Funded by
European Union

Grant Agreement 952111
H2020 WIDESPREAD-05-2020 (Twinning)

Standard course – Modelling
Lesson SM1.4 – How to build a data-driven model - exercises

Marta Magnani – Institute of Geoscience and Earth Resources (CNR, Italy)

- Exercise 1

```
R = corrcoef(data,'Rows','complete');
```

	NEE	ER	Pressure	Soil moisture	Soil temp.	Air misture	Air temp.	Irradiance	GPP
NEE	1.00	-0.65	-0.27	-0.32	-0.35	-0.18	-0.34	-0.39	0.93
7. ER	-0.65	1.00	0.06	0.60	0.36	0.19	0.21	0.50	-0.89
Pressure	-0.27	0.06	1.00	0.02	-0.07	0.04	0.32	-0.21	-0.20
Soil moisture	-0.32	0.60	0.02	1.00	-0.02	0.21	0.04	0.20	-0.49
Soil temp.	-0.35	0.36	-0.07	-0.02	1.00	-0.32	0.73	0.55	-0.39
Air moisture	-0.18	0.19	0.04	0.21	-0.32	1.00	-0.54	-0.22	-0.20
Air temp	-0.34	0.21	0.32	0.04	8. 0.73	-0.54	1.00	0.30	-0.31
Irradiance	-0.39	0.50	-0.21	0.20	0.55	-0.22	0.30	1.00	-0.48
7. GPP	0.93	-0.89	-0.20	-0.49	-0.39	-0.20	-0.31	-0.48	1.00

• Exercise 2

ER models

```
clear y x1 X
y=A(:,9);
x1=A(:,15);
X=[ones(size(x1)),x1];

b=regress(y,X);

x2=A(:,31);
X=[X,x2];
b2=regress(y,X);
```

Model:

1. $ER = a_0 + a_1\ soil\ moisture$

$a_0 = 0.23$ mol/m²/day
$a_1 = 0.013$ mol/m²/day

2. $ER = a_0 + a_1\ soil\ moisture + a_2\ irradiance$

$a_0 = -0.07$ mol/m²/day
$a_1 = 0.011$ mol/m²/day
$a_2 = 4.6 * 10^{-4}$ mol/W/day

• Exercise 2

GPP models

```
clear y x1 X
y=A(:,7)-A(:,9);
x1=A(:,15);
X=[ones(size(x1)),x1];

b=regress(y,X);

x2=A(:,31);
X=[X,x2];
b2=regress(y,X);
```

Model:

1. GPP $= a_0 + a_1\ soil\ moisture$

$a_0 = -0.57$ mol/m²/day
$a_1 = -\ 0.018$ mol/m²/day

2. GPP $= a_0 + a_1\ soil\ moisture + a_2\ irradiance$

$a_0 = 0.14$ mol/m²/day
$a_1 = -\ 0.015$ mol/m²/day
$a_2 = -0.0011$ mol/W/day

- Exercise 3

ER model

```
clear y x1 b model beta0

y=A(:,9);
x1=A(:,27);

model = @(b, x) (b(1) * exp (b(2) * x));
beta0=[1;1];

[b, R, J, covb, mse] = nlinfit (x1, y, model, beta0);

N=length(y);
k=length(b)+1;
AIC=N*log(sum(R.^2)/N)+2*k;
```

	Estimate
b1	0.38
b2	0.03

AIC = -149.66

• Exercise 3

GPP model

```
clear y x1 b model beta0

y=A(:,7)-A(:,9);
x1=A(:,31);

model=@(b, x) (b(1) * b(2) * x/( b(1) + b(2) * x));
beta0=[-3;-0.003];

[b, R, J, covb, mse] = nlinfit (x1, y, model, beta0);

N=length(y);
k=length(b)+1;
AIC=N*log(sum(R.^2)/N)+2*k;
```

	Estimate
b1	-3.3
b2	-0.0022

AIC = -61.74

• Exercise 4

ER model

```
clear y x1 x2 b model beta0

y=A(:,9);
x1=A(:,27);
x2=A(:,15);

model = @(b, x) (exp (b(1) * x(:,1))*(b(2)+b(3)*x(:,2)));
beta0=[1;1;1];

[b, R, J, covb, mse] = nlinfit ([x1,x2], y, model, beta0);

N=length(y);
k=length(b)+1;
AIC=N*log(sum(R.^2)/N)+2*k;
```

	Estimate
b1	0.019
b2	0.18
b3	0.0098

AIC = -174.70

The AIC is lower than the one obtained for the univariate regression. Hence, the multi regression model is better.

• Exercise 4

GPP model

```
clear y x1 x2 b model beta0

y=A(:,7)-A(:,9);
x1=A(:,31);
x2=A(:,15);

model=@(b, x) ((b(1) * b(2) * x(:,1)/( b(1) + b(2) * x(:,1)))*(b(3)+b(4)*x(:,2)));
beta0=[-1;-0.0002;2;0.05];

[b, R, J, covb, mse] = nlinfit ([x1,x2], y, model, beta0);

N=length(y);
k=length(b)+1;
AIC=N*log(sum(R.^2)/N)+2*k;
```

	Estimate
b1	-0.08
b2	-0.0003
b3	3.8
b4	0.07

AIC = -81.34

The AIC is lower than the one obtained for the univariate regression. Hence, the multi regression model is better.