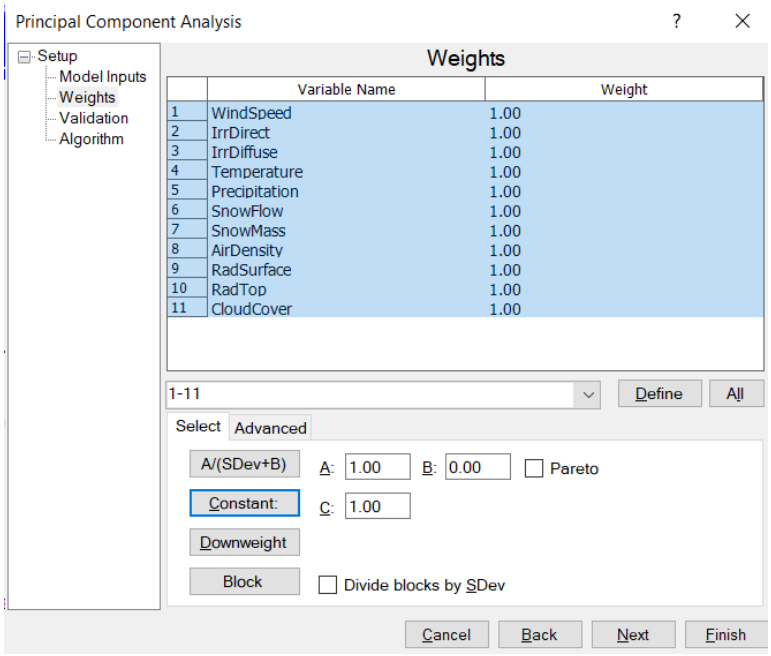
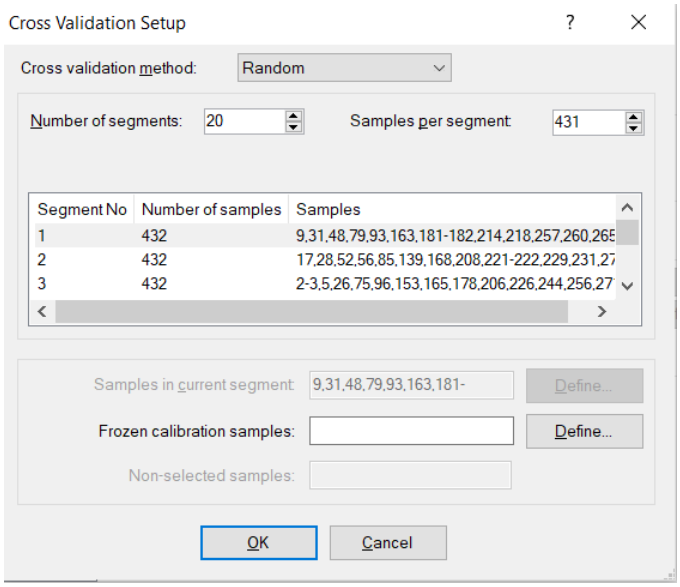
Open the file WindSolarPower.unsb

- Make a PCA model on the samples set “2016” with the column set “X”
- Tasks-Analyze-Principal Component Analysis

