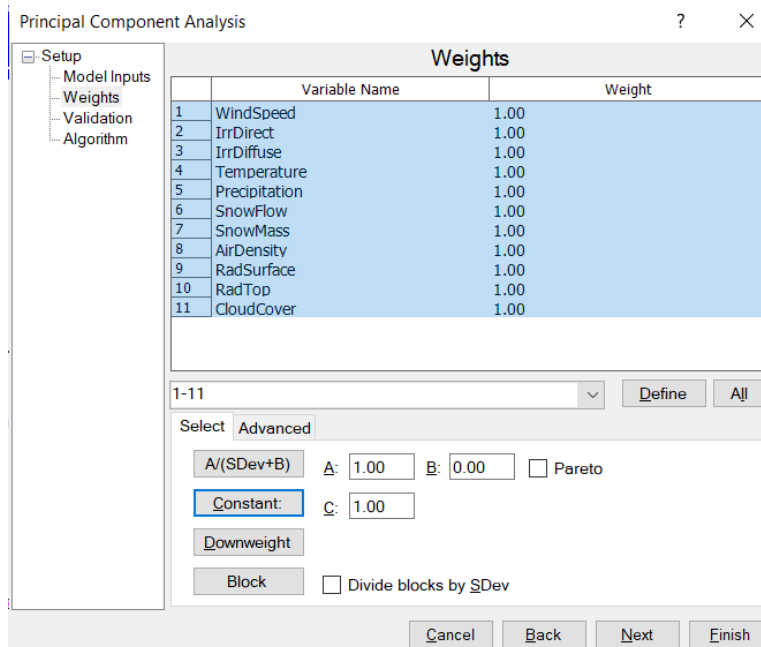
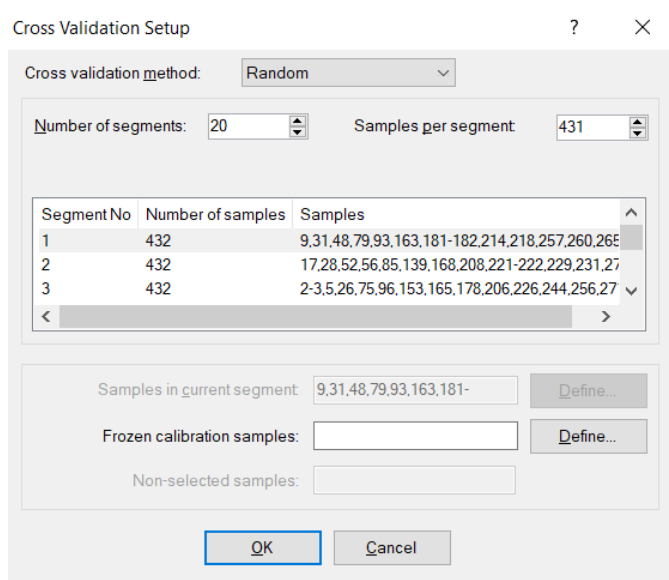


Open the file WindSolarPower.unsb

- 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

Sample Grouping and Labeling

☒ Show sample groups

☐ Cross validation segments
☐ Row set
☒ Value of variable: Category Variables {8760} Month cat
☐ External variable:

Number of groups: 2

Separate with


☐ Colors
☐ Symbols
☒ Both

Advanced...

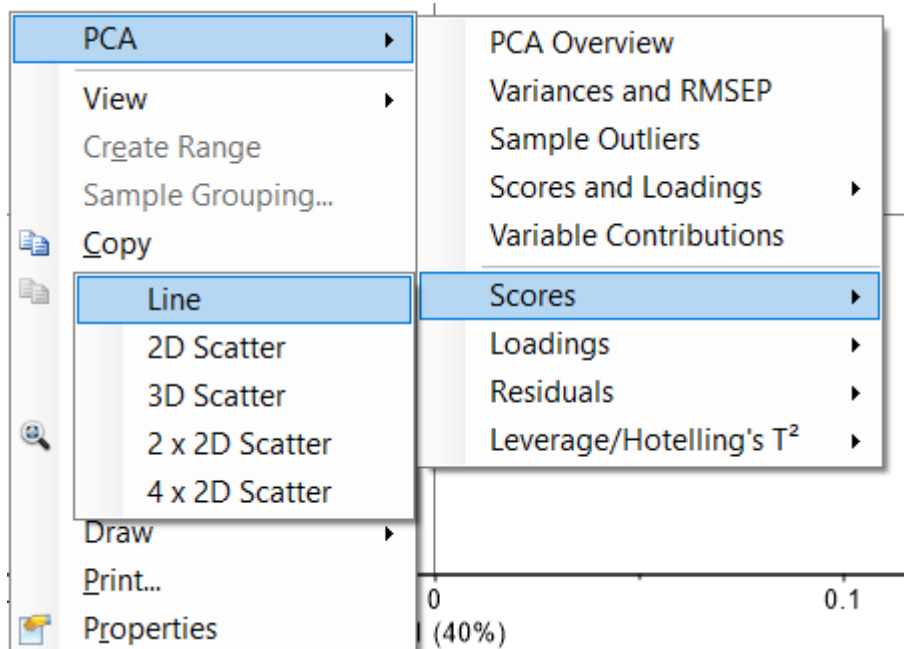
Labels


☐ Name
☒ Group
☐ Value of variable:
☐ External variable:
☐ Number (Position)
☐ Position in source matrix

OK Cancel Apply

- 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)

Cross Validation Setup

Cross validation method: Category Variable

Number of segments: 12 Samples per segment: 10

Category Variable: Month cat

Segment No	Number of samples	Samples
1	5	246,248-251
2	13	6579-6585,6856,7122-7126
3	12	7528,7808-7811,7816-7822

Samples in current segment: 246,248-251 Define...

Frozen calibration samples: Define...

Non-selected samples:

OK Cancel

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 Hotelling'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

Project To Latent Space

Select model: PCA Pretreatment

Components: 3 Outlier Limits

☐ Identify Outliers

Data Matrix: Data {26304x20}

Row: 2016 {8784} Col: X {11} Define.

OK Cancel

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.