Supplementary Material

Bayesian linear regression

Bayesian linear regression is an approach to ordinary linear regression (OLR) within a Bayesian framework. The OLR model is typically denoted as:

$$y = \beta_0 + \beta_1 x + \varepsilon \tag{A1}$$

where x are the explanatory variables, y the dependent variables, β_0 and β_1 the regression coefficients, and ε are the regression error, assumed to be independent and identically distributed, expressed as $\sim N(0, \sigma^2)$. Another assumption of OLR is that the x are independent of ε. The objective of OLR is to calculate values of β_0 and β_1 that minimize (in the least-square sense) the error between the observed values of y and β_0 and $\beta_1 x$. In a Bayesian framework, this model structure can also be represented as a Normal probability model:

$$y \sim N(X\beta, \ \sigma^2 I) \tag{A2}$$

where X is matrix of dependent variables, limited to $[x_1 \ x_2]$ here with $x_1 = I$ to allow for an intercept, and x_2 are the observations, the β are the corresponding regression coefficients, σ^2 is the variance of the dependent variable y given the observations (*i.e. $y|\beta$, X*), and lastly, I is the identity matrix. In a Bayesian framework, this function ($y|\beta$, σ^2, X) is the likelihood function. Note that Bayes rule states that the posterior distribution is proportional to the product of the likelihood function and the prior distribution. Thus, OLR when viewed through a Bayesian framework can be written as:

$$p(\beta, \sigma^2 \mid y, X) \propto p(y \mid \beta, \sigma^2, X) p(\beta, \sigma^2) \tag{A3}$$

where the term on the left hand side represents the posterior, and the right hand side is the product of the likelihood and the prior. The likelihood in this case is a multivariate Normal, and is often more convenient to be parameterized using the following terms (the derivation is not shown here but available in many econometric textbooks including Zellner, 1971; Birkes and Dodge, 1993; Koop, 2003; Greenberg, 2008; Gelman et al., 2014):

$$\hat{\beta} = (X^T X)^{-1} X^T y \tag{A4}$$

$$s^2 = (y - X\hat{\beta})^T (y - X\hat{\beta}) / v \tag{A5}$$

These values are identical to the least-squares estimates of degrees of freedom ($v = N - 2$) (with N the total number of observations), β and the standard error, s^2.

Natural conjugate priors

In Bayesian analysis, the prior is meant to reflect any previously available data or information about the parameters of interest. The choice of prior used determines whether or not there is an analytical solution for the posterior distribution. In the case where no analytical solution is possible, numerical integration techniques, using Monte Carlo simulations, such as the Gibb's sampler are needed. A particular class of priors, known as *natural conjugate priors*, are typically used because when they are combined with the likelihood function, these yield a posterior of the same class of distributions (Birkes and Dodge, 1993; Koop, 2003). This ensures that an analytical solution for the posterior is possible. For the Normal regression model, the natural conjugate prior can be stated as:

$$p(\beta, \sigma^2) = p(\beta \mid \sigma^2) p(\sigma^2) \tag{A6}$$

or:

$$p(\beta, \tau) = p(\beta \mid \tau) p(\tau) \tag{A7}$$

where $\tau = 1/\sigma^2$, and $(\beta|\tau) \sim N(\underline{\beta}, \tau^{-1}\underline{V})$ and $\tau \sim \Gamma(\underline{s}^{-2}, \underline{v})$, and the underlined terms represent the prior hyper-parameters. This means that the joint prior is Normal-Gamma: $(\beta, \tau) \sim N\Gamma(\underline{\beta}, \underline{V}, \underline{s}^{-2}, \underline{v})$). Multiplying this prior with the likelihood function results in the joint posterior of β and τ that is also Normal-Gamma (as expected because of the choice of *natural conjugate priors*):

$$(\beta, \tau \mid y) \sim N\Gamma(\overline{\beta}, \ \overline{V}, \ \overline{s}^{-2}, \ \overline{v}) \tag{A8}$$

where the bar above parameter denotes the "posterior" value of the hyper-parameters. These hyper-parameters can then be estimated as follows:

$$\overline{V} = (\underline{V}^{-1} + X^T X)^{-1} \tag{A9}$$

$$\overline{\beta} = \overline{V}(\underline{V}^{-1}\underline{\beta} + X^T X \beta) \tag{A10}$$

$$\overline{v} = \underline{v} + N \tag{A11}$$

And $\overline{s}^{-2}$ can be calculated from:

$$\overline{V}\overline{s}^{-2} = \underline{v}\underline{s}^{-2} + vs^{-2}(\hat{\beta} - \beta)^T[\underline{V} + (X^T X)^{-1}]^{-1}(\hat{\beta} - \beta) \tag{A12}$$

In essence, this form of BLR results in model estimates (of β and τ) that are weighted averages of the provided priors and the least-square estimates (West, 1984). This formulation is the standard BLR form with a *natural conjugate prior*.

