Main program

```
clear;clc;
data_CMA = load("D:\CSRDD_code\Data_CMA.mat");
% 1 台站号、2 纬度、3 经度、4 海拔、5 年、6 月、7 日、8 日平均气温、9 日最高气温、10
日最低气温、11 日照时数、12 大气压强、13 天文辐射、14GHI、15DHI
%% 气象数据质量控制
% 质量控制
station = unique(data_CMA(:,1));

overview_m = [];                        % 气象数据概况
data_m = [];                            % 经过质量控制的气象数据
for i = 1:length(station)
    i
    i_station = station(i);
    idx = find(data_CMA(:,1)==i_station);
    i_data = data_CMA(idx,:);
    num_days = length(idx);                          % 总数据量（天数）
    % 缺测、漏测数据以 37644 或 32766 表示
    idx_t1 = find(i_data(:,8)>30000);                %平均气温
    idx_t2 = find(i_data(:,9)>30000);                %最高气温
    idx_t3 = find(i_data(:,10)>30000);                %最低气温
    idx_s = find(i_data(:,11)>30000);                %日照时数
    idx_p = find(i_data(:,12)>30000);                %大气压强
    % 获得 overview_m
    idx_abn_all = unique([idx_t1;idx_t2;idx_t3;idx_s;idx_p;]); % 异常数据所在
行
    num_abn = length(idx_abn_all);   %异常梳理 abnormal
    new = [i_station,num_days,num_abn,num_abn/num_days];  % 台站号、总数据
量、异常数据量、异常比例
    overview_m = [overview_m;new];
    % 获得 data_m
    i_data(idx_abn_all) = [];
    data_m = [data_m;i_data];
end

%% 国家气象局 CMA 的 GHI 数据质量控制
station_GHI = [50136   50468   50527   50742   50834   50873   50953   51076
    51133   51431   51463   51567   51573   51628   51644   51709   51777
    51828   52203   52267   52418   52533   52681   52754   52818   52866
    52983   53068   53336   53464   53487   53543   53545   53614   53772
    53817   53845   53923   53963   54102   54135   54161   54292   54324
```

```
    54342   54511   54527   54539   54662   54764   54765   54823   54915
    54936   55228   55299   55591   56029   56043   56137   56146   56173
    56187   56196   56385   56386   56651   56666   56691   56739   56778
    56959   56985   57083   57131   57178   57186   57245   57411   57432
    57461   57494   57516   57604   57649   57687   57816   57874   57957
    57993   58141   58208   58238   58265   58321   58362   58367   58457
    58467   58531   58606   58665   58737   58847   59082   59287   59316
    59431   59485   59644   59758   59948   59981
];
overview_GHI = [];   % 国家气象局 CMA 的 GHI 数据概况
data_GHI = [];       % 质量控制后的 CMA 的 GHI 数据
for i = 1:length(station_GHI)
    i
    i_station = station_GHI(i);
    idx = find(data_qxfs(:,1)==i_station);
    i_data = data_qxfs(idx,:);
    num_days = length(idx);                          % 总数据量
    % 异常判别标准：1）kt>1；2）kt<0.015
    idx_g1 = find(i_data(:,14)<30000 &
i_data(:,14)*0.01./i_data(:,13)>1);      % 保证有观测值情况下，kt>1
    idx_g2 = find(i_data(:,14)<30000 &
i_data(:,14)*0.01./i_data(:,13)<0.015);   % kt<0.015
    idx_abn_all = unique([idx_g1;idx_g2])
    num_abn = length(idx_abn_all);   % 异常梳理 abnormal
    % 获得 overview_GHI
    new = [i_station,num_days,num_abn,num_abn/num_days];   % 台站号、总数据
量、异常数据量、异常比例
    overview_GHI = [overview_GHI;new];
    % 获得 data_GHI
    i_data(idx_abn_all) = [];
    data_GHI = [data_GHI;i_data];
end

%% 国家气象局 CMA 的 DHI 数据质量控制
station_DHI = [50953   51463   51709   52818   54342   54511   56778   57083
    57494   59287   55591   57816   58362   52267   59948   50136   53545
    54765   53464   51644   56691   57432   58367   58467   59082   59485
    56386   57411   56187   52983   55228   52418   52681   56137   56959
    51076   51431   51573   51777   51828   52203   56029   50468   50873
    52866   53068   53487   53772   53963   54161   54823   56385   56739
    57461   57957   57993   58321   58606   58847   59316   59431   59758
    55299   54527   58238   58457   53614   57516   56196   56985   53817
    57687
];
```

```
overview_DHI = [];
for i = 1:length(station_DHI)
    i
    i_station = station_DHI(i);
    idx = find(data_qxfs(:,1)==i_station);
    i_data = data_qxfs(idx,:);
    i_kt = i_data(:,14)./i_data(:,13)*0.01;      % 晴空指数
    i_kdf = i_data(:,14)./i_data(:,15);          % 散射比
    % 使用子函数（包络线方法）进行质量控制
    i_output = subfun1_DHI_control_baoluoxian(i_kt,i_kdf);
    new = [i_station,i_output];  % 台站号、总数量、异常数、异常比例
    overview_DHI = [overview_DHI;new];
end

%% 太阳辐射分区 Solar Radiation zoning
% 提取特征值
X = data_m(:, 2:end);
% 计算距离/相似度矩阵
D = pdist(X, 'euclidean');
% 构建连接矩阵
Z = linkage(D, 'ward');
% 剪枝并获取聚类标签
k = 8;
T = cluster(Z, 'maxclust', k);
% 将聚类标签添加到原始数据中
data_with_clusters = [data_m ones(size(data_m,1), 1) * NaN];
data_with_clusters(:, end) = T;

%% 计算 GHI 和 DHI
G = myfun_GHI(ssh,ssh0,T);
D = myfun_DHI(ssh,ssh0,T);

%% 评估、误差分析 Technical Validation
G_err = [];    % GHI 的误差分析：台站号、RMSE、RMSE%
D_err = [];    % DHI 的误差分析：台站号、RMSE、RMSE%
data_CSRDD = load("D:\CSRDD_code\Data_CSRDD.mat");
% 1 台站号、2 纬度、3 经度、4 海拔、5 年、6 月、7 日、8GHI、9DHI
for i = 1:length(station_GHI)
    i
    i_station = station_GHI(i);
    index = find(data_CMA(:,1)==i_station);
    % 单位转换,将实测的 32744 和 32766 转为 NaN，将计算的异常值转为 NaN
    GHI_msr = data_CMA(index,14)*0.01;  GHI_msr(GHI_msr > 50) = NaN;  % CMA
观测的数据单位是 0.01MJ/m2
```

```
    DHI_msr = data_CMA(index,15)*0.01;  DHI_msr(DHI_msr > 50) = NaN;
    GHI_calcu = data_CSRDD(index,8);    GHI_calcu(GHI_calcu > 50) = NaN;
    DHI_calcu = data_CSRDD(index,9);    DHI_calcu(DHI_calcu > 50) = NaN;

    i_data = [data_CMA(index,1:7),GHI_msr,DHI_msr,GHI_calcu,DHI_calcu];
    % (2)实测的 GHI  DHI 均为 0 表示无观测
    idx = find(i_data(:,8)==0 & i_data(:,9)==0);
    i_data(idx,:)=[];
    [G_rmse,G_rmsep] = subfun2_error(i_data(:,8),i_data(:,10));
    G_err = [G_err;i_station,G_rmse,G_rmsep];

    % (3)对于散射，要再剔除 NaN（113 个站，已经保证了总辐射均有观测值）
    idx1 = find(isnan(i_data(:,9)));
    i_data(idx1,:)=[];
    [D_rmse,D_rmsep] = myfun231030_error(i_data(:,9),i_data(:,11));
    D_err = [D_err;i_station,D_rmse,D_rmsep];
end
```

