictive density (NLPD) which is estimated using Monte Carlo. In addition, we compute average ranking (M-RANK) over RMSE and NLPD and mean forecast variance (F-VAR).

**Results** Results for all models are presented at Table 1. For solar, GGP maintains or improves point accuracy when compared to best performing benchmarks on both RMSE and NLPD individually. For RMSE, accuracy under GGP differs by less than one percent relative to GPRN, and similarly matches or improves on NLPD relative to LCM and other benchmarks. GGP performs well in terms of overall accuracy across both measures, consistently achieving the highest average rank across both measures (M-RANK). In contrast, competing baselines either perform well on RMSE at the expense of poor performance under NLPD or *vice versa*. The benefit of regularization under the GGP is clear when considering mean forecast variance, which is lower under GGP than all benchmark models for all experiments. Compared to GPRN (LCM), the most accurate GGP model reduces solar forecast variance by 18 to 24 (13 to 40) percent. Orientation of spatial kernels and compact kernel support were found to vary markedly across groups, confirming the relevance of flexible, site-specific parameterization.

We test statistical significance of differences in performance via 95 percent intervals estimated by Monte Carlo. Results show that RMSE under GPRN is statistically significantly lower than under GGP for solar datasets. In all other cases, RMSE is either not significantly different or significantly higher than under GGP. Results are similar for NLPD, which is statistically significantly lower under LCM for two of three datasets, and otherwise higher or not significantly different.

Table 1: Forecast accuracy and variance of GGP and benchmark models using diagonal (D) and full (F) Gaussian posteriors. M-RANK is mean of RMSE and NLPD ranks and F-VAR is mean forecast variance. Lower values indicate better performance for all measures. \*indicates significantly different from best GGP model ((D) or (F)) based on 95 percent interval from Monte Carlo analysis.

	ADEL-AUTM				ADEL-SUMM			
	RMSE	NLPD	M-RANK	F-VAR	RMSE	NLPD	M-RANK	F-VAR
GGP (D)	0.282	0.243	<b>2.5</b>	0.140	0.318	0.323	<b>2.5</b>	0.118
GGP (F)	0.288 *	0.265 *	4.0	<b>0.136</b> *	0.321 *	0.352 *	4.0	<b>0.113</b> *
LCM (D)	0.294 *	0.240	4.0	0.162 *	0.325 *	0.332 *	4.5	0.165 *
LCM (F)	0.293 *	<b>0.240</b> *	3.0	0.160 *	0.323 *	<b>0.323</b>	3.0	0.158 *
GPRN (D)	<b>0.278</b> *	0.311 *	3.0	0.173 *	<b>0.315</b> *	0.376 *	3.0	0.152 *
GPRN (F)	0.283	0.320 *	4.5	0.174 *	0.316 *	0.382 *	4.0	0.152 *
MTG (D)	0.301 *	0.337 *	7.0	0.174 *	0.444 *	0.675 *	10.0	0.256 *
MTG (F)	0.304 *	0.376 *	9.0	0.206 *	0.441 *	0.674 *	9.0	0.267 *
IGP (D)	0.315 *	0.368 *	9.0	0.177 *	0.341 *	0.415 *	7.5	0.153 *
IGP (F)	0.314 *	0.370 *	9.0	0.183 *	0.343 *	0.414 *	7.5	0.156 *

	SYD-AUTM				WIND			
	RMSE	NLPD	M-RANK	F-VAR	RMSE	NLPD	M-RANK	F-VAR
GGP (D)	0.284	<b>0.257</b>	<b>2.5</b>	0.157	0.454 *	0.670 *	2.5	0.282 *
GGP (F)	0.298 *	0.286 *	6.0	<b>0.142</b> *	<b>0.450</b>	<b>0.661</b>	<b>1.0</b>	<b>0.281</b>
LCM (D)	0.310 *	0.273 *	6.5	0.180 *	0.464 *	0.672 *	4.0	0.300 *
LCM (F)	0.302 *	0.257	5.5	0.178 *	0.465 *	0.676 *	5.0	0.305 *
GPRN (D)	0.281 *	0.323 *	3.5	0.185 *	0.453 *	0.685 *	3.5	0.301 *
GPRN (F)	0.284	0.326 *	5.5	0.187 *	0.461 *	0.700 *	5.0	0.304 *
MTG (D)	<b>0.280</b>	0.342 *	5.0	0.207 *	0.474 *	0.737 *	10.0	0.353 *
MTG (F)	0.283	0.360 *	6.5	0.219 *	0.473 *	0.729 *	8.0	0.337 *
IGP (D)	0.286	0.340 *	7.5	0.204 *	0.472 *	0.721 *	7.0	0.336 *
IGP (F)	0.286	0.335 *	6.5	0.202 *	0.473 *	0.731 *	9.0	0.345 *

For the WIND dataset, GGP outperforms all other models on all measures including point accuracy (NLPD and RMSE), overall accuracy (M-RANK) and forecast variance. Consistent reductions in variance are observed, ranging from 7 to 25 percent improvements over competing models. As for solar, confidence intervals are constructed via Monte Carlo. For WIND, all differences in model performance are confirmed to be statistically significant.

