



(12) 发明专利

(10) 授权公告号 CN 107067100 B

(45) 授权公告日 2020.12.04

(21) 申请号 201710055628.5

G06Q 50/06 (2012.01)

(22) 申请日 2017.01.25

G06K 9/62 (2006.01)

(65) 同一申请的已公布的文献号
申请公布号 CN 107067100 A

(56) 对比文件

CN 104766175 A, 2015.07.08

CN 105486945 A, 2016.04.13

(43) 申请公布日 2017.08.18

Man Xu. Application of change-point analysis to abnormal wind power data detection.《2014 IEEE PES General Meeting | Conference & Exposition》.2014,

(73) 专利权人 国网冀北电力有限公司
地址 100053 北京市西城区枣林前街32号
专利权人 国家电网公司 清华大学

王博. 非平稳序列的疏系数混合回归模型.《水电能源科学》.1999, 第17卷(第4期),

(72) 发明人 崔正湃 乔颖 王靖然 吴晓刚
王若阳 徐海翔 鲁宗相 孙荣富

审查员 孙旭

(74) 专利代理机构 北京华进京联知识产权代理有限公司 11606

代理人 王程

(51) Int. Cl.

G06Q 10/04 (2012.01)

权利要求书3页 说明书9页 附图7页

(54) 发明名称

风电功率异常数据辨识方法及辨识装置

(57) 摘要

本发明涉及一种风电功率异常数据辨识方法,其中,所述方法包括:根据异常数据的来源和时间序列的不规则变化,将异常数据进行分类,得到异常数据分类结果;根据异常数据的分类结果,获取异常数据判断标准;根据异常数据判断标准,对同时段风电功率序列和风速序列分别进行概率变点分析,辨识时间序列中的不规则突变;根据风电功率序列和风速序列建立混合回归模型,根据各回归变量重要性排序的变化,获得风电功率序列中的模型变点。本发明还涉及一种风电功率异常数据辨识装置。本发明提供的风电功率异常数据辨识方法和装置能够对风电功率异常数据进行准确辨识。

根据异常数据的来源和风电功率序列的不规则变化,将异常数据进行分类,得到异常数据分类结果

获取异常数据的分类结果与风速序列的对应关系,建立基于风速序列的异常数据判断标准

根据异常数据判断标准,对位于同时段的风电功率序列和风速序列分别进行概率变点分析,得到风电功率序列和风速序列中的概率变点,通过概率变点辨识风电功率序列以及风速序列中的不规则突变,根据该风速序列的不规则突变与风电功率序列中不规则突变的对应关系识别风电功率异常数据的类型

根据风电功率序列和风速序列建立混合回归模型,根据混合回归模型中各回归变量重要性排序的变化,获得风电功率序列中的模型变点,并根据模型变点识别剩余风电功率异常数据的类型

1. 一种风电功率异常数据辨识方法,其特征在于,所述方法包括:

根据异常数据的来源和风电功率序列的不规则变化,将异常数据进行分类,得到异常数据分类结果;其中,根据异常数据的产生机理将异常数据分为错误数据和不规则数据两类;所述错误数据为在数据量测、传输和保存环节中由于干扰导致的失真数据;所述不规则数据包括第一类不规则数据和第二类不规则数据,所述第一类不规则数据为正常记录存储但由于极端天气过程引起的异常数据,所述第二类不规则数据为正常记录存储但由于人工控制干预而表现异常的数据;

获取异常数据的分类结果与风速序列的对应关系,建立基于风速序列的异常数据判断标准;其中,基于风速序列的异常数据判断标准包括:

若风电功率序列和风速序列中同时存在的不规则突变则属于第一类不规则数据;

若只在风电功率序列中存在的不规则突变则属于错误数据;

若风速序列不变而风电功率序列中存在的预定时长的不规则变化,为第二类不规则数据;

根据异常数据判断标准,对位于同时段的风电功率序列和风速序列分别进行概率变点分析,得到风电功率序列和风速序列中的概率变点,通过概率变点辨识风电功率序列以及风速序列中第一不规则数据和错误数据的不规则突变,根据该风速序列的不规则突变与风电功率序列中不规则突变的对应关系识别风电功率异常数据的类型;

根据风电功率序列和风速序列建立混合回归模型,根据混合回归模型中各回归变量重要性排序的变化,获得风电功率序列中第二类不规则数据的模型变点,并根据模型变点识别剩余风电功率异常数据的类型;其中,所述回归变量重要性根据对应所述回归变量的残差变量得到;