Non-informative priors

It can be readily seen from the above formulation that the impact of the selection of priors can be extremely significant, and can severely influence the posterior values of the model parameters. Thus, often a single value for a prior is rarely used; instead an upper or lower limit is used to get an estimate of parameters. However, in many cases there is no prior information available, and thus this type of prior is unrealistic or subjective. Thus, an objective prior is needed that can represent complete uncertainty regarding the model parameters – this type of prior is known as a *non-informative* prior. To represent this in the formulation above, the values of $\underline{v}$ and $\underline{V}^{-1}$ are both set at 0. This results in posterior estimates of the hyper-parameters as follows:

$$\overline{V} = (X^T X)^{-1} \tag{A13}$$

$$\overline{\beta} = \beta \tag{A14}$$

$$\overline{v} = N \tag{A15}$$

And $\overline{s}^{-2}$ can be calculated from:

$$\overline{v}\overline{s}^{-2} = vs^{-2} \tag{A16}$$

Note that when using non-informative priors, the posterior estimates of the hyper-parameters reduce to the familiar least-squares parameter estimates β and $\overline{s}^{-2}$. Using this formulation the predictive distribution of y can be found analytically and is multi-variate *t*:

$$\left(y^* \mid y\right) \sim t(X^*\overline{\beta}, \ \overline{s}^2[I + X^*\overline{V}X^{*T}], \overline{v}) \tag{A17}$$

where X^* are new observation used for prediction, and y^* are the predicted values.

Independent priors

Apart from the natural conjugate prior, and the non-informative prior described above, another option for the form of prior used in BLR is the independent prior. Under this prior the parameters β and τ are independent, *i.e.* p(β, τ) = p(β) p(τ), which compared to the original relationship between the two parameters removes the conditional on β (*n.b.* the variance of β was listed as $\tau^{-1}\underline{V}$, again showing the dependence of β on τ). By removing this conditional, the independent priors can be expresses as: $\beta \sim$ N($\underline{\beta}$, $\underline{V}$) and $\tau \sim$ $\Gamma(\underline{s}^{-2}, \underline{v})$. Unlike the first case using the natural conjugate priors, multiplying these priors with the likelihood function results in an expression for (β, $\tau|y$) that is no longer NΓ, or in the form of a known density function. Thus, an analytical solution for the joint distribution for both posterior parameters and the predictive distribution for y^* is not possible. However, a numerical solution using a Monte Carlo integration method along with the conditional posterior distributions of the parameters is possible. The conditional posterior distribution for β and τ can be calculated as:

$$(\beta \mid \tau, \ y) \sim N(\overline{\beta}, \ \overline{V}) \tag{A18}$$

$$(\tau \mid \beta, \ y) \sim \Gamma(\overline{s}^{-2}, \ \overline{v}) \tag{A19}$$

where the hyper-parameters can be calculated as follows:

$$\overline{V} = (\underline{V}^{-1} + \tau X^T X)^{-1} \tag{A20}$$

2

$$\overline{\beta} = \overline{V}(\underline{V}^{-1}\underline{\beta} + \tau X^T y) \tag{A21}$$

$$\overline{v} = \underline{v} + N \tag{A22}$$

$$\overline{s}^{-2} = \frac{(y - X\beta)^T(y - X\beta) + \underline{v}\underline{s}^{-2}}{\overline{v}} \tag{A23}$$

A Gibb's sampling algorithm can be used to sample from the posterior conditionals to find the joint posterior distribution ($\boldsymbol{\beta}$, $\boldsymbol{\tau}$) numerically (adapted from Greenberg, 2008). First a random value of $\boldsymbol{\tau}$ is selected; this value is used to calculate one realisation of $\overline{V}$ then $\overline{\beta}$. A value of $\boldsymbol{\beta}$ is then sampled from $(\boldsymbol{\beta}\mid\boldsymbol{\tau},\boldsymbol{y}) \sim \mathrm{N}(\overline{\beta},\overline{V})$, and this value is used to sample $(\boldsymbol{\tau}\mid\boldsymbol{\beta},\boldsymbol{y}) \sim \Gamma(\overline{s}^{-2},\overline{v})$. This process is repeated until convergence and typically a number of samples considered to be "burn-in" are discarded (selected as 10%, or 100 samples for this research). Each pair of sampled values of $\boldsymbol{\beta}$ and $\boldsymbol{\tau}$ represents an approximation of the joint posterior. To predictive probability of $\boldsymbol{y}$ can be estimate numerically; for each pair of $\boldsymbol{\beta}$ and $\boldsymbol{\tau}$, a sample for $\boldsymbol{y}$* can be drawn from the predictive distribution:

$$(y*) \sim N(X*\beta, \frac{1}{\tau}) \tag{A24}$$

where $\boldsymbol{X}$* are new observation used for prediction and $\boldsymbol{y}$* are the predicted values. The posterior conditional distributions from the Gibb's sampling routine can then be used as the priors for any subsequent models.

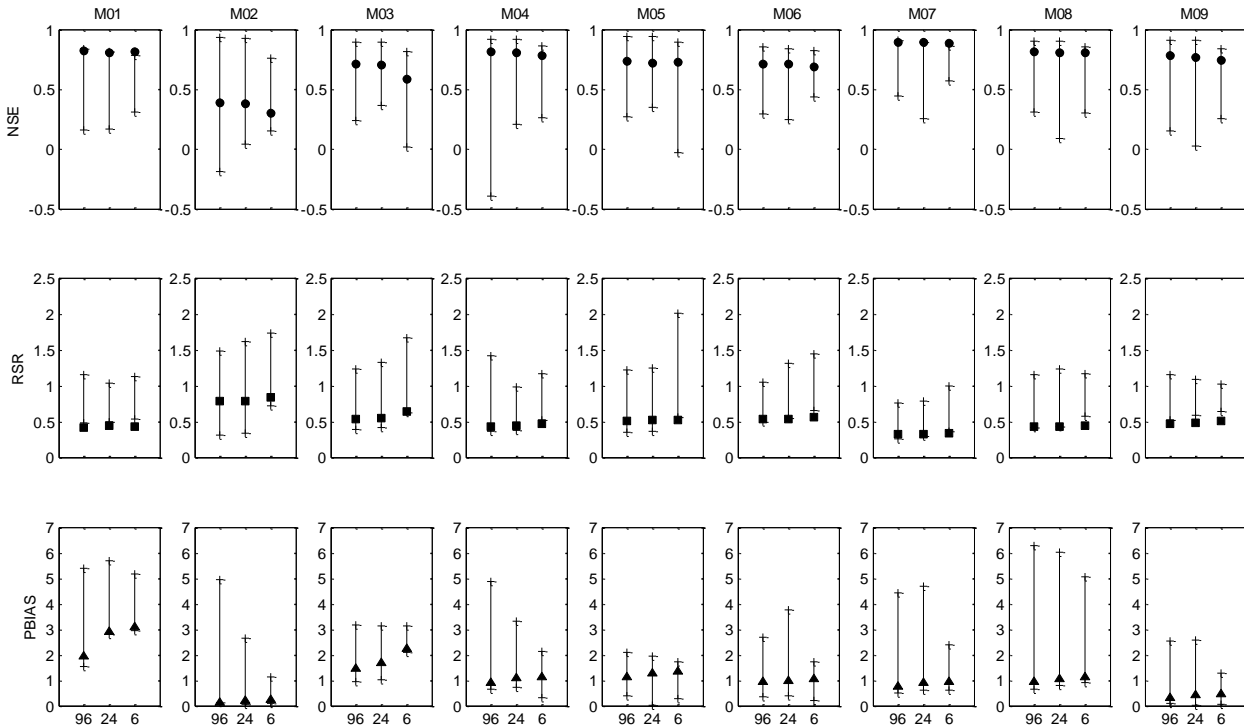## Summary of performance metrics for all models

