Laboratory 4

Parametric system identification of a bike sharing process

System and data description. Bike sharing systems represent the evolution of traditional bike rentals where the whole process from rental and return back has become automatic: the user is able to easily rent a bike from a particular position and return back at another position, see Figure 1.2. The count of rental bikes is highly correlated to the weather conditions

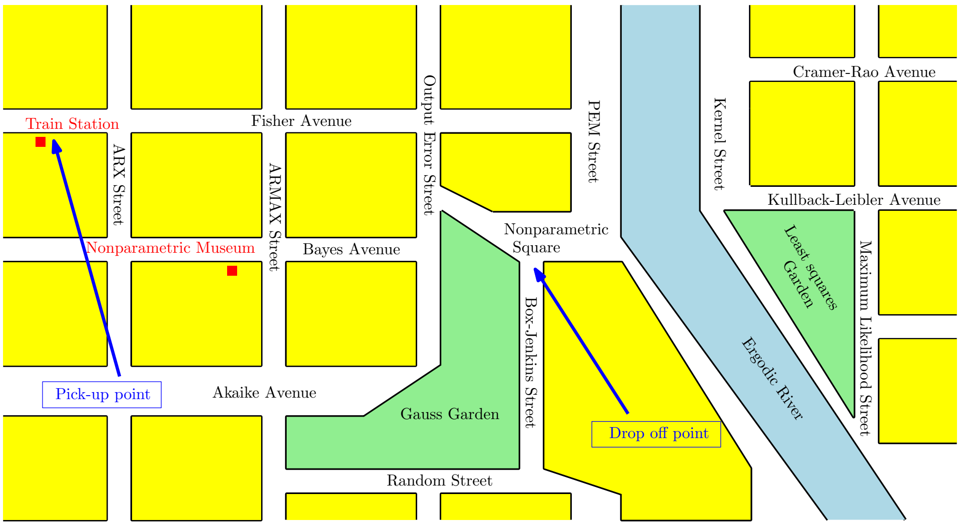

Figure 1.2: Bike sharing system.

and to the hours of the day. Therefore, to design an efficient bike sharing system it is necessary to understand how the count of rental bikes depends on these variables.

Let $y(t)$ and $u(t)$ denote the count rental bikes (output) and the windspeed (input) at time t. Upload the file `bike.mat` which contains the historical log corresponding to the period 13 April 2012 - 28 May 2012 from *Capital Bikeshare system*, Washington D.C., USA, see [1] for more details:

$$u^N = [\, u(1) \dots u(N) \,]^T$$
$$y^N = [\, y(1) \dots y(N) \,]^T$$

with $N = 1104$. These data are aggregated hourly. Missing values (i.e. variables not measured at a certain hour) have been replaced with 0. The unit measure of the input (windspeed) is Km/h.

Data plotting. Plot the data y^N and u^N.

Data detrending. Detrend the data.

Model structure design. From physics it is very difficult to extrapolate a relation between the windspeed and the count of rental bikes. For this reason, we adopt a black-box modeling approach. More precisely, we adopt the Fisherian viewpoint and consider the following four BJ model structures:

Model structure	Degrees
$\mathcal{M}_1(\theta): \quad \boldsymbol{y}(t) = \frac{B(z)}{F(z)}u(t-1) + \frac{C(z)}{D(z)}\boldsymbol{e}(t)$	$n_B = 1$, $n_C = 1$ $n_D = 1$, $n_F = 1$
$\mathcal{M}_2(\theta): \quad \boldsymbol{y}(t) = \frac{B(z)}{F(z)}u(t-1) + \frac{C(z)}{D(z)}\boldsymbol{e}(t)$	$n_B = 2$, $n_C = 2$ $n_D = 2$, $n_F = 1$
$\mathcal{M}_3(\theta): \quad \boldsymbol{y}(t) = \frac{B(z)}{F(z)}u(t-1) + \frac{C(z)}{D(z)}\boldsymbol{e}(t)$	$n_B = 4$, $n_C = 4$ $n_D = 4$, $n_F = 3$
$\mathcal{M}_4(\theta): \quad \boldsymbol{y}(t) = \frac{B(z)}{F(z)}u(t-1) + \frac{C(z)}{D(z)}\boldsymbol{e}(t)$	$n_B = 6$, $n_C = 6$ $n_D = 6$, $n_F = 5$

Training step. Compute the PEM estimate using $\mathcal{M}_i$, $i = 1, 2, 3, 4$. We denote by $\mathcal{M}_i(\hat{\theta}_{PEM,i})$, $i = 1, 2, 3, 4$, the four estimated models.

Validation step

Zero-Pole cancellation analysis. Plot zeros and poles of

$$\mathcal{F}_{\hat{\theta}_{PEM,i}}(z) = \frac{B_{\hat{\theta}_{PEM,i}}(z)}{F_{\hat{\theta}_{PEM,i}}(z)}$$

for the four models we found.

Criteria with complexity terms. Compute the SURE index (we use the given Matlab function `sure.m`), the AIC index and the BIC index (we use the given Matlab function `bic.m`).

Hold-out Cross-validation. Let $\hat{\theta}_{PEM,i}^{CV}$ be the PEM estimate using $\mathcal{M}_i$, with $i = 1, 2, 3, 4$, using the data:

$$u_T = [\ u(1) \quad \dots \quad u(\tfrac{N}{2})\ ]^T, \ \ y_T = [\ y(1) \quad \dots \quad y(\tfrac{N}{2})\ ]^T.$$

Compute the fit percent of the h-step ahead prediction and using the data

$$\begin{aligned} u_V &= [\ u(\tfrac{N}{2}+1) \quad \dots \quad u(N)\ ]^T, \\ y_V &= [\ y(\tfrac{N}{2}+1) \quad \dots \quad y(N)\ ]^T, \end{aligned}$$

with $h = 1, 2, 3, 4$.

Question: In view of the results found in the validation step, which is the best model describing the bike sharing process? Motivate your choice.