所述根据混合回归模型中各回归变量重要性排序的变化,获得风电功率序列中第二类不规则数据的模型变点,包括:根据所述混合回归模型中各回归变量重要性排序的变化,所述混合回归模型在进行削去变换过程中,削去所述残差变量最小值对应的所述回归变量,获得风电功率序列中的模型变点。

2. 根据权利要求1所述的风电功率异常数据辨识方法,其特征在于,根据风电功率的不规则变化与风速序列的不规则变化对应关系,异常数据包括三种类型:

类型一:若不规则变化变化是由于风资源/风电功率的非平稳、非线性特征导致,则风电功率序列中的异常突变对应于相同或相近时刻风速序列中的突变,为第一类不规则数据;

类型二:若风电功率序列中的异常突变与相同或相近时刻风速序列中的突变不对应,则风电功率序列中的异常突变属于错误数据;以及

类型三:人工控制干预导致的风电功率序列异常数据,为第二类不规则数据。

3. 根据权利要求1所述的风电功率异常数据辨识方法,其特征在于,所述根据风电功率序列和风速序列建立混合回归模型,根据各回归变量重要性排序的变化,获得风电功率序列中的模型变点的步骤包括:

以风电功率序列的一阶差分结果作为因变量数组,风速序列一阶差分作为自变量数组,建立混合回归模型;

根据各回归变量重要性排序的变化,分析辨识风电功率序列中的模型变点。

4. 根据权利要求1所述的风电功率异常数据辨识方法,其特征在于,对位于同时段的风电功率序列和风速序列分别进行概率变点分析,得到风电功率序列和风速序列中的概率变点包括:

当获得一个概率变点后,将原风电功率序列根据该概率变点位置拆分为前后两个子序列;

分别对每个子序列进行辨识,直到没有新的概率变点被发现,获得风电功率序列中所有的概率变点。

5. 根据权利要求1所述的风电功率异常数据辨识方法,其特征在于,进一步包括:

利用特殊影响因子对异常数据进行分析,根据正态分布均值的假设检验,获得异常数据的来源与特殊影响因子之间的相关关系。

6. 一种风电功率异常数据辨识装置,其特征在于,所述装置包括:

异常数据分类模块,用于根据异常数据的来源和风电功率序列的不规则变化,将异常数据进行分类,得到异常数据分类结果;其中,根据异常数据的产生机理将异常数据分为错误数据和不规则数据两类;所述错误数据为在数据量测、传输和保存环节中由于干扰导致的失真数据;所述不规则数据包括第一类不规则数据和第二类不规则数据,所述第一类不规则数据为正常记录存储但由于极端天气过程引起的异常数据,所述第二类不规则数据为正常记录存储但由于人工控制干预而表现异常的数据;

标准建立模块,用于根据异常数据的分类结果,建立基于风速序列的异常数据判断标准;其中,基于风速序列的异常数据判断标准包括:

若风电功率序列和风速序列中同时存在的不规则突变则属于第一类不规则数据;

若只在风电功率序列中存在的不规则突变则属于错误数据;

若风速序列不变而风电功率序列中存在的预定时长的不规则变化,为第二类不规则数据;

第一异常数据分析模块,用于根据预设的异常数据判断标准,对位于同时段的风电功率序列和风速序列分别进行概率变点分析,得到风电功率序列和风速序列中第一不规则数据和错误数据的概率变点,通过概率变点辨识风电功率序列以及风速时间序列中的不规则突变,根据该风速序列的不规则突变与风电功率序列中不规则突变的对应关系识别风电功率异常数据的类型;

第二异常数据分析模块,用于根据风电功率序列和风速序列建立混合回归模型,根据混合回归模型中各回归变量重要性排序的变化,获得风电功率序列中第二类不规则数据的模型变点,并根据模型变点识别剩余风电功率异常数据的类型;其中,所述回归变量重要性根据对应所述回归变量的残差变量得到;

所述根据混合回归模型中各回归变量重要性排序的变化,获得风电功率序列中第二类不规则数据的模型变点,包括:根据所述混合回归模型中各回归变量重要性排序的变化,所述混合回归模型在进行削去变换过程中,削去所述残差变量最小值对应的所述回归变量,获得风电功率序列中的模型变点。

7. 如权利要求6所述的装置,其特征在于,所述第二异常数据分析模块包括:

回归模型获取单元,用于以风电功率序列的一阶差分结果作为因变量数组,风速序列一阶差分作为自变量数组,建立混合回归模型;