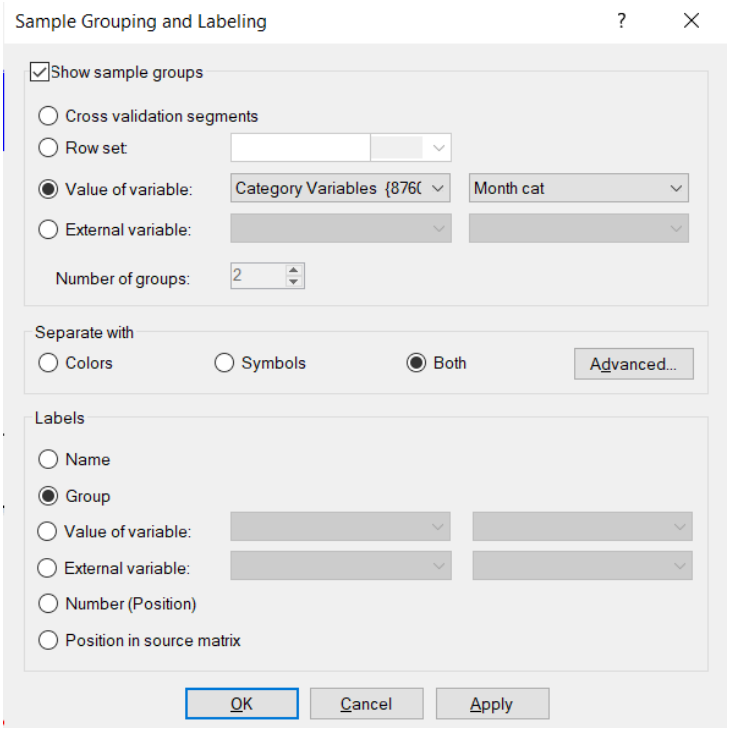


- First use weights = 1
- Select random cross validation



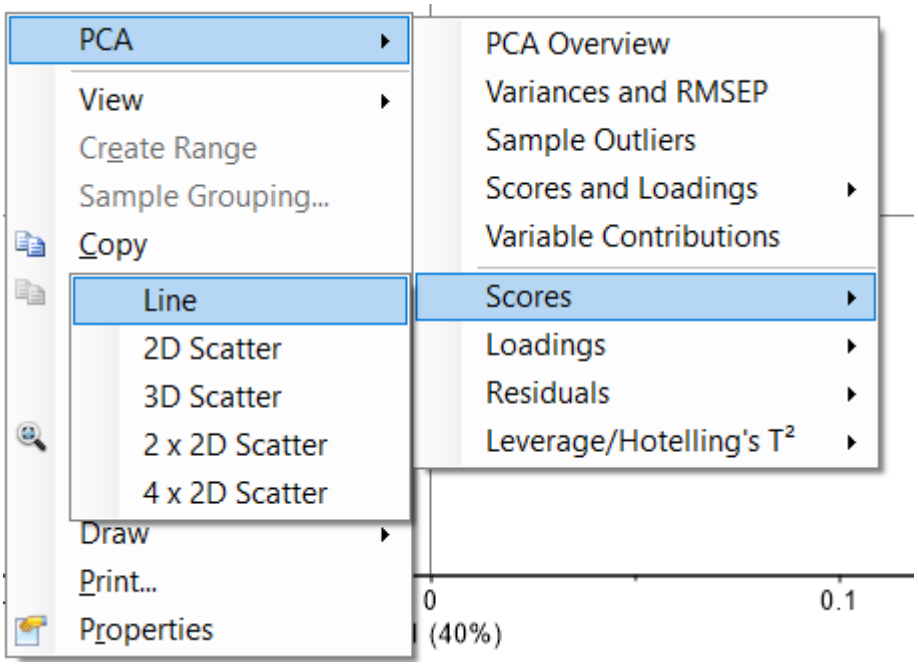
- Interpret scores loadings and explained variance

Right click and select Sample grouping to group e.g. on month



- Click on the button  to show correlation loadings (first click in the loading plot to make it active). Explain why the plot has changed. You may add the Hotelling's T^2 ellipse.
- Now recalculate with setting Weights to 1/Std.Dev
- Interpret the model again. Decide on the optimal number of PCs

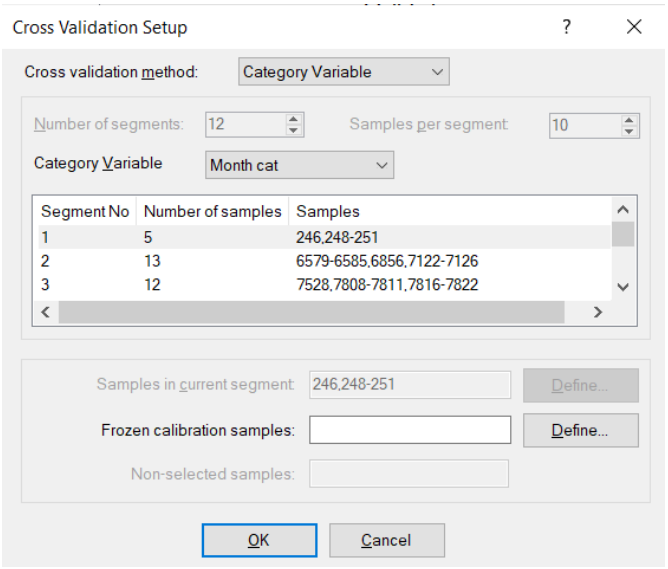
Hint: For time series data it is a good idea to look at scores as line plots



Zoom in to show only scores for some days (use the Frame scale () . Do you see the daily systematic pattern?

Try other cross validation set-ups:

- Systematic (111 222 333)
- Category variables (day/night, month)



Next steps:

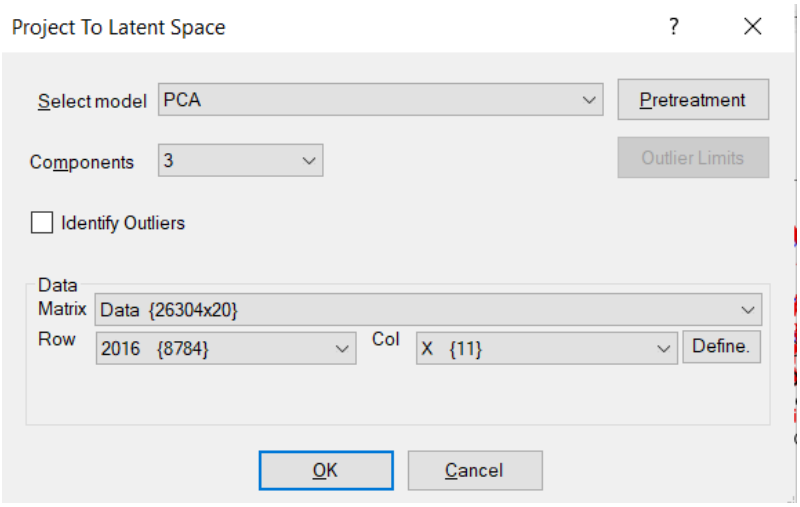
- Compare the explained validation variance for these models
- Decide on the optimal number of components (compare to the model with random CV above)
- Look into the Hoteling's T^2 and F-residual plots if there are any outliers (NB! Decide on the optimal number of PCs first)

Discuss if it is conceptually viable to include year, month and/or day/night if the purpose is to project new samples onto a training model for detecting changes in the x-variables.

- You may make a model on the column set "X incl day_night year and month" while discussing this.

Project the row set "2017" onto the model you selected as the "best"

Tasks – Predict - Projection



See if the projected samples are within the critical limits from the 2016 data.

Now make a model on all data and cross validate over year.