Subfunction

(1) subfun1_DHI_control_baoluoxian

```
function output = subfun1_DHI_control_baoluoxian(kt,kdf)
% DHI 包络线质量控制,kt 为晴空指数，kdc 为散射比

    % 0. 对 kt 和 kdf 初步质量控制
    idx = find((kt>1) | (kdf>=1) | isnan(kt) | isnan(kdf));
    kt(idx)=[];    kdf(idx)=[];

    % 1. 找出 kdf 的最大值、最小值，分成 10 个区间，得到每个区间的上下限
    range = (max(kt) - min(kt)) / 10;
    intervals = cell(10, 2);
    for i = 1:10
        intervals{i, 1} = min(kt) + (i-1) * range;
        intervals{i, 2} = min(kt) + i * range;
    end

    % 2. 对每个元素进行所在区间的判定，得到各元素的区间号码 idx_10sect
    idx_10sect = zeros(length(kt),1);
    for i = 1:length(kt)
        for j = 1:10
            if kt(i) >= intervals{j, 1} && kt(i) < intervals{j, 2}
                idx_10sect(i) = j;
                break;
            end
        end
    end

    % 3. 将元素按区间编号分为 10 组，得到每组的 kdf 均值和标准差
    means = zeros(10,1);
    stdDevs = zeros(10,1);
    num_all = zeros(10,1);
    for i = 1:10
        indices = find(idx_10sect == i);
        if ~isempty(indices)
            means(i) = mean(kdf(indices));
            stdDevs(i) = std(kdf(indices));
            num_all(i) = length(kdf(indices));
        end
    end
```

```
    % 4. 10 个区间
    f_low_f_up  = [means-2*stdDevs,means+2*stdDevs];

    % 5. 判断有多少在区间内
    num_in_sect = zeros(10,1);
    for i = 1:10
        indices = find(idx_10sect == i);
        if ~isempty(indices)
           num_in_sect(i) = length(find(kdf(indices)>f_low_f_up(i,1) &
kdf(indices)<f_low_f_up(i,2)));
        end
    end

    % 6.输出
    abno_pct = (1-sum(num_in_sect)/sum(num_all))*100;    %异常比例，单位是百分
比
    output = [sum(num_all),sum(num_all)-sum(num_in_sect),abno_pct];
end
```

