```
> 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. Progress indicators are surpressed.

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

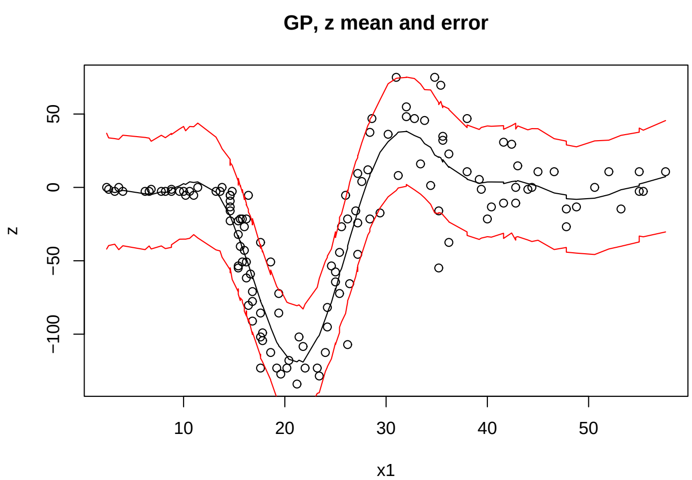


Figure 12: Posterior predictive distribution using `bgp` on the motorcycle accident data: mean and 90% credible interval

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)
```

Figure 13 shows the resulting piecewise linear predictive surface and MAP partition ($\hat{\mathcal{T}}$).

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`.

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

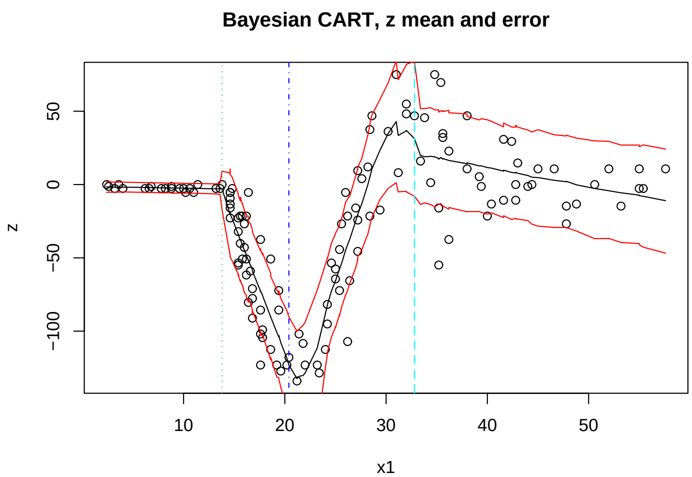


Figure 13: Posterior predictive distribution using `btlm` on the motorcycle accident data: mean and 90% credible interval

```
> 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 the *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.

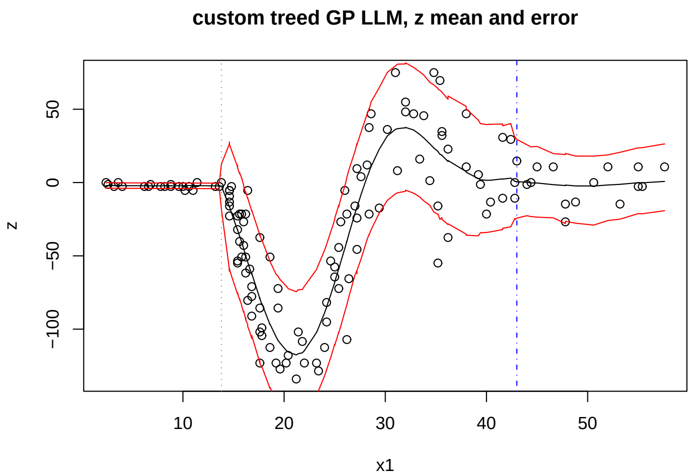
3.5 Friedman data

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

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

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

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




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

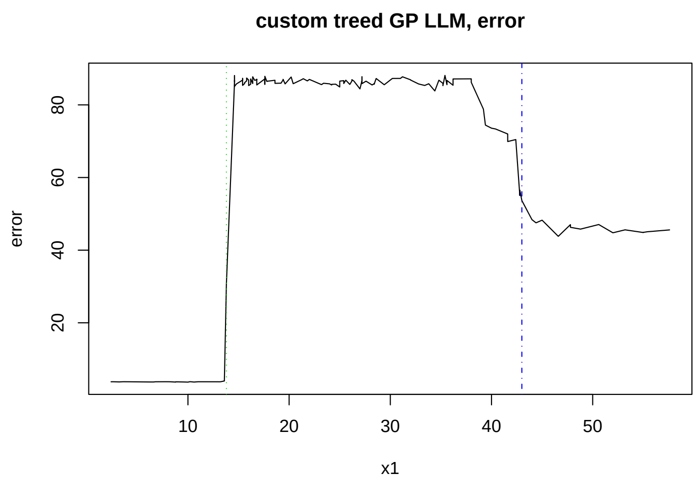


Figure 14: *top* Posterior predictive distribution using a custom parameterized **tgp** call on the motorcycle accident data: mean and 90% credible interval; *bottom* Quantile-norm (90%-5%) showing input-dependent noise.

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.