

# 基于 EMD-LMD-LSSVM 联合模型的 逐时太阳辐照度预测

田翠霞, 黄 敏, 朱启兵

(江南大学轻工过程先进控制教育部重点实验室, 无锡 214122)

**摘 要:** 由于太阳辐射的随机性很大,同时考虑到传统方法建立的单一最小二乘支持向量机(least square support vector machine, LSSVM)模型精度不高,该文提出一种基于经验模态分解(empirical mode decomposition, EMD)局部均值分解(local mean decomposition, LMD)和与机器学习方法 LSSVM 联合的逐时太阳辐照度预测模型。先利用信号处理算法 EMD 及 LMD 将时间顺序数据分解成一系列相对平稳的分量序列,再对各子序列分别建立 LSSVM 预测模型,最后将各子序列预测结果进行叠加得到最终预测值。仿真结果表明,该模型取得比单一模型更好的预测效果,均方根误差精度可提高 24.59%。

**关键词:** 太阳辐照度; 预测; 机器学习; 信号处理; 时间序列

**中图分类号:** P422.1; TP181

**文献标识码:** A

## 0 引 言

光伏发电被认为是目前转换效率最高、使用期较长、可贡献大量电力的太阳能利用方式<sup>[1]</sup>。太阳能发电受到气象条件、地理位置等因素的影响,具有间歇性和不可控性,由此产生的波动性易对电网造成冲击<sup>[2,3]</sup>。对光伏出力进行有效预测有助于电力系统调度部门制定合理的调度计划,实现光伏发电与传统电源的联合优化运行<sup>[4]</sup>。太阳辐射是影响光伏发电最主要的因素<sup>[5]</sup>,因此研究高精度的辐射预测模型对光伏发电具有重要意义。

影响太阳辐射的主要因素有地表、大气、天文和人类活动 4 个方面。地表因素主要包括地形和下垫面状况,其中,地形影响主要指海拔高度和坡地坡向、坡度及起伏程度,下垫面状况主要指地表的物理性质与其覆盖状况;大气因素对辐射的影响主要通过空气分子(如 O<sub>3</sub>、CO<sub>2</sub>等)、云、水汽以及气溶胶粒子等对太阳辐射的散射、反射和吸收作用来体现;天文因素对辐射状况的影响主要通过太阳高度角、太阳倾角、日地距离以及地理纬度来体现;人类活动对太阳辐射产生的影响主要通过改变大气

中某些气体成分、气溶胶含量及地形状况来表现<sup>[6]</sup>。

国内外学者在太阳辐射模型方面已做了大量研究。目前已积累了许多具有明确物理意义的传统模型,如 ASHRAE 模型、北京模型、半正弦模型及 Collares-Pereira & Rabl 模型<sup>[7]</sup>。由于实际环境的多变性,利用这些确定模型进行辐照度预测时,模型只能机械地按特定规律计算辐射值,不能反映辐射序列的随机性,结果的精确性很难保证<sup>[8]</sup>。文献[8]介绍了一种用于预测太阳逐时总辐射的基于混沌优化神经网络的预测模型,取得了一定效果,但该模型采用试凑法得到近似最优的网络结构,在确定隐含层节点上有很大的随意性,泛化能力有待进一步加强。

近年来,支持向量机(SVM)在预测方面得到广泛研究和应用<sup>[9-11]</sup>。SVM 算法建立在统计学习理论基础之上,通过求解局域凸二次规划,从理论上推导得到全局最优解,回归预测性能良好。最小二乘支持向量机(least square support vector machine, LSSVM)是在 SVM 基础上提出的回归分类方法,通过构造损失函数将二次寻优变为求解线性方程,降低了计算的复杂性,求解速度得到提高,在回归建

收稿日期: 2016-01-05

基金项目: 国家自然科学基金(61271384; 61275155)

通信作者: 黄 敏(1974—)女,博士、教授,主要从事复杂工业过程建模与智能控制方面的研究。huangmzqb@163.com

模方面取得了良好的效果<sup>[12]</sup>。然而,真实的时间序列往往波动性大,融入一些智能算法对原始时间序列进行分解可把波动性强的序列分解成有限个相对平稳的分量序列,从而提高预测精度。例如,周松林等<sup>[13]</sup>根据风速时间序列的特点,利用小波变换对原序列进行分解,提出一种WT-SVM的预测模型;叶林等<sup>[14]</sup>进一步提出一种基于经验模态分解的EMD-SVM风电功率预测模型;樊保龙等<sup>[15]</sup>提出基于局部均值分解的LMD-SVM瓦斯涌出量预测模型。这些联合模型考虑到SVM预测模型中输入向量的选择问题,一定程度上提高了回归预测的精度和稳定性。

经验模态分解(empirical mode decomposition, EMD)的自适应分解性等优点<sup>[16]</sup>使它在地震信号分析、海洋信号分析、机械故障诊断等领域已得到广泛应用<sup>[17-19]</sup>,但EMD方法在理论上还存在一些问题,如端点效应、过包络、欠包络和模态混淆,这些问题目前仍在研究中。由于运算机理不同,局部均值分解(local mean decomposition, LMD)的迭代次数通常比EMD少,在一定程度上可抑制端点效应<sup>[20]</sup>,此外LMD还解决了欠包络和过包络问题,其PF分量比EMD的IMF分量保存了更多频率及包络信息。但作为一种新的非平稳信号处理方法,LMD相对EMD方法也有明显不足,如平滑次数较多时,信号会发生滞后或提前,且在平滑时步长不能最优确定,无快速算法。