模型变点获取单元,根据各回归变量重要性排序的变化,分析辨识风电功率序列中的模型变点。

8.如权利要求6所述的装置,其特征在于,所述第一异常数据分析模块还用于:

当获得一个概率变点后,将原风电功率序列根据该概率变点位置拆分为前后两个子序列;

分别对每个子序列进行辨识,直到没有新的概率变点被发现,获得风电功率序列中所有的概率变点。

9.如权利要求6所述的装置,其特征在于,所述装置还包括:

相关关系建立模块,用于利用特殊影响因子对异常数据进行分析,根据正态分布均值的假设检验,获得异常数据的来源与特殊影响因子之间的相关关系。

10.如权利要求6所述的装置,其特征在于,所述异常数据包括三种类型:

类型一:若不规则变化变化是由于风资源/风电功率的非平稳、非线性特征导致,则风电功率序列中的异常突变对应于相同或相近时刻风速序列中的突变,为第一类不规则数据;

类型二:若风电功率序列中的异常突变与相同或相近时刻风速序列中的突变不对应,则风电功率序列中的异常突变属于错误数据;以及

类型三:人工控制干预导致的风电功率序列异常数据,为第二类不规则数据。

风电功率异常数据辨识方法及辨识装置

技术领域

[0001] 本发明涉及一种风电功率异常数据辨识方法及辨识装置,尤其涉及一种基于变点分析的风电功率异常数据辨识方法及辨识装置,属于电力系统预测与控制技术领域。

背景技术

[0002] 风电大规模并网背景下,其具有的波动性和随机性给传统电力系统的安全稳定运行带来巨大挑战,风电功率预测技术是帮助解决这一问题的重要手段。目前用于生产实践的功率预测系统大多采用数值天气预报(NWP)和统计学相结合的方法,不仅用到未来的天气预报情况,还用到了历史实际运行数据。历史样本是统计学方法的研究对象,其质量好坏对风电功率预测统计学方法建模效果影响很大。原始数据预处理可以为串行模式下充分发挥样本统计特性对预测精度的提升作用提供可靠的数据基础,也是风电功率预测系统的必备环节。

[0003] 以往的研究当中常用控制图方法来辨识风电功率数据中的异常值,然而,基于控制图原理的风电功率异常数据辨识方法中,大部分分散点,不论是错误数据还是不规则数据,都被辨识为异常点去除,无法适用于精细化功率预测,使得风电功率的预测的准确度不够高。

发明内容

[0004] 综上所述,确有必要提出一种能够准确对风电功率异常数据进行辨识的方法及装置。

[0005] 一种风电功率异常数据辨识方法,其中,所述方法包括:

[0006] 根据异常数据的来源和风电功率序列的不规则变化,将异常数据进行分类,得到异常数据分类结果;

[0007] 根据异常数据的分类结果,建立基于风速序列的异常数据判断标准;

[0008] 根据预设的异常数据判断标准,对位于同时段的风电功率序列和风速序列分别进行概率变点分析,得到风电功率序列和风速序列中的概率变点,通过概率变点辨识风电功率序列以及风速时间序列中的不规则突变,根据该风速序列的不规则突变与风电功率序列中不规则突变的对应关系识别风电功率异常数据的类型;

[0009] 根据风电功率序列和风速序列建立混合回归模型,根据混合回归模型中各回归变量重要性排序的变化,获得风电功率序列中的模型变点,并根据模型变点识别剩余风电功率异常数据的类型。

[0010] 在其中一个实施例中,异常数据包括错误数据和不规则数据;错误数据包括在数据量测、传输和保存等环节中由于干扰导致的失真数据;不规则数据包括正常记录存储的数据,但由于极端天气过程引起的第一类不规则数据或人工控制干预而表现异常的数据第二类不规则数据。

[0011] 在其中一个实施例中,根据风电功率的不规则变化与风速序列的不规则变化对应

关系,异常数据包括三种类型:

[0012] 类型一:若不规则变化是由于风资源/风电功率的非平稳、非线性特征导致,则风电功率序列中的异常突变对应于相同或相近时刻风速序列中的突变,为第一类不规则数据;

[0013] 类型二:若风电功率序列中的异常突变与相同或相近时刻风速序列中的突变不对应,则风电功率序列中的异常突变属于错误数据;以及

[0014] 类型三:人工控制干预导致的风电功率序列异常数据,为第二类不规则数据。

[0015] 在其中一个实施例中,基于风速序列的异常数据判断标准包括:

[0016] 若风电功率序列和风速序列中同时存在的不规则突变则属于第一类不规则数据;

