function [OUT] = F4\_CoreSF(data)

% This function will calculate a number of core sensor features for any

% data sensors time series that is contained in 'data'. Examples of these

% sensors are MILK YIELD, ACTIVITY, BODY WEIGHTS and RUMINATION, and should

% contain of 'daily' measurements (as high-frequent as possible).

% The calculation of these core sensor features will take into account herd

% peers, meaning that first the average 'lactation curve' within herd and

% parity will besubtracted (= relative curve) from the raw data before

% calculating the core sensor features. This means that we are basically

% characterizing the 'deviation from the herd and lactation mean' rather

% than the raw sensor time series. To this end, some of the core features

% are calculated using the relative curve (REL), while some of the features

% are calculated using the residuals (RESREL) from the regression line

% through this relative curve; the exact calculation will further be

% specified below. The output table can be used as predictors for

% regressing against the resilience scores/ranking

%

% INPUTS: data table that contains the following columns:

% - CowID unique cow identification numbers

% - DIM days in milk of each lactation

% - Lac lactation/parity number

% - Sensor daily sensor data, BW/ACT/RUM/MY

%

% OUTPUTS: OUT an output table with the core SF for the sensor

% contains following columns:

% - AVG mean value of the relative curve (REL)

% - AC autocorrelation of the relative curve (REL)

% - MIN minimum value of RESREL

% - MAX maximum value of RESREL

% - STD standard deviation of RESREL

% - SKEW skewness of the RESREL

% - SLOPE slope of the RESREL

%

% The script includes following steps:

% STEP 0: check which variable/sensor is used

% STEP 1: calculate average 'lactation curve' and subtract from data

% (= REL)

% STEP 2: regress the relative curve & calculate residuals (RESREL)

% STEP 3: calculate the core SF

% STEP 4: output the results in OUT table

**%% STEP 0: Check variable/sensor to be calculated (colnames)**

% find the sensor column in order to indentify the sensor variable - change

% colname to general colname

try

VN = find(string(data.Properties.VariableNames) == 'TMY'); % col position of TMY (total milk yield)

catch

try

VN = find(string(data.Properties.VariableNames) == 'ACT'); % col position of ACT if no TMY

catch

try

VN = find(string(data.Properties.VariableNames) == 'RUM'); % col position of RUM

catch

VN = find(string(data.Properties.VariableNames) == 'BW'); % col position of Calving

end

end

end

data.Properties.VariableNames{VN} = 'SENS'; % change variable name of sensor column

**%% STEP 1: Calculate average lactation curve & subtract data (REL)**

% In this step, we will calculate the mean for the sensor for each day in

% lactation; for example for yield, this will be the average lactation

% curve within parity. It can be calculated independent of 'completeness'

% of the lactation, as long as there is a bit of overlap (e.g. several cows

% have lactation in days in milk (=DIM) 20, then you can calculate an

% average value for DIM 20 and correct for it. If your data does not have

% any agreement in lactation stage, you might want to skip this step, but

% then it should be reported.

%

% ASSUMPTION: ONLY take at most 300 days into account

% ASSUMPTION: lactation counting starts from DIM 0 (calving date)

% (important for accumulating arrays (= calculate mean) with

% accumarray)

% explore which parities are present to average per parity

LAC = unique(data.Lac); % all unique parity numbers in the dataset;

% find all data with parity

for i = LAC % all unique lactations- for averageing per parity

Sens\_avg(LAC(i),1:301) = NaN; % prepare average structure; size = unique lactations x 306 (DIM 0-305)

val = data.SENS(data.Lac == LAC(i)); % values to aggregate (calculate average for)

subs = data.DIM(data.Lac == LAC(i))+1; % to calculate the mean indices of all data from lactation i, subs = positive integer (e.g. mean value for all data recorded on DIM 3, 10, ...)