太阳辐照度序列是具有一定周期性和随机性的非平稳时间序列,采用单一LSSVM预测方法所得到辐照度的预测误差较大。为了进行有效预测,提高预测精度,同时考虑到EMD和LMD方法各自的优缺点,本文尝试融合这2种算法对原始辐射序列进行分解,利用LMD-EMD-LSSVM联合模型对辐照度进行预测。实验仿真表明,此联合模型具有优于其他单一模型的预测精度。

## 1 实验原理与方法

### 1.1 经验模态分解

EMD是一种处理非线性、非平稳数据序列的分解方法,是对以平稳和线性假设为前提的傅里叶变换和小波分析的突破<sup>[21]</sup>。其基本思想是:若待处理数据序列的极小值或极大值的数量比上跨零点(或

下跨零点)多2个或2个以上,则该数据序列就需进行平稳化处理。EMD方法通过筛过程(sifting process)提取反映时间序列内在特征的本征模态函数(intrinsic mode function, 简称IMF),把复杂的时序数据分解成几个简单的单分量信号,刻画研究对象在不同尺度上的波动信息。

对于给出的原始时间序列样本数据 $\{x(t), t=1, 2, \dots, N\}$ ,首先计算 $x(t)$ 的局部最大值和最小值,通过插值拟合得到由局部最大值和最小值组成的上下包络 $e_{\max}(t)$ 和 $e_{\min}(t)$ 序列,接着计算上下包络序列的均值序列 $m_1(t)$ :

$$m_1(t) = \frac{e_{\max}(t) + e_{\min}(t)}{2} \quad (1)$$

得到1个移除低频的新数据序列 $h_1(t)$ :

$$h_1(t) = x(t) - m_1(t) \quad (2)$$

$h_1(t)$ 通常并不是IMF分量,因此需对 $h_1(t)$ 重复以上处理过程进行 $k$ 次筛选,直至所得到的平均曲线趋于零,此时得到的序列 $h_{1k}(t)$ 为:

$$h_{1k}(t) = x(t) - m_{1k}(t) \quad (3)$$

式中, $h_{1k}(t)$ ——第 $k$ 次筛选得到的数据。采用限制标准差 $SD$ 来判断每次筛选的结果是否为IMF分量, $SD$ 定义为<sup>[22]</sup>:

$$SD = \sum_{k=1}^T \frac{|h_{1(k-1)}(t) - h_{1k}(t)|^2}{h_{1(k-1)}^2(t)} \quad (4)$$

式中, $SD$ ——限制标准差, $SD$ 既能使 $h_{1k}(t)$ 足够接近IMF要求,又能控制分解次数,其在0.2~0.3之间取值; $T$ ——当前筛选序列的长度; $h_{1(k-1)}(t)$ ——第 $k-1$ 次筛选得到的数据。

当 $h_{1k}(t)$ 满足 $SD$ 要求时,令 $IMF_1 = h_{1k}(t)$ , $IMF_1$ 即为原始序列 $x(t)$ 的第1个IMF分量,它代表了原序列 $x(t)$ 高频成分。减去 $x(t)$ 中第1个IMF分量,得到1个移除了最高频成分的差值序列 $r_1 = x(t) - IMF_1$ ,继续上述筛过程,可依次得到后续IMF分量 $IMF_2, IMF_3, \dots, IMF_n$ 及残余项 $r_n$ ,原始序列 $x(t)$ 可表示为:

$$x(t) = \sum_{i=1}^n IMF_i + r_n \quad (5)$$

式中, $n$ ——IMF的数量。

EMD分解的本质是把数据序列 $x(t)$ 分解为各种不同特征尺度上波动的叠加,每个IMF分量可以是非线性的也可以是线性的。

## 1.2 局部均值分解

LMD 方法的实质是从原始序列中分离出纯调频信号和包络信号,将纯调频信号和包络信号相乘便能得到一个瞬时频率具有物理意义的 PF 分量,迭代至分离出所有 PF 分量,便可得到原始信号的时频分布。对于任意信号  $x(t)$ ,其分解过程为<sup>[23]</sup>:

1)找出原始信号的所有局部极值点  $n_i$ ,求出所有相邻的极值点的平均值  $m_i$ :

$$m_i = \frac{n_i + n_{i+1}}{2} \quad (6)$$

将所有相邻的均值点  $m_i$  用直线连接起来,接着用滑动平均法对其进行平滑,得到局部均值函数  $m_{11}(t)$ 。

2)求出包络估计值  $a_i$ :

$$a_i = \frac{|n_i - n_{i+1}|}{2} \quad (7)$$

将所有相邻的两个包络估计值  $a_i$  用直线连接,接着采用滑动平均法对其平滑,得到包络估计函数  $a_{11}(t)$ 。

3)从原始序列  $x(t)$  中分离出局部均值函数  $m_{11}(t)$ ,得到:

$$h_{11}(t) = x(t) - m_{11}(t) \quad (8)$$

4)对  $h_{11}(t)$  进行解调,用  $h_{11}(t)$  除以包络估计函数  $a_{11}(t)$  得到:

$$s_{11}(t) = h_{11}(t)/a_{11}(t) \quad (9)$$

对  $s_{11}(t)$  重复上述过程便可得  $s_{11}(t)$  的包络估计函数  $a_{12}(t)$ ,若  $a_{12}(t)$  不等于 1,表示  $s_{11}(t)$  不是纯调频信号,需重复上述迭代过程  $n$  次,直到  $s_{1n}(t)$  为 1 个纯调频信号,即  $s_{1n}(t)$  的包络估计函数  $a_{1(n+1)}(t)=1$ ,因此,有:

$$\begin{cases} h_{11}(t) = x(t) - m_{11}(t) \\ h_{12}(t) = s_{11}(t) - m_{12}(t) \\ \vdots \\ h_{1n}(t) = s_{1(n-1)}(t) - m_{1n}(t) \end{cases} \quad (10)$$