Figure A1. NSE, RSR, and PBIAS results for M01 to M09 using BLR and FLR models for the 1 day lag at all resolutions for the validation dataset. The black markers are the results for the mean Bayesian predictions, whereas the solid black lines are the upper and lower limits of the fuzzy prediction.
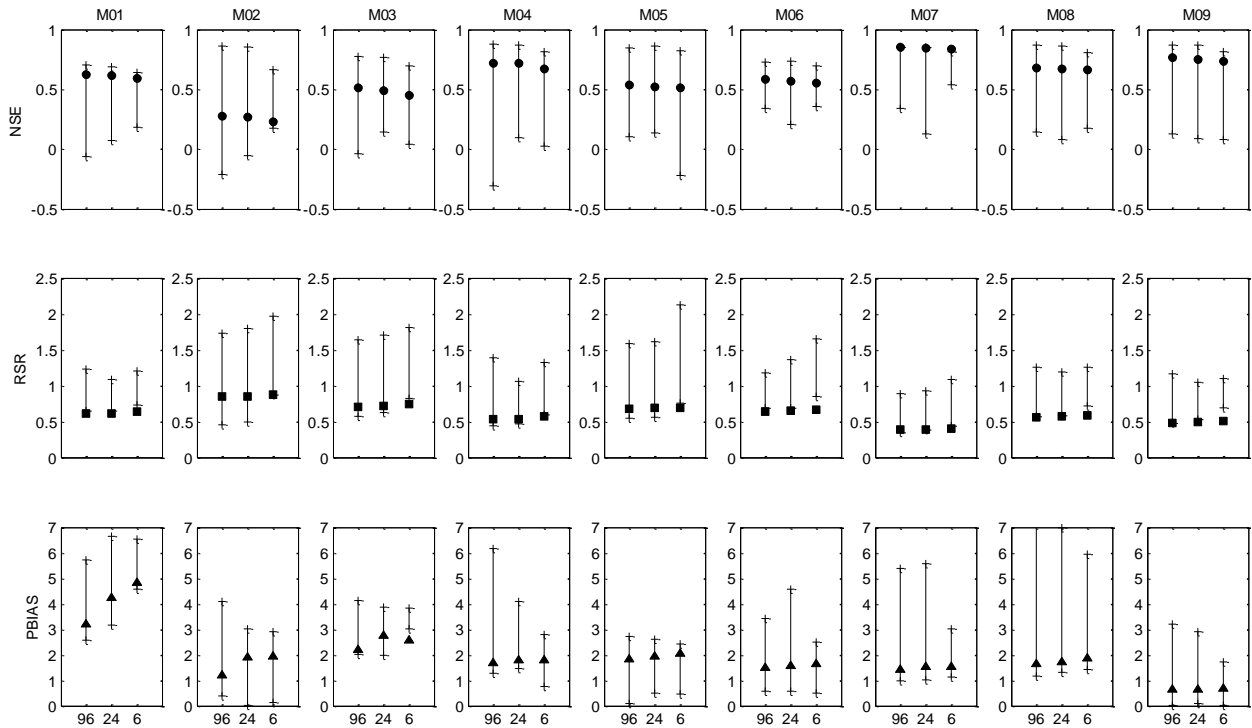
Figure A2. NSE, RSR, and PBIAS results for M01 to M09 using BLR and FLR models for the 2 day lag at all resolutions for the validation dataset. The black markers are the results for the mean Bayesian predictions, whereas the solid black lines are the upper and lower limits of the fuzzy prediction.
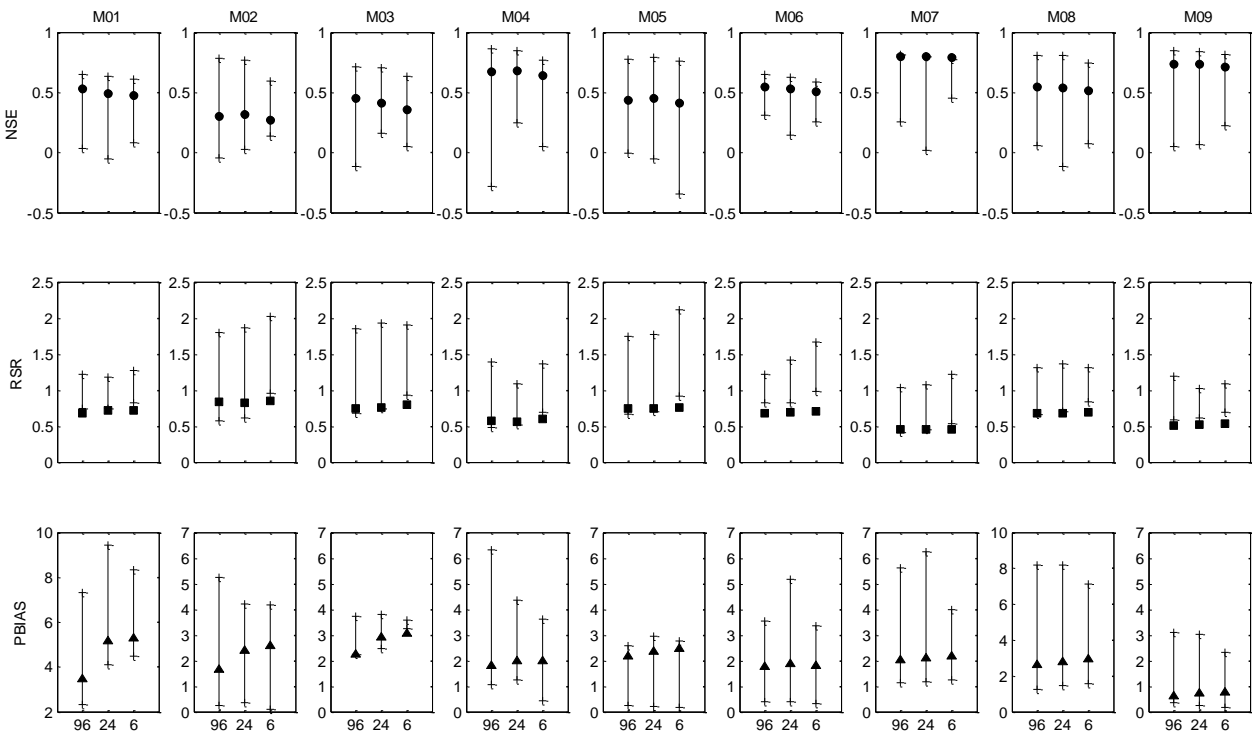