[0017] 若只在风电功率序列中存在的不规则突变则属于错误数据;

[0018] 若风速序列不变而风电功率序列中存在的预定时长的不规则变化,为第二类不规则数据。

[0019] 在其中一个实施例中,所述根据风电功率序列和风速序列建立混合回归模型,根据各回归变量重要性排序的变化,获得风电功率序列中的模型变点的步骤包括:

[0020] 以风电功率序列的一阶差分结果作为因变量数组,风速序列一阶差分作为自变量数组,建立混合回归模型;

[0021] 根据各回归变量重要性排序的变化,分析辨识风电功率序列中的模型变点。

[0022] 在其中一个实施例中,对位于同时段的风电功率序列和风速序列分别进行概率变点分析,得到风电功率序列和风速序列中的概率变点包括:

[0023] 当获得一个概率变点后,将原风电功率序列根据该概率变点位置拆分为前后两个子序列;

[0024] 分别对每个子序列进行辨识,直到没有新的概率变点被发现,获得风电功率序列中所有的概率变点。

[0025] 在其中一个实施例中,进一步包括:

[0026] 利用特殊影响因子对异常数据进行分析,根据正态分布均值的假设检验,获得异常数据的产生与特殊影响因子之间的相关关系。

[0027] 一种风电功率异常数据辨识装置,其中,所述装置包括:

[0028] 异常数据分类模块,用于根据异常数据的来源和风电功率序列的不规则变化,将异常数据进行分类,得到异常数据分类结果;

[0029] 标准建立模块,用于根据异常数据的分类结果,建立基于风速序列的异常数据判断标准;

[0030] 第一异常数据分析模块,用于根据异常数据判断标准,对位于同时段的风电功率序列和风速序列分别进行概率变点分析,得到风电功率序列和风速序列中的概率变点,通过概率变点辨识风电功率序列以及风速序列中的不规则突变,根据该风速序列的不规则突变与风电功率序列中不规则突变的对应关系识别风电功率异常数据的类型;

[0031] 第二异常数据分析模块,用于根据风电功率序列和风速序列建立混合回归模型,根据混合回归模型中各回归变量重要性排序的变化,获得风电功率序列中的模型变点,并根据模型变点识别剩余风电功率异常数据的类型。

[0032] 在其中一个实施例中,所述第二异常数据分析模块包括:

[0033] 回归模型获取单元,用于以风电功率序列的一阶差分结果作为因变量数组,风速序列一阶差分作为自变量数组,建立混合回归模型;

[0034] 模型变点获取单元,根据各回归变量重要性排序的变化,分析辨识风电功率序列中的模型变点。

[0035] 在其中一个实施例中,所述第一异常数据分析模块还用于:

[0036] 当获得一个概率变点后,将原风电功率序列根据该概率变点位置拆分为前后两个子序列;

[0037] 分别对每个子序列进行辨识,直到没有新的概率变点被发现,获得风电功率序列中所有的概率变点。

[0038] 在其中一个实施例中,所述装置还包括:

[0039] 相关关系建立模块,用于利用特殊影响因子对异常数据进行分析,根据正态分布均值的假设检验,获得异常数据的来源与特殊影响因子之间的相关关系。

[0040] 上述实施例根据异常数据分析判断原则对其成因进行分类,并结合了概率变点和模型变点的时序关系,符合风资源/风电序列的时序相依特性,能够对风电功率异常数据进行准确辨识。

附图说明

[0041] 图1为一个实施例提供的风电功率异常数据辨识方法的流程框图;

[0042] 图2为风电功率异常数据辨识方法中多变点分析流程的示意图;

[0043] 图3为滚动建模子样本形成过程的示意图;

[0044] 图4为其中一个实施例中概率变点分析的示意图;

[0045] 图5为另一个实施例中概率变点分析的示意图;

[0046] 图6为实施例提供的风电功率异常数据辨识方法中采用的样本曲线的示意图;

[0047] 图7为不同延迟步长变量的排序情况,其中 $n_0=16$, $n=96$, $p=4$ 。

[0048] 图8为一个实施例中电功率异常数据辨识方法与传统方法的异常点辨识效果对比;

[0049] 图9为一个实施例提供的风电功率异常数据辨识装置的结构框图。

具体实施方式

[0050] 下面根据说明书附图并结合具体实施例对本发明的技术方案进一步详细表述。

[0051] 本发明提供的基于变点分析的风电功率异常数据辨识方法,用于为系统后续预测建模提供可靠的数据基础。本发明提供的风电功率异常数据辨识方法,通过将待分析的时间序列看作某非线性系统的输出,根据不同类型异常数据的产生机理制定分析判断原则,从概率、模型、特殊影响因子三个角度开展异常数据辨识和分析。