式中,

$$\begin{cases} s_{11}(t) = h_{11}(t)/a_{11}(t) \\ s_{12}(t) = h_{12}(t)/a_{12}(t) \\ \vdots \\ s_{1n}(t) = h_{1n}(t)/a_{1n}(t) \end{cases} \quad (11)$$

迭代的终止条件为:

$$\lim_{n \rightarrow \infty} a_{1n}(t) = 1 \quad (12)$$

在实际应用中,以不影响分解效果为前提,为

了降低迭代次数,缩短运算时间,可用式(13)作为迭代的终止条件。

$$a_{1n}(t) \approx 1 \quad (13)$$

5)对迭代过程中产生的所有包络估计函数求乘积,得到包络信号(瞬时幅值函数):

$$a_1(t) = a_{11}(t)a_{12}(t) \cdots a_{1n}(t) = \prod_{q=1}^n a_{1q}(t) \quad (14)$$

6)将纯调频信号  $s_{1n}(t)$  和包络信号  $a_1(t)$  相乘,得到原始序列的第 1 个 PF 分量:

$$PF_1(t) = a_1(t)s_{1n}(t) \quad (15)$$

它包含了原始信号中的最高频成分,为 1 个单分量的调幅-调频信号,它的瞬时幅值就是包络信号  $a_1(t)$ ,其瞬时频率  $f_1(t)$  则可由纯调频信号  $s_{1n}(t)$  求出,即:

$$f_1(t) = \frac{1}{2\pi} \frac{d[\arccos(s_{1n}(t))]}{dt} \quad (16)$$

7)将第 1 个 PF 分量  $PF_1(t)$  从原始信号  $x(t)$  中分离出来,可得 1 个新的信号  $u_1(t)$ ,将  $u_1(t)$  作为原始序列重复以上过程,循环  $k$  次,直到  $u_k$  为单调函数为止。

$$\begin{cases} u_1(t) = x(t) - PF_1(t) \\ u_2(t) = u_1(t) - PF_2(t) \\ \vdots \\ u_k(t) = u_{k-1}(t) - PF_k(t) \end{cases} \quad (17)$$

原始序列  $x(t)$  能够被所有 PF 分量及  $u_k$  重构,即:

$$x(t) = \sum_{p=1}^k PF_p(t) + u_k(t) \quad (18)$$

表明 LMD 分解未丢失原信号信息。

## 1.3 最小二乘支持向量机

LSSVM 用不等式约束替代了 SVM 的等式约束,以误差平方和损失函数作为训练集的经验损失,从而把问题转换成一个线性矩阵求解问题,降低了计算复杂性,加快了求解速度,如果初始空间具有以下形式的模型:

$$y(x) = \omega^T \psi(x) + b \quad (19)$$

式中, $x$ ——输入样本数据,  $x \in \mathbf{R}^n$ ,  $\mathbf{R}$  为实数集; $y$ ——相对应的输出,  $y \in \mathbf{R}$ ;  $\psi(x)$ ——把  $x$  映射到高维特征空间的非线性映射函数; $b$ ——位移项,决定了超平面与原点之间的距离。

$\{x_i, y_i\}_{i=1}^n$  是给定的训练集,根据结构风险最小化原理,最优化问题可表示为:

$$\min_{\omega, b, e} J_p(\omega, e) = \frac{1}{2} \omega^T \omega + \frac{1}{2} \gamma \sum_{i=1}^n e_i^2$$

$$\text{s. t. } y_i[\omega^T \psi(x_i) + b] = 1 - e_i, i = 1, 2, \dots, n \quad (20)$$

式中,  $e_i$ ——误差变量。

$\omega$  的维数趋于无穷时, 可建立解决上述问题的拉格朗日对偶问题。该问题的解为<sup>[12]</sup>:

$$y(x) = \sum_{i=1}^n a_i K(x, x_i) + b \quad (21)$$

式中,  $K(x, x_i)$ ——核函数。

本文选取径向基(RBF)作为核函数, 其定义为:

$$K(x, x_i) = \psi(x)^T \psi(x_i) = \exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right) \quad (22)$$

式中,  $\gamma$ ——正则化参数;  $\sigma^2$ ——核函数参数。这 2 个参数的选取将影响 LSSVM 的学习能力。

#### 1.4 虚假邻点法求嵌入维数

虚假邻点法<sup>[24]</sup>的基本思想是: 当维数从  $m$  变成  $m+1$  时, 考察轨线  $Y(i)$  的邻点中哪些是真实的邻点和虚假的邻点。无虚假邻点时, 可认为吸引子的几何结构被完全展开。设  $Y_m(n(i, m))$  为  $Y_m(i)$  的最近邻点, 距离为  $Y_m(i)$ , 当嵌入维数由  $m$  增至  $m+1$  时, 距离为  $\|Y_{m+1}(n(i, m)) - Y_{m+1}(i)\|$ 。若  $\|Y_{m+1}(n(i, m)) - Y_{m+1}(i)\|$  比  $\|Y_m(n(i, m)) - Y_m(i)\|$  大很多, 可认为它们是高维吸引子中两个不相邻的点, 在投影到低维轨线上时变成了相邻的两点, 这样的邻点是虚假的。因此, 若:

$$\left( \frac{\|Y_{m+1}(n(i, m)) - Y_{m+1}(i)\| - \|Y_m(n(i, m)) - Y_m(i)\|}{\|Y_m(n(i, m)) - Y_m(i)\|} \right) > R_r \quad (23)$$

