```
> plot(exp.btgpllm, main = "treed GP, LLM")
```

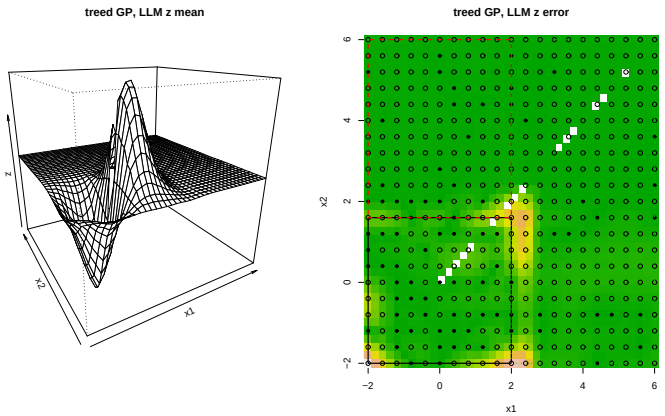


Figure 11: *Left:* posterior predictive mean using `btgpllm` on synthetic exponential data; *right* image plot of posterior predictive variance with data locations `X` (dots) and predictive locations `XX` (circles).

3.4 Motorcycle Accident Data

The Motorcycle Accident Dataset [21] is a classic nonstationary data set used in recent literature [18] to demonstrate the success of nonstationary models. The data consists of measurements of acceleration of the head of a motorcycle rider as a function of time in the first moments after an impact. In addition to being nonstationary, the data has input-dependent noise, which makes it useful for illustrating how the treed GP model handles this nuance. There are at least two, and perhaps three regions where the response exhibits different behavior both in terms of the correlation structure and noise level.

The data is included as part of the `MASS` library in `R`.

```
> library(MASS)
```

Figure 12 shows how a stationary GP is able to capture the nonlinearity in the response but fails to capture the input dependent noise, and increased smoothness (perhaps linearity) in parts of the input space.

```
> moto.bgp <- bgp(X = mcycle[, 1], Z = mcycle[, 2], m0r1 = TRUE)
```

Since the responses in this data have a wide range, it helps to translate and rescale them so that they have a mean of zero, and a range of one. The `m0r1` argument to `b*` and `tgp` functions automates this procedure. All progress indicators are surpressed for this example.

A Bayesian Linear CART model is able to capture the input dependent noise but fails to capture the waviness of the "whiplash"—center— segment of th the response.

```
> moto.btlm <- btlm(X = mcycle[, 1], Z = mcycle[, 2], m0r1 = TRUE)
```

```
> plot(moto.bgp, main = "GP,")
```

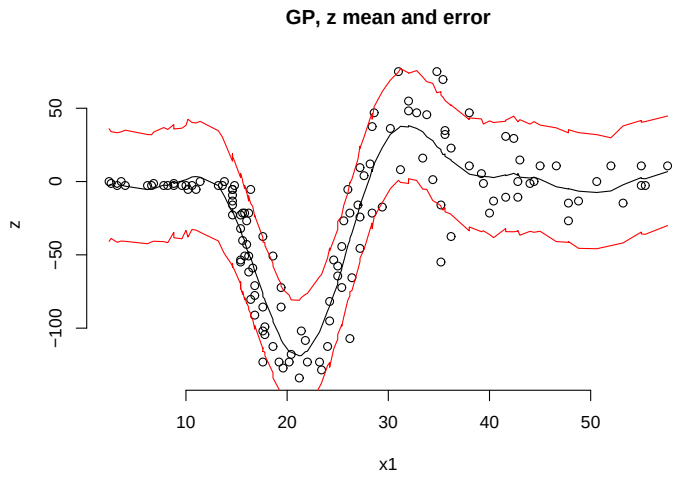


Figure 12: Posterior predictive distribution using `bgp` on the motorcycle accident data: mean and 90% credible interval

```
> plot(moto.btlm, main = "Bayesian CART,")
```

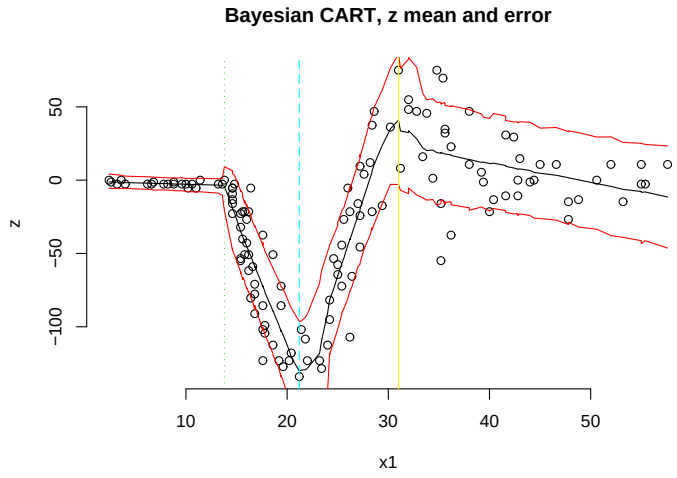


Figure 13: Posterior predictive distribution using `btlm` on the motorcycle accident data: mean and 90% credible interval

predictive surface and MAP partition ($\hat{T}$).