(2) subfun2_assessment

```
function [rmse,rmse_percent] = subfun2_assessment(predictions, targets)
    % 计算均方根误差百分比
    % predictions: 预测值
    % targets: 实际值

    % 删除 nan
    idx = find(isnan(predictions));
    predictions(idx,:) = [];
    targets(idx,:) = [];
    idx1 = find(isnan(targets));
    predictions(idx1,:) = [];
    targets(idx1,:) = [];

    % 计算均方误差
    mse = mean((predictions - targets).^2);

    % 计算均方根误差
    rmse = sqrt(mse);

    % 计算均方根误差百分比
    rmse_percent = (rmse / mean(targets)) * 100;
end
```

（3）subfun3_GHImodel

```
function GHI = subfun3_GHImodel(T,ssh,ssh0,Tmax,Tmin,pre)
    % 区域模型
    G1 = 0.218 + 0.52*(ssh./ssh0);
    G2 = 0.247 + 0.58*(ssh./ssh0);
    G3 = 0.044 + 0.039*(Tmax - Tmin) + (0.252 +
2080.56*(1./pre)).*(ssh./ssh0);
    G4 = 0.104 + 0.034*(Tmax - Tmin) + (-0.189 +
6487.65*(1./pre)).*(ssh./ssh0);
    G5 = 0.003 + 0.048*(Tmax - Tmin) + (0.125 +
3389.29*(1./pre)).*(ssh./ssh0);
    G6 = -0.240 + 0.104*(Tmax - Tmin) + 0.375.*(ssh./ssh0).^0.644;
    G7 = 0.208 + 0.52*(ssh./ssh0);
    G8 = 0.018 + 0.032*(Tmax - Tmin) + 0.499.*(ssh./ssh0).^0.672;

    G = [G1,G2,G3,G4,G5,G6,G7,G8];

    for type = 1:8
        if T(1) == type
            GHI = G(:,type);
        end
    end
end
```

（4）subfun4_DHImodel

```
function DHI = subfun4_DHImodel(T,ssh,ssh0,GHI_CMA,G0)
    % 区域模型
    ssh_p = ssh./ssh0;
    kt = GHI_CMA./G0;
    %%%% model A %%%%%
    D_A = zeros(length(T),8);
    coef_A = [0.908 1.196   -3.892  2.28    -0.707  1.065   -0.818
              0.997    0.374   -1.811  0.94    -0.797  0.742   -0.492
              0.937    1.054   -4.29   2.845   -0.455  0.58    -0.513
              0.99 0.235   -2.014  1.458   -0.61   0.354   -0.214
              0.956    0.821   -3.748  2.267   -0.347  0.507   -0.471
              0.931    1.065   -4.133  2.573   -0.401  0.462   -0.441
              0.946    0.973   -4.292  2.886   -0.45   0.349   -0.267
              0.942    1.016   -4.375  3.057   -0.456  0.477   -0.419
              ];

    for i = 1:8
        D_A(:,i) = coef_A(i,1) + coef_A(i,2)*kt + coef_A(i,3)*kt.^2 +
coef_A(i,4)*kt.^3 + coef_A(i,5)*ssh_p + coef_A(i,6)*ssh_p.^2 +
coef_A(i,7)*ssh_p.^3;
    end

    %%%%% model B %%%%%
    D_B = zeros(length(T),8);
    coef_B = [0.21  0.297   -0.08   -0.404
              0.191    0.371   -0.336  -0.149
              0.161    0.528   -0.503  -0.073
              0.18 0.48    -0.721  0.206
              0.161    0.707   -0.968  0.239
              0.176    0.559   -0.842  0.291
              0.165    0.675   -1.099  0.361
              0.15 0.686   -1.02   0.334
              ];
    for i = 1:8
        D_B(:,i) = coef_B(i,1) + coef_B(i,2)*ssh_p + coef_B(i,3)*ssh_p.^2 +
coef_B(i,4)*ssh_p.^3;
    end

    % 选择
    DHI = zeros(length(T),1);
    for day = 1:length(T)
        if GHI_CMA(i)>0 & GHI_CMA(i)<300000
```

```
            D = D_A;
        else
            D = D_B
        end

        for type = 1:8
            if T(day) == type
                DHI(day) = D(day,type);
            end
        end

    end

end
```