则  $Y_m(n(i, m))$  为  $Y_m(i)$  的虚假最近邻点。经验证, 阈值  $R_r$  可在  $[10, 50]$  间选取<sup>[24]</sup>。对于精确的无限长数据, 按上述标准可获得较好的结果。对于具有噪声的有限长数据, 补充下列标准, 若:

$$\frac{\|Y_{m+1}(n(i, m)) - Y_{m+1}(i)\|}{\sqrt{\frac{1}{n} \sum_{j=1}^n (x_j - \bar{x})^2}} > 2, \quad \bar{x} = \frac{1}{n} \sum_{j=1}^n x_j \quad (24)$$

此时  $Y_m(n(i, m))$  也认为是  $Y_m(i)$  的虚假最近邻点。

对实际时间序列, 取  $R_r$  为  $[10, 50]$  间的某一值,  $m$  从 1 开始, 计算虚假最近邻点所占比例。逐步增加  $m$ , 当虚假最近邻点的比重低于 5%, 或虚假最近邻点开始不随  $m$  增加而减少时, 可认为吸引子的几何结构被完全打开, 此时的  $m$  即为嵌入维数。

#### 1.5 太阳辐照度联合预测模型

基于 EMD-LMD-LSSVM 的太阳辐照度预测流程如图 1 所示。

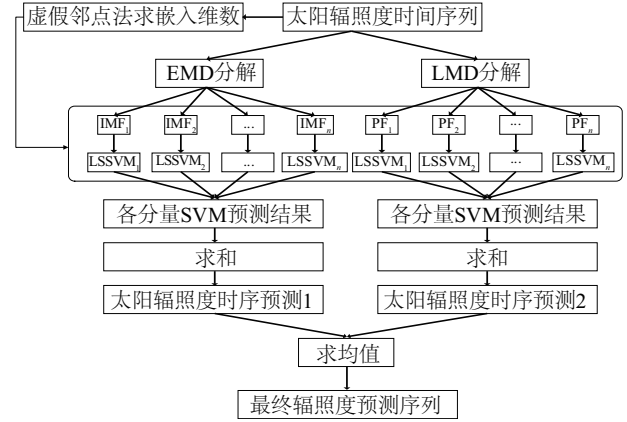


图1 太阳辐照度预测流程图

Fig. 1 The flow chart of solar irradiance prediction

具体步骤为:

1) 样本划分, 确定输入输出数据。模型的原理是利用前  $m$  时点的数据来预测下一时点的辐照度。即网络的输入为前  $m$  时点的辐照度, 而网络的输出为所预测的下一时点的辐照度, 即对于  $t$  时点辐照度  $X(t)$  为:

$$X(t) = f(X(t-1), X(t-2), \dots, X(t-m)) \quad (25)$$

本文假设  $1 \leq m \leq 30$ , 即下一时点的辐照度与之前 1~30 个时点的辐照度有关。对单一 LSSVM 模型, 利用虚假邻点法, 选出最优的  $m$  作为嵌入维数。

设  $y(1), y(2), \dots, y(n)$  为原始序列经 EMD 或 LMD 分解后的某一分量序列, 其 LSSVM 输入输出结构如下:

$$X = \begin{pmatrix} y(1) & y(2) & \dots & y(m) \\ y(2) & y(3) & \dots & y(m+1) \\ \vdots & \vdots & \ddots & \vdots \\ y(M) & y(M+1) & \dots & y(M+m-1) \end{pmatrix}$$

$$Y = \begin{pmatrix} y(m+1) \\ y(m+2) \\ \vdots \\ y(M+m) \end{pmatrix} \quad (26)$$

2) 将原始数据分别进行 EMD 分解和 LMD 分解, 得到若干个 IMF 分量和 1 个剩余分量以及若干个 PF 分量和 1 个残余分量。

3) LSSVM 训练与预测。对每个 IMF 分量和 PF 分量进行预处理, 将样本数据归一化到  $[0, 1]$  范

围内,采用上述输入输出结构作为 LSSVM 训练集,对归一化后的每个分量分别进行 LSSVM 建模,建模过程中采用十折交叉验证选择参数  $\gamma$  和  $\sigma^2$ 。得到每个分量序列的预测结果。然后分别将 EMD 及 LMD 每个分量的预测结果进行叠加,得到 EMD-LSSVM 及 LMD-LSSVM 的预测结果。最后求 2 个模型预测结果的均值,即可得到太阳辐照度下一时段的预测值。

4) 预测结果评价。根据预测结果,参考文献[25]的评价指标,采用均方根误差(RMSE)以及误差标准差 SD 对预测结果进行评价:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{x}(t) - x(t))^2}{n}} \quad (27)$$

$$SD = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (28)$$

式中,  $\hat{x}(t)$  ——预测值;  $x(t)$  ——真实值;  $n$  ——预测点的数量;  $x_i$  ——误差值;  $\bar{x}$  ——平均误差值。

## 2 结果与讨论

实验以美国国家能源部可再生能源实验室(national renewable energy laboratory,简称 NREL,网址链接: [http://www.nrel.gov/midc/srrl\\_bms/](http://www.nrel.gov/midc/srrl_bms/))采集的科罗拉多州地区 2014 年 1 月 1 日~3 月 28 日太阳辐照度数据为例进行分析,辐照度测量仪器为 CM22 总辐射表。CM22 采用了质量较好的石英圆顶,光谱范围宽,方向响应时间快,热偏移误差小,其光谱范围宽为 200~3600 nm。由于这些圆顶的光学质量高,因此方向误差被降到 5 W/m<sup>2</sup> 以下,精确度非常高。仪器采集频率为每分钟采集 1 次。