[0052] 请参阅图1,本发明实施例提供的风电功率异常数据的辨识方法,主要包括以下步骤:

[0053] S10,根据异常数据的来源和风电功率序列的不规则变化,将异常数据进行分类,得到异常数据分类结果;

[0054] S20,获取异常数据的分类结果与风速序列的对应关系,建立基于风速序列的异常

数据判断标准；

[0055] S30,根据异常数据判断标准,对位于同时段风电功率序列和风速序列分别进行概率变点分析,风电功率序列和风速序列中的概率变点,通过概率变点辨识风电功率序列以及风速序列中的不规则突变,根据该风速序列的不规则突变与风电功率序列中不规则突变的对应关系识别异常数据的类型；

[0056] S40,根据风电功率序列和风速序列建立混合回归模型,根据各回归变量重要性排序的变化,获得风电功率序列中的模型变点,并根据模型变点识别异常数据的类型。

[0057] 在其中一个实施例中,在步骤S10中,可以根据异常数据的产生机理将异常数据进行分类。

[0058] 其中,根据异常数据的产生机理可将异常数据分为错误数据和不规则数据两类;错误数据指在数据量测、传输和保存等环节中由于干扰导致的失真数据,不规则数据则是指正常记录存储的数据,但由于极端天气过程引起的第一类不规则数据或人工控制干预等原因而表现异常引起的第二类不规则数据。

[0059] 进一步,风电功率异常数据的表现都可以归纳为时间序列的不规则变化,即根据风电功率的不规则变化与风速序列的不规则变化对应关系,异常数据可包括三种类型:

[0060] 类型一:若不规则变化变化是由于风资源/风电功率的非平稳、非线性特征导致,则风电功率序列中的异常突变对应于相同或相近时刻风速序列中的突变,为第一类不规则数据。

[0061] 类型二:若风电功率序列中的异常突变与相同或相近时刻风速序列中的突变不对应,则风电功率序列中的异常突变属于错误数据。

[0062] 类型三:人工控制干预导致的风电功率序列异常数据,为第二类不规则数据。

[0063] 虽然风速序列中不会有所反应,但这类功率控制措施下功率序列的变化,受技术和安全等条件限制,相对于错误数据的即时产生,一般是一个较为缓变的过程,且一旦发生,会持续一段时间,即对应于第二类不规则数据,风速序列不变而风电功率序列中存在的预定时长的不规则变化。

[0064] 在步骤S20中,获取异常数据的分类结果与风速序列的对应关系,建立异常数据基本判断标准如下:

[0065] 对风电功率和风速序列中的不规则变化进行识别;

[0066] 若风电功率和风速中同时存在的不规则突变则属于不规则数据;

[0067] 若只在风电功率中存在的不规则突变则属于错误数据;

[0068] 若风速序列不变而风电功率序列中存在的预定时长的不规则变化,为第二类不规则数据。

[0069] 可以理解,变点问题的分析方法很多,根据不同类型风电功率异常数据的产生机理,可通过概率变点和模型变点对类型一、类型二以及类型三进行辨识。具体的,可选取概率变点作为类型一、类型二中不规则突变的辨识方法,选择模型变点作为类型三中较缓慢不规则变化的辨识方法,并利用特殊影响因子变点分析异常数据产生的遥相关性。

[0070] 作为其中一个实施例,在步骤S30中,基于概率变点的不规则数据辨识,对同时段风电功率序列和风速序列分别进行概率变点分析,获得风电功率序列和风速序列中的概率变点,通过概率变点辨识风电功率序列中的不规则突变。

[0071] 具体的,可根据累积概率分析变点问题,基于原假设下的渐近分布对其进行检验。考虑一类简单情况,假设时间序列中只有一个变点,事件的发生概率开始保持在 p_1 ,时刻 m_c 下发生概率突然变化为 p_2 ,则 m_c 即为变点发生时间。为了估计 m_c ,不同时刻事件累计发生次数 C_n 计算如下:

$$[0072] \quad C_n = \sum_{i=1}^n z_i \quad (1)$$

[0073] 式中 $\{z_i\}, i=1, \dots, n$,是表示每个时刻下事件是否发生的时间序列, $z_i=1$ 表示事件在 i 时刻发生, $z_i=0$ 表示没有发生。统计量 T_n 为:

$$[0074] \quad T_n = n\left(\frac{C_n}{n} - \frac{C_N}{N}\right), n=1, 2, \dots, N \quad (2)$$

