**EMSC tutorial companion**

## Licence

The extended multiplicative signal correction (EMSC) and Saisir MATLAB code are available free of charge. For EMSC there was a patent limiting commercial use until October 2013. The code is now distributed under GPL-2 licencing. Functions, codes, data sets and documents are presented as is, and are to be used at the user’s own risk.

## Description

This document accompanies the EMSC code used in the paper “Extended multiplicative signal correction in vibrational spectroscopy, a tutorial”. It is intended to explain briefly how to apply the code to spectroscopic data in general and to recreate the plots in the paper.

## Contents

[Licence 1](#_Toc361129252)

[Description 1](#_Toc361129253)

[Contents 1](#_Toc361129254)

[Saisir 2](#_Toc361129255)

[(0) Preparations 2](#_Toc361129256)

[(1) MSC 2](#_Toc361129257)

[(2) Basic EMSC model 2](#_Toc361129258)

[(3) Higher order polynomial 2](#_Toc361129259)

[(4) Reference spectrum (baseline) 2](#_Toc361129260)

[(5) Constituent spectra 2](#_Toc361129261)

[(6) Replicate correction 2](#_Toc361129262)

## Saisir

Saisir is a data structure having three elements:

* d: A data matrix, typically containing spectroscopic data
* v: A character matrix containing variable names, e.g. wavenumbers/-lengths on a format that is directly available to str2num().
* i: A character matrix containing fixed length samples names, including replicate information.

A large range of functions accompany the Saisir structure, covering many aspects from data import, visualisation, processing and analyses of different sorts. This is described in a separate document.

## Data sets

### Raman

These are raw spectra from Raman spectroscopy of salmon oils. The data set consists of 3 repeated measurements of in total 45 salmon oil samples.

### FT-IR

These are FT-IR microscopy spectra of meat tissue sections recorded under varying amounts of water vapour in the air.

## Code description

### (0) Preparations

Before application of EMSC one has to import and format ones data to fit the Saisir structure. Saisir functions depend on variable names that can be translated using str2num() and sample names of fixed length. Therefore the variable names of the supplied Raman data need to be reformatted to remove extraneous signs, and the sample names need to be adapted.

It is always advisable to use plotting as a visual confirmation of successful data handling. After import and reformatting we therefore plot the raw Raman spectra.

The supplied FT-IR data consists of several data sets that all follow the naming conventions needed by Saisir.

### (1) MSC

A simple multiplicative signal correction is performed on the Raman data. Corrected spectra and residuals from the correction are plotted together with the raw spectra for comparison.

### (2) Basic EMSC model

A basic EMSC model with linear and quadratic terms is created and applied to the Raman spectra.

### (3) Higher order polynomial

A fourth order EMSC model with linear, quadratic, cubic and quartic terms is created and applied to the Raman spectra.

### (4) Reference spectrum (baseline)

A basic EMSC model with linear and quadratic terms is created. Then the mean spectrum is replaced by one of the raw spectra as the reference spectrum. The model is applied to the data and plots are made.

### (5) Constituent spectra

Weights are constructed to minimize the effect of the area from 2450 cm-1 to 2200 cm-1. EMSC models are made for the complete FT-IR data set (all conditions) and the reduced dataset with only large differences in water vapour levels. The complete data is corrected separately with each model (A and B, respectively), including the weights.

Data set A is divided into two spectra with high levels of O-H stretching and low levels of O-H stretching. Difference spectra between the first spectrum of data set B and each of the other spectra are calculated, and the two first (non-centred) principal component loadings are estimated. These two loadings are added as constituent spectra to the EMSC model used on data set A, and data set A is reprocessed with the extended model (C).

A difference spectrum between two of the spectra in data set B having large water vapour differences is also calculated. The EMSC model is further extended with this spectrum and the reprocessed data set C is again reprocessed with the final EMSC model (D).

Plots of data set A and the final processed spectra are made (D), in addition to the first two principal components of data set A and D are made, together with the single difference spectrum.

### (6) Replicate correction

A basic EMSC model is made. This is extended with a replicate correction model. Here, principal component loadings are made for each set of replicates. Another principal component analysis is performed on these loadings, and the first few loadings of this model are included in the EMSC model as typical replicate variation. The model is applied to the Raman data and results are plotted.