科罗拉多州为山地州,气候复杂多变。采集点地处北纬 39.74°,西经 105.18°,海拔 1828.80 m。1 月份日平均气温为 2.06 °C,最高气温为 16.82 °C,最低气温为 -18.01 °C,日平均相对湿度为 44.73%,最大相对湿度为 103.30%,最小相对湿度为 3.43%;2 月份日平均气温为 -0.61 °C,最高气温为 17.91 °C,最低气温为 -25.09 °C,日平均相对湿度为 57.54%,最大相对湿度为 104.70%,最小相对湿度为 5.60%;3 月份日平均气温为 5.16 °C,最高气温为 21.87 °C,最低气温为 -15.62 °C,日平均相对湿度为 43.61%,最大相对湿度为 107.00%,最小相对湿度为 6.23%。可见观测环境波动性很强,对模型的精度

是不小的挑战。

该数据详细记录了每天中每个时点的太阳辐照度,采样周期为 1 h,观测日期采用时间序数排列,共有 2088 个采样点,所得辐照度时间序列如图 2 所示,其中,前 1872 个(1 月 1 日~3 月 19 日)采样点用作训练集,后 216 个(3 月 20 日~3 月 28 日)用作测试集。

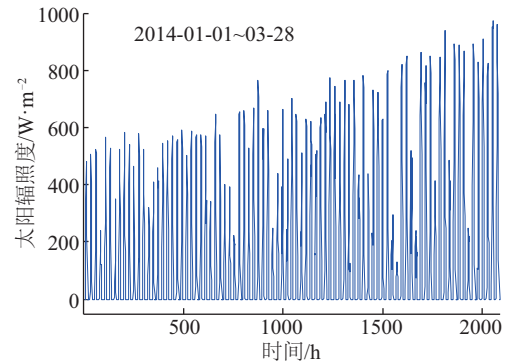


图2 太阳辐照度时间序列

Fig. 2 Solar irradiance time series

利用虚假邻近点法计算得到的嵌入维数和虚假最邻近点比例如图 3 所示,从图 3 可看出,  $m$  增加到 15 时,虚假最近邻点基本不再减少,因而选取 15 作为嵌入维数。

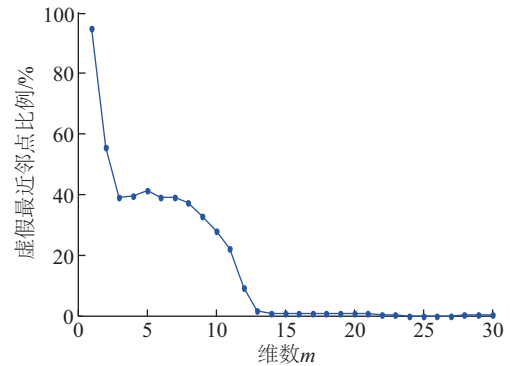


图3 虚假最近邻点比例与嵌入维数关系

Fig. 3 Relationship between the proportion of false nearest neighbor points and the embedding dimension

为了验证本文所采用模型的优越性,分别采用 LSSVM、EMD-LSSVM、LMD-LSSVM、EMD-LMD-LSSVM 共 4 种模型对超短期负荷进行预测,比较其效果。

从图 2 中可看出,时间序列具有明显的非平稳性,为了得到更好的预测效果,首先对其进行 EMD 分解,得到 7 个 IMF 分量和 1 个剩余分量,如图 4

所示。由图4可看出,IMF1~IMF2波动明显较大,较复杂。由于径向基核函数是局部核函数,有较好的局部自适应能力,学习能力较强,通过调整核参数,就能很好地模拟出相应的信号,所以选用径向基核函数较合适。然后,将各个分量预测的结果进行叠加得到最后的预测结果,进一步得到辐照度曲线图。

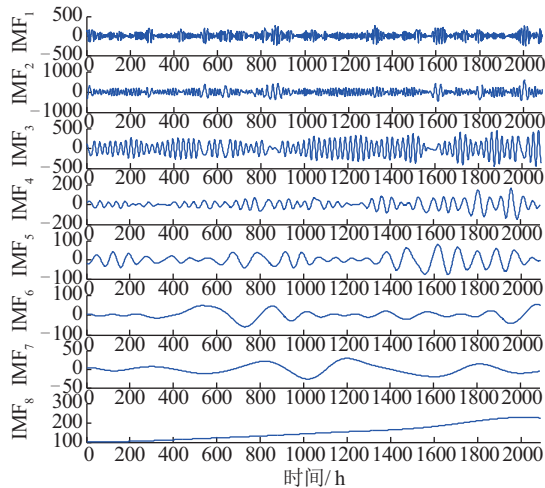


图4 太阳辐照度序列EMD分解

Fig. 4 The EMD of solar irradiance sequence

对每个IMF分量分别进行LSSVM建模,得到的各个分量参数取值以及误差如表1所示。由表1可得出前2个高频分量波动大,因而误差相对较大;其余各个分量的预测效果相对较好。最后把每个分量预测的结果相加,得到EMD-LSSVM最终预测结果,此模型RMSE的值为61.24 W/m<sup>2</sup>。