## 4 Discussion

We have proposed a general multi-task GP model, where groups of functions are coupled *a priori*. Our approach allows for input-varying covariance across tasks governed by kernels and features and, by building upon sparse variational methods and exploiting Kronecker structures, our inference method is inherently scalable to a large number of observations.

We have shown the applicability of our approach to forecasting short term distributed solar power and wind speed at multiple locations, where it matches or improves point forecast performance of single-task learning approaches and other multi-task baselines under similar computational constraints while improving quantification of predictive variance. In general, the GGP strikes a balance between flexible, task-specific parameterization and effective regularization via structure imposed in the prior.

### Acknowledgments

This research was conducted with support from the Cooperative Research Centre for Low-Carbon Living in collaboration with the University of New South Wales and Solar Analytics Pty Ltd.

### References

- [1] Álvarez, M. A., Rosasco, L., and Lawrence, N. D. (2012). Kernels for vector-valued functions: A review. *Found. Trends Mach. Learn.*, 4(3):195–266.

- [2] Antonanzas, J., Osorio, N., Escobar, R., Urraca, R., de Pison, F. M., and Antonanzas-Torres, F. (2016). Review of photovoltaic power forecasting. *Solar Energy*, 136:78 – 111.
- [3] Cressie, N. and Wikle, C. K. (2011). *Statistics for spatio-temporal data*. John Wiley & Sons.
- [4] Dahl, A. and Bonilla, E. (2017). Scalable Gaussian process models for solar power forecasting. In Woon, W. L., Aung, Z., Kramer, O., and Madnick, S., editors, *Data Analytics for Renewable Energy Integration: Informing the Generation and Distribution of Renewable Energy*, pages 94–106. Springer International Publishing.
- [5] Inman, R. H., Pedro, H. T., and Coimbra, C. F. (2013). Solar forecasting methods for renewable energy integration. *Progress in energy and combustion science*, 39(6):535–576.
- [6] Jordan, M. I., Ghahramani, Z., Jaakkola, T. S., and Saul, L. K. (1998). *An introduction to variational methods for graphical models*. Springer.
- [7] Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980.
- [8] Krauth, K., Bonilla, E. V., Cutajar, K., and Filippone, M. (2017). AutoGP: Exploring the capabilities and limitations of Gaussian process models. In *Artificial Intelligence and Statistics*.
- [9] Rasmussen, C. E. and Williams, C. K. I. (2006). *Gaussian processes for machine learning*. The MIT Press.
- [10] Voyant, C., Notton, G., Kalogirou, S., Nivet, M.-L., Paoli, C., Motte, F., and Fouilloy, A. (2017). Machine learning methods for solar radiation forecasting: A review. *Renewable Energy*, 105:569–582.
- [11] Widén, J., Carpman, N., Castellucci, V., Lingfors, D., Olauson, J., Remouit, F., Bergkvist, M., Grabbe, M., and Waters, R. (2015). Variability assessment and forecasting of renewables: A review for solar, wind, wave and tidal resources. *Renewable and Sustainable Energy Reviews*, 44:356–375.
- [12] Wilson, A. G., Knowles, D. A., and Ghahramani, Z. (2012). Gaussian process regression networks. In *International Conference on Machine Learning*.
- [13] Yang, D., Kleissl, J., Gueymard, C. A., Pedro, H. T., and Coimbra, C. F. (2018). History and trends in solar irradiance and pv power forecasting: A preliminary assessment and review using text mining. *Solar Energy*.

## A Supplementary Information

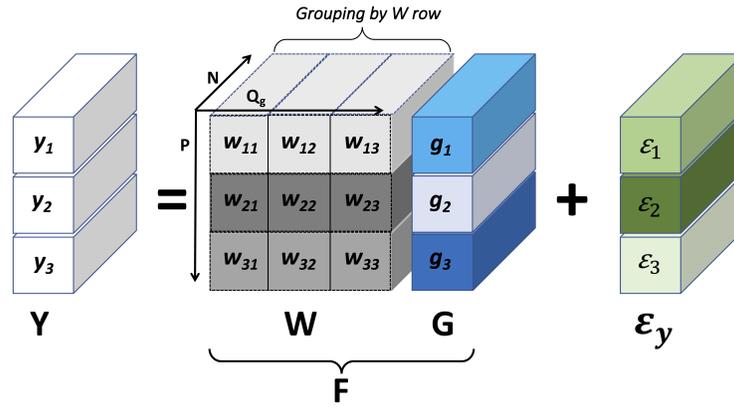


Figure 1: Gaussian process regression network model where  $\mathbf{Y}$  is a linear combination of node and weight latent functions comprising  $\mathbf{F}$ . In the grouped Gaussian process (GGP) framework, latent functions may be grouped arbitrarily. A grouping scheme is illustrated where weight functions in  $\mathbf{W}$  are grouped by rows (grouped functions are shown in the same shade) and given a fully-coupled prior, while node functions in  $\mathbf{G}$  are independent. Here  $N$  is the number of observations per task;  $P$  is the number of tasks; and  $Q_g$  is the group size.