Figure A3. NSE, RSR, and PBIAS results for M01 to M09 using BLR and FLR models for the 3 day lag at all resolutions for the validation dataset. The black markers are the results for the mean Bayesian predictions, whereas the solid black lines are the upper and lower limits of the fuzzy prediction.
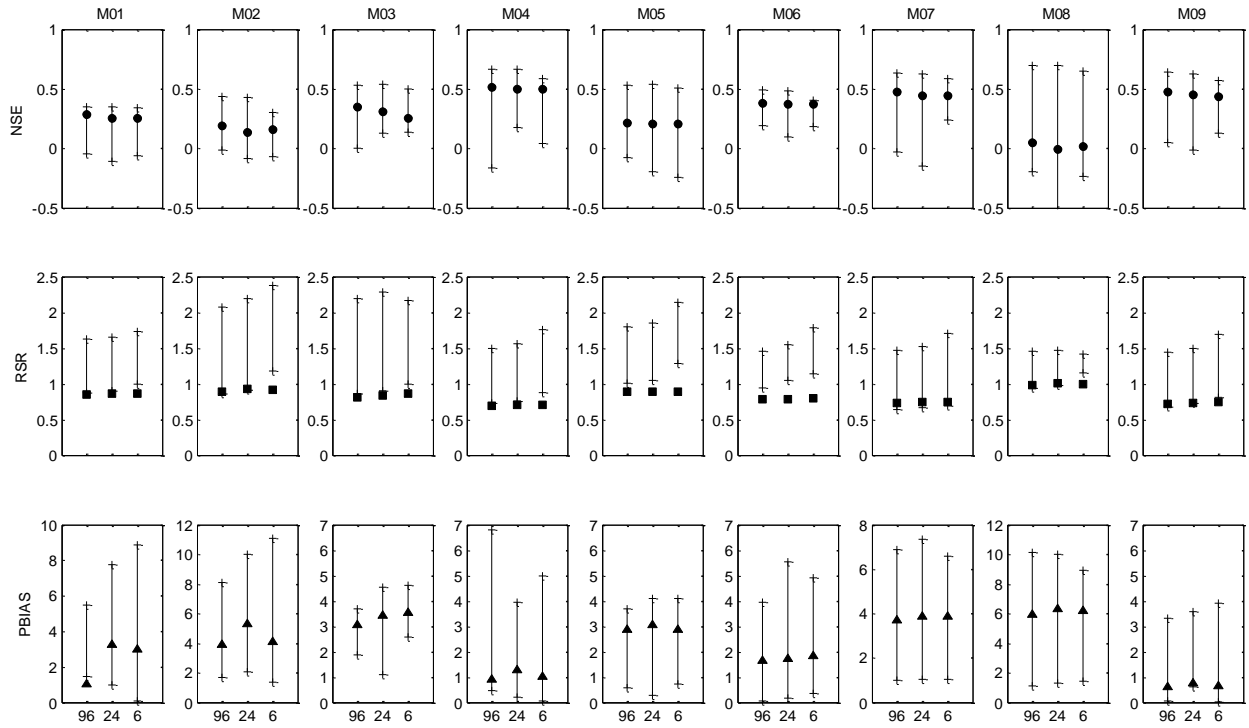
4

Figure A4. NSE, RSR, and PBIAS results for M01 to M09 using BLR and FLR models for the 7 day lag at all resolutions for the validation dataset. The black markers are the results for the mean Bayesian predictions, whereas the solid black lines are the upper and lower limits of the fuzzy prediction.