表1 EMD各IMF分量参数取值以及误差

Table1 Each IMF's parameter selection and error of the EMD

分量	$\sigma^2$	$\gamma$	$RMSE/W \cdot m^{-2}$
IMF <sub>1</sub>	9.28	19.27	49.99
IMF <sub>2</sub>	12.67	417.20	43.23
IMF <sub>3</sub>	34.37	$2.16 \times 10^5$	2.02
IMF <sub>4</sub>	651.51	$4.76 \times 10^{12}$	0.52
IMF <sub>5</sub>	2.16	$2.63 \times 10^9$	0.13
IMF <sub>6</sub>	0.21	$1.19 \times 10^6$	16.54
IMF <sub>7</sub>	0.11	$4.90 \times 10^8$	0.03
IMF <sub>8</sub>	1.33	$2.57 \times 10^{12}$	0.48

其次,对原始辐照度进行LMD分解,得到7个PF分量和1个残余分量,如图5所示。对8个PF

分量分别进行LSSVM建模,然后将每个PF的预测结果叠加得到最终预测结果。

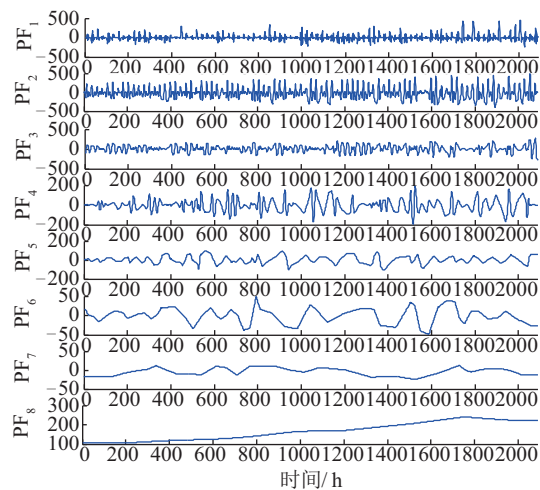


图5 太阳辐照度LMD分解

Fig. 5 The LMD of solar irradiance sequence

各个PF参数取值以及误差如表2所示。最后把每个分量预测的结果相加,得到LMD-LSSVM最终预测结果, RMSE的值为58.31 W/m<sup>2</sup>。

表2 LMD各PF分量参数取值以及误差

Table 2 Each PF's parameter selection and error of the LMD

分量	$\sigma^2$	$\gamma$	$RMSE/W \cdot m^{-2}$
PF <sub>1</sub>	7.71	4.16	52.90
PF <sub>2</sub>	30.99	655.38	17.59
PF <sub>3</sub>	36.86	743.21	11.40
PF <sub>4</sub>	139.94	$1.23 \times 10^4$	3.51
PF <sub>5</sub>	40.15	$2.00 \times 10^4$	0.57
PF <sub>6</sub>	4.42	$4.21 \times 10^4$	0.03
PF <sub>7</sub>	0.10	$6.36 \times 10^4$	0.36
PF <sub>8</sub>	0.02	$2.25 \times 10^6$	5.00

分别采用4种模型进行预测,得到的均方根误差RMSE和标准差SD及各模型相对于单一模型提高的精度如表3所示,预测误差量对比如图6所示。

由表3可得出,EMD-LMD-LSSVM的联合模型的均方根误差为51.51 W/m<sup>2</sup>, RMSE相对于单一LSSVM模型、EMD-LSSVM和LMD-LSSVM精度分别提高了24.59%、15.89%和11.66%,可见联合模型的预测误差小,精度高。另外,联合模型的误差标准差为50.81 W/m<sup>2</sup>, SD相对于单一LSSVM模型、

EMD-LSSVM 和 LMD-LSSVM 精度分别提高了 24.48%、16.90%和 10.47%,表明联合模型的预测误差波动性小,总体平稳,预测较其他模型更为准确。

表 3 4 种模型的预测误差对比

Table 3 The comparison of each model's error value

模型名称	RMSE/ $W \cdot m^{-2}$	相对单 一模型 精度提 高率/%	SD/ $W \cdot m^{-2}$	相对单 一模型 精度提 高率/%
LSSVM	68.31	—	67.28	—
EMD-LSSVM	61.24	10.35	61.14	9.13
LMD-LSSVM	58.31	14.64	56.75	15.65
联合模型	51.51	24.59	50.81	24.48

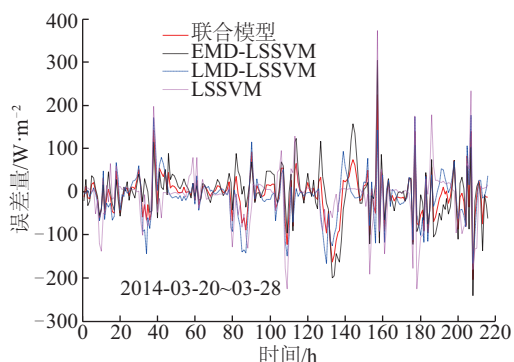


图 6 各模型误差量对比图

Fig. 6 The comparison chart of each model's error value

通过上述比较,显然,EMD-LMD-LSSVM 的预测结果要优于其他 3 种模型。由图 6 可知,联合模型的误差明显比单一模型的误差要小,且较平稳。表明本文使用的方法可在一定程度上提高辐照度预测的准确性。

图 7 中对应预测时间段为 2014 年 3 月 20 日~3 月 28 日,其中 3 月 23 日和 3 月 26 日有少许降雨。3 月 22 日云量最高,为 44.28%,云量次高的日期是 3 月 24 日,为 23.19%,云量最少的日期是 3 月 23 日,为 0.95%,可见 22 日的云量远高于其他日期。以上分析也反映出辐射值极易受云量的影响。从图 7 可看出,除去无用的夜间时段(即图中幅值为零的波段),本文使用的方法在整体上能较好地预测辐照度的变化趋势,模型不仅能在晴天进行高精度预测,且在其他天气状况、甚至太阳辐射较少的多云天气(即图 7 第 3 个半波处,其日期为 2014 年 3 月 22 日,当天的云量非常高)模型预测值