A treed GP model seems appropriate because it can model input dependent smoothness *and* noise. A treed GP LLM is probably most appropriate since the left-hand part of the input space is likely linear. One might further hypothesize

that the right-hand region is also linear, perhaps with the same mean as the left-hand region, only with higher noise. The `b*` and `tgp` functions can force an i.i.d. hierarchical linear model by setting `bprior=b0`. Moreover, instead of rescaling the responses with `m0r1`, one might try encoding a mixture prior for the nugget in order to explicitly model region-specific noise. This requires direct usage of `tgp`.

```
> p <- tgp.default.params(2)
> p$bprior <- "b0"
> p$nug.p <- c(1, 0.1, 10, 0.1)
> moto.tgp <- tgp(X = mcycle[, 1], Z = mcycle[, 2], params = p,
+     BTE = c(2000, 22000, 2))
```

The resulting posterior predictive surface is shown in *top* half of Figure 14. The *bottom* half of the figure shows the norm (difference) in predictive quantiles, clearly illustrating the treed GP's ability to capture input-specific noise in the posterior predictive distribution.

Other permutations of possible models, functions and arguments, for this data is contained in the `b*` and `tgp` examples sections of the respective `R` help files.

3.5 Friedman data

This Friedman data set is the first one of a suite that was used to illustrate MARS (Multivariate Adaptive Regression Splines) [9]. There are 10 covariates in the data ($\mathbf{x} = \{x_1, x_2, \ldots, x_{10}\}$). The function that describes the responses (Z), observed with standard Normal noise, has mean

$$E(Z|\mathbf{x}) = \mu = 10\sin(\pi x_1 x_2) + 20(x_3 - 0.5)^2 + 10x_4 + 5x_5, \qquad (16)$$

but depends only on $\{x_1, \ldots, x_5\}$, thus combining nonlinear, linear, and irrelevant effects. Comparisons are made on this data to results provided for several other models in recent literature. Chipman et al. [4] used this data to compare their linear CART algorithm to four other methods of varying parameterization: linear regression, greedy tree, MARS, and neural networks. The statistic they use for comparison is root mean-square error (RMSE)

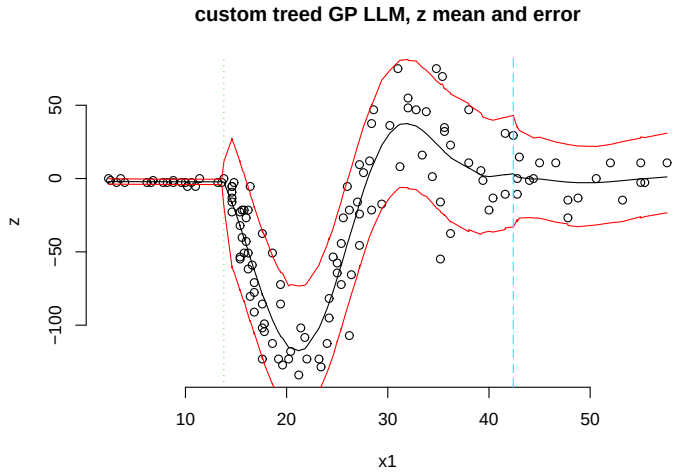
$$\text{MSE} = \textstyle\sum_{i=1}^{n}(\mu_i - \hat{z}_i)^2/n \qquad\qquad \text{RMSE} = \sqrt{\text{MSE}}$$

where $\hat{z}_i$ is the model-predicted response for input $\mathbf{x}_i$. The $\mathbf{x}$'s are randomly distributed on the unit interval.

Input data, responses, and predictive locations of size $N = 200$ and $N' = 1000$, respectively, can be obtained by a function included in the `tgp` package.

```
> f <- friedman.1.data(200)
> ff <- friedman.1.data(1000)
> X <- f[, 1:10]
> Z <- f$Y
> XX <- ff[, 1:10]
```

```
> plot(moto.tgp, main = "custom treed GP LLM,")
```




```
> main <- "quantile difference,"
> plot(moto.tgp$X[, 1], moto.tgp$Zp.q, type = "l", main = main)
```

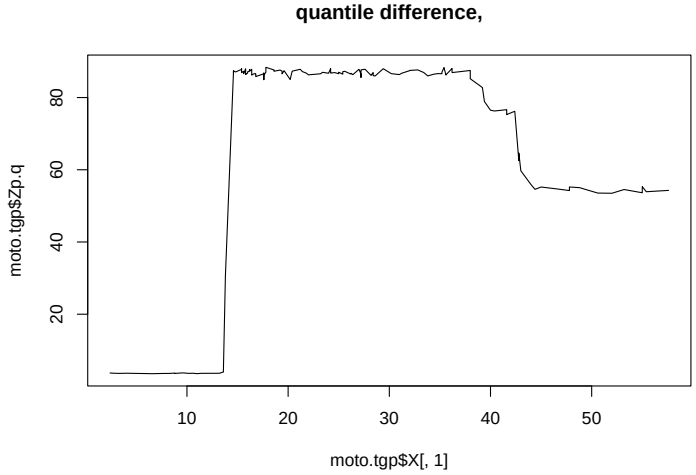


Figure 14: *top* Posterior predictive distribution using a custom parameterized `tgp` call on the motorcycle accident data: mean and 90% credible interval; *bottom* Quantile-norm (90%-5%) showing input-dependent noise.

This example compares Bayesian linear CART with Bayesian GP LLM (not treed), following the RMSE experiments of Chipman et al. It helps to scale the responses so that they have a mean of zero and a range of one. First, fit the Bayesian linear CART model, and obtain the RMSE.