Sens\_avg(LAC(i), subs) = accumarray(subs, val, [], @mean); % calculate mean and store them in Sens\_avg

end

% RESULT: variable 'Sens\_avg': each row contains the average data of a lactation

% each col contains the average data of a DIM

% subtract the average curve from the raw sensor data

data.REL(:,1) = NaN; % prepare column to fill in the relative data

for i = LAC % unique lactations

for j = 0:305 % unique DIM

ind = find(data.DIM == j & data.Lac == i); % find all data with DIM = i and lactation number = j

if isempty(ind) == 0 % if there is data for that DIM and that parity (not empty)

data.REL(ind,1) = data.SENS(ind) - Sens\_avg(i,j); % subtract average value for that DIM and that parity

end

end

end

% RESULT: new column in the dataset containing the relative value compared

% to the peers (DIM/parity) of that same herd

**%% STEP 2: Regress the relative curve and calculate residuals (RESREL)**

% In this step, we calculate the regression line through the relative curve

% for the sensor data

% identify the unique cow lactations in the dataset

cowlac = unique([data.CowID data.Lac]); % contains identifiers of the unique cows

% RESULT: cowlac = [cow1 lac1

% cow1 lac2

% cow2 lac1]; etc.

% add parameter to cowlac to store slope of relative curve

cowlac(:,3) = NaN; % to fill in the slope of the regression of the relative curve

data.REGRES(:,1) = NaN; % to fill in the regression of the relative curve

data.RELRES(:,1) = NaN; % to fill in the residuals from the regression of the relative curve

for i = 1:length(cowlac(:,1)) % all unique cow lactations

ind = find(data.CowID == cowlac(i,1) & data.Lac == cowlac(i,2)); % find each unique lactation

X = data.DIM(ind); % select DIM for that cow lactation

Y = data.REL(ind); % select relative (peer-corrected) sensor data for that cow lactation

b = regress(Y,X); % obtain the regression coefficients on the residual curves for this

sensor/cow/lactation

% if you have very little points,

% consider using a robust fit

data.REGRES(ind,1) = X\*b; % regression line for this cow lactation

data.RELRES(ind,1) = Y-X\*b; % residuals from the regression relative curve

cowlac(i,3) = b(2); % fill in slope

end

% RESULT: data table with additional columns containing the regression line

% through the data and the residuals against this regression line

**%% STEP 3: Calculate the core SF**

% In this step, we will use REGRES and RELRES to calculate the core sensor

% features of the sensor data for each lactation

% prepare OUT = output dataset with the core features calculated

OUT.CowID(:,1) = cowlac(:,1); % cow ID

OUT.Lac(:,1) = cowlac(:,2); % lactation ID

OUT.AVG(:,1) = NaN; % mean value of the relative curve (REL)

OUT.AC(:,1) = NaN; % autocorrelation of the relative curve (REL)

OUT.MIN(:,1) = NaN; % minimum value of RESREL

OUT.MAX(:,1) = NaN; % maximum value of RESREL

OUT.STD(:,1) = NaN; % standard deviation of RESREL

OUT.SKEW(:,1) = NaN; % skewness of the RESREL

OUT.SLOPE(:,1) = cowlac(:,3); % slope of the REL as stored before (in b)

numLags = 1; % number of lags for which the autocorrelation is calculated (i.e. 1)

for i = 1:length(cowlac(:,1))

ind = find(data.CowID == cowlac(i,1) & data.Lac == cowlac(i,2) & isnan(data.REL(ind)) == 0); % find each

unique lactation

OUT.AVG(:,1) = nanmean(data.REL(ind)); % mean value of the relative curve (REL)

OUT.AC(:,1) = autocorr(data.REL(ind),numLags); % autocorrelation of the relative curve (REL)

OUT.MIN(:,1) = min(data.RESREL(ind)); % minimum value of RESREL

OUT.MAX(:,1) = max(data.RESREL(ind)); % maximum value of RESREL

OUT.STD(:,1) = std(data.RESREL(ind)); % standard deviation of RESREL

OUT.SKEW(:,1) = skewness(data.RESREL(ind),1); % skewness of the RESREL (correct for bias)

end

% RESULT = OUTPUT TABLE with for each cow/lactation the different core

% sensor parameters calculated.

**%% STEP 4: Output table**

% Be aware: an additional outlier detection step or normalization or

% standardisation step might be needed before regression these agains the

% resilience ranking or scores!

% To work in with only first lactation cows you will need to select 1st

% lactations using the LAC column