也基本和真实值一致,表明联合模型受天气影响小,验证了该方法的有效性。

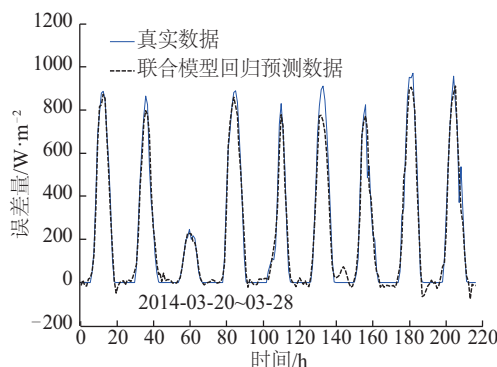


图 7 EMD-LMD-LSSVM 联合模型预测结果与真实数据对比图

Fig. 7 The comparison chart of EMD-LMD-LSSVM model forecast result and real data

### 3 结 论

太阳辐照度是一个非平稳信号,运用 EMD 和 LMD 对其平稳化处理,能得到一系列的不同频率的平稳分量,这些分量能更准确地反映原始的特征信息。根据这些分量的变化规律建立不同的 LSSVM 子模型,最后通过对子模型预测结果进行相加得到最终预测结果。试验仿真结果表明,与单一的 EMD 及 LMD 相比,EMD 及 LMD 的联合建模更适合对太阳辐照度时间序列进行处理,是一种行之有效的预测方法。

不过,LMD 作为一种新的非平稳信号处理方法,存在着最优平滑步长的选择,运算的实时性等问题。同时,本文仅针对太阳辐照度做了单步预测(也就是利用当前及以前时点的辐射数据,预测下一个时点的辐照度),如何对相关算法加以改进,并应用于太阳辐射的长期预测,将是下一步需要解决的问题。

#### [参考文献]

- [1] 张庆阳. 国外太阳能的开发利用及其借鉴[J]. 气象科技合作动态, 2009, (5): 28—32.
- [1] Zhang Qingyang. Development and utilization of foreign solar energy and its reference [J]. Meteorological Science and Technology Cooperation, 2009, (5): 28—32.
- [2] 李 宽, 李兴源, 陈 实, 等. 光伏并网抑制由直流

- 输电引起的次同步振荡的可行性分析[J]. 电力自动化设备, 2015, 35(3): 41—46.
- [2] Li Kuan, Li Xingyuan, Chen Shi, et al. Analysis on feasibility of damping HVDC-induced sub synchronous oscillation by photovoltaic grid-connection [J]. Electric Power Automation Equipment, 2015, 35(3): 41—46.
- [3] 张卫东, 刘祖明, 王振洲, 等. 基于全年辐照数据的多路径光伏柔性并网的经济性对比[J]. 电力自动化设备, 2014, 34(9): 123—129, 136.
- [3] Zhang Weidong, Liu Zuming, Wang Zhengzhou, et al. Economic comparison among multiple schemes of photovoltaic flexible grid-connection based on annual irradiance data[J]. Electric Power Automation Equipment, 2014, 34(9): 123—129, 136.
- [4] 赵唯嘉, 张 宁, 康重庆, 等. 光伏发电出力的条件预测误差概率分布估计方法[J]. 电力系统自动化, 2015, 39(16): 8—15.
- [4] Zhao Weijia, Zhang Ning, Kang Chongqing, et al. A method of probabilistic distribution estimation of conditional forecast error for photovoltaic power generation [J]. Automation of Electric Power Systems, 2015, 39(16): 8—15.
- [5] 蒋亚娟. 光伏电池建模及其在光伏发电预测中的应用 [D]. 武汉: 华中科技大学, 2011.
- [5] Jiang Yajuan. A thesis submitted in partial fulfillment of the requirements for the degree of master of engineering [D]. Wuhan: Huazhong University of Science and Technology, 2011.
- [6] 马雪清. 基于NASA/POWER数据的太阳辐射估算模型[D]. 重庆: 西南大学, 2015.
- [6] Ma Xueqing. Estimation of solar radiation using NASA/POWER data [D]. Chongqing: Southwest University, 2015.
- [7] 林星春. 基于神经网络的太阳辐射复合预测技术研究 [D]. 上海: 东华大学, 2006.
- [7] Lin Xingchun. Study of methodology of solar irradiance forecasting using compound forecast techniques on the basis of neural network [D]. Shanghai: Donghua University, 2006.
- [8] 曹双华, 曹家枞. 太阳逐时总辐射混沌优化神经网络预测模型研究[J]. 太阳能学报, 2006, 27(2): 164—169.
- [8] Cao Shuanghua, Cao Jiacong. Study of chaos optimization neural networks for the forecast of hourly total solar irradiation [J]. Acta Energiae Solaris Sinica, 2006, 27(2): 164—169.
- [9] Lingras P, Butz C J. Rough support vector regression [J]. European Journal of Operational Research, 2010, 206(2): 445—455.
- [10] Javed F, Arshad N, Wallin F, et al. Forecasting for demand response in smart grids: An analysis on use of anthropologic and structural data and short term multiple loads forecasting [J]. Applied Energy, 2012, 96: 150—160.
- [11] Sudheer C, Anand N, Panigrahi B K, et al. Streamflow forecasting by SVM with quantum behaved particle swarm optimization [J]. Neurocomputing, 2013, 101: 18—23.
- [12] Yuan Xiaohui, Chen Chen, Yuan Yanbin, et al. Short-term wind power prediction based on LSSVM-GSA model [J]. Energy Conversion and Management, 2015, 101: 393—401.
- [13] 周松林, 茆美琴, 苏建徽. 基于小波分析与支持向量机的风速预测[J]. 太阳能学报, 2012, 33(3): 452—456.
- [13] Zhou Songlin, Mao Meiqin, Su Jianwei. Wind speed forecasting based on wavelet analysis and support vector machine [J]. Acta Energiae Solaris Sinica, 2012, 33(3): 452—456.
- [14] 叶 林, 刘 鹏. 基于经验模态分解和支持向量机短期风电功率组合预测模型[J]. 中国电机工程学报, 2011, 31(31): 102—108.
- [14] Ye Lin, Liu Peng. Combined model based on EMD-SVM for short-term wind power prediction [J]. Proceedings of the CSEE, 2011, 31(31): 102—108.
- [15] 樊保龙, 白春华, 李建平. 基于LMD-SVM的采煤工作面瓦斯涌出量预测[J]. 采矿与安全工程学报, 2013, 30(6): 946—952.
- [15] Fan Baolong, Bai Chunhua, Li Jianping. Forecasting model of coalface gas emission based on LMD-SVM method [J]. Journal of Mining & Safety Engineering, 2013, 30(6): 946—952.
- [16] Moghtaderi A, Flandrin P, Borgnat P, et al. Trend filtering via empirical mode decompositions [J]. Computational Statistics and Data Analysis, 2013, 58: 114—126.
- [17] 程军圣, 于德介, 杨 宇. 基于EMD的能量算子解调方法及其在机械故障诊断中的应用[J]. 机械工程学报, 2004, 40(8): 115—118.
- [17] Cheng Junsheng, Yu Dejie, Yang Yu. Energy operator demodulating approach based on EMD and its application in mechanical fault diagnosis [J]. Chinese

- Journal of Mechanical Engineering, 2004, 40 (8) : 115—118.
- [18] Loh C H, Wu T C, Huang N E. Application of the empirical mode decomposition-Hilbert spectrum method to identify near-fault ground-motion characteristics and structural response[J]. Bulletin of the Seismological Society of American, 2001, 91(5): 1339—1352.
- [19] Marcus D, Torsten S. Performance and limitations of the Hilbert-Huang transformation (HHT) with an application to irregular water waves[J]. Ocean Engineering, 2004, 31(14-15): 1783—1834.
- [20] 程军圣, 张 亢, 杨 宇, 等. 局部均值分解与经验模式分解的对比研究[J]. 振动与冲击, 2009, 28(5): 13—15.
- [20] Cheng Junsheng, Zhang Kang, Yang Yu, et al. Comparison between the local mean decomposition and empirical mode decomposition[J]. Journal of Vibration and Shock, 2009, 28(5): 13—15.
- [21] 谢启伟, 轩 波, 李建平, 等. 经验模式分解算法的分析及应用[J]. 系统工程理论与实践, 2009, 29 (11): 168—176.
- [21] Xie Qiwei, Xuan Bo, Li Jianping, et al. Analysis and application of empirical mode decomposition algorithm [J]. Systems Engineering—Theory & Practice, 2009, 29(11): 168—176.
- [22] 阮连法, 包洪洁. 基于经验模态分解的房价周期波动实证分析[J]. 中国管理科学, 2012, 20(3): 41—46.
- [22] Ruan Lianfa, Bao Hongjie. An empirical analysis on periodic fluctuations of real estate price based on EMD [J]. Chinese Journal of Management Science, 2012, 20 (3): 41—46.
- [23] Smith J S. The local mean decomposition and its application to EEG perception data [J]. Journal of the Royal Society Interface, 2005, 2(5): 443—454.
- [24] 王丽婕, 冬 雷, 胡国飞, 等. 基于多嵌入维数的风力发电功率组合预测模型[J]. 控制与决策, 2010, 25 (4): 577—580.
- [24] Wang Lijie, Dong Lei, Hu Guofei, et al. Combined prediction of wind power generation in multi-dimension embedding phase space[J]. Control and Decision, 2010, 25(4): 577—580.
- [25] 张素宁, 田胜元. 太阳辐射逐时模型的建立[J]. 太阳能学报, 1997, 18(3): 273—277.
- [25] Zhang Suning, Tian Shengyuan. The institution of the hourly solar radiation model [J]. Acta Energiæ Solaris Sinica, 1997, 18(3): 273—277.

## HOURLY SOLAR IRRADIANCE FORECAST BASED ON EMD-LMD-LSSVM JOINT MODEL

Tian Cuixia, Huang Min, Zhu Qibing

(Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), Jiangnan University, Wuxi 214122, China)

**Abstract:** Due to the large randomness of solar radiation and considering the low accuracy of the single least squares support vector machine (LSSVM) model established by the traditional method, this paper presents a new hourly solar radiation forecasting model based on empirical mode decomposition (EMD) and local mean decomposition (LMD) combining with machine learning method LSSVM. Firstly, signal processing methods EMD and LMD are used to decompose the time series into a series of relatively stationary component sequences. Then LSSVM prediction models are established for each subsequence, and finally the prediction results of each subsequence are summed to obtain the final prediction value. The simulation results show that the model has better prediction performance than the single model, and the average root mean square error can be improved by 24.59%

**Keywords:** solar irradiance; forecasting; machine learning; signal processing; time series