[0075] 式中 C_n/n 是截止到时刻 n 的事件发生频次, C_N/N 是总事件发生频次。 T_n 的期望值为:

$$[0076] \quad E(T_n) = \begin{cases} nN^{-1}(N - m_c + 1)(p_1 - p_2), 1 \leq n \leq m_c - 1 \\ N^{-1}(N - n)(m_c - 1)(p_1 - p_2), m_c \leq n \leq N \end{cases} \quad (3)$$

[0077] 一开始, $|E(T_n)|$ 的值随着 n 的增加而增大,在 $n=m_c-1$ 时到达最大值,然后 $|E(T_n)|$ 开始减小。因此,变点判断流程可以总结为:

[0078] S31:计算统计量 T_n 的绝对值 $|T_n|, n=1, 2, \dots, N$

[0079] S32:找到变点发生时间 m^* ,使 $|T_{m^*}| = \max(|T_n|), n=1, 2, \dots, N$

[0080] S33:将 $\hat{m}_c = m^* + 1$ 作为变点发生的时刻。

[0081] 假设 $p_1 = p_2$ 下,统计量 $Nn_1^{-1}n_2^{-1}|T_{m^*}|$ 的渐近分布与斯米尔诺夫检验(CMNPFOB testing)统计量相同,其中 $n_1 = C_N, n_2 = N - C_N$ 。因此,如果有:

$$[0082] \quad \frac{N}{n_1 n_2} |T_{m^*}| > l_\alpha \quad (4)$$

[0083] 式中 l_α 是 n_1, α 限制下的斯米尔诺夫检验临界值,则对于给定置信度水平 α 下,原假设 $p_1 = p_2$ 不成立,变点 \hat{m}_c 存在,且其估计可信度可由检验置信水平保证。

[0084] 具体的,作为其中一个实施例,请参阅图2,当检验得到一个概率变点后,将原风电功率序列根据变点位置拆分为前后两个子序列,分别对每个子序列重复上述辨识过程,直到没有新的概率变点被发现,获得所有的概率变点。另外,可以根据相邻两点间的跃变是否超过一定比例将时间序列转化为0-1序列 $\{z_i\}, i=1, \dots, n$ 。然后将上述方法用于检验风电功率、风速序列中是否存在突变点。

[0085] 在步骤S40中,可以风电功率序列的一阶差分结果作为因变量数组,风速序列一阶差分作为自变量数组,建立混合回归模型,根据各回归变量重要性序列的变化,来分析辨识风电功率序列中的模型变点,该模型变点用于识别经过概率变点分析之后剩余的风电功率异常数据,且风电功率异常数据的第二类不规则数据。

[0086] 作为其中一个实施例,可采用疏系数混合回归模型来建立输入、输出变量之间的关系:

$$\begin{aligned}
 [0087] \quad & y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_p y_{t-p} \\
 [0088] \quad & + \beta_{1,1} x_{1,t-1} + \beta_{1,2} x_{1,t-2} + \dots + \beta_{1,p} x_{1,t-p} + \dots \\
 [0089] \quad & + \beta_{m,1} x_{m,t-1} + \beta_{m,2} x_{m,t-2} + \dots + \beta_{m,p} x_{m,t-p} + \varepsilon_t \quad (5)
 \end{aligned}$$

[0090] 式中 $\{y_t\}$, $t=1, 2, \dots, N$ 是因变量序列, $\{x_{i,t}\}$, $i=1, \dots, m, t=1, 2, \dots, N$, 是 m 种相关影响变量序列, $\alpha_1, \dots, \alpha_p$ 是因变量延迟变量系数, $\beta_{1,1}, \dots, \beta_{1,p}, \dots, \beta_{m,p}$ 是相关影响变量延迟变量系数, p 是最大延迟步长, ε_t 是白噪声序列。式 (5) 中回归模型既包含了对因变量自回归特性的考虑, 也包含了对影响变量混合回归特性的考虑。

[0091] 下面简要介绍模型估计过程, 将式 (5) 整理为一般回归模型因变量和自变量形式:

$$[0092] \quad y_t = \alpha_1' x_{t,1}' + \alpha_2' x_{t,2}' + \dots + \alpha_s' x_{t,s}' + \varepsilon_t \quad (6)$$

[0093] 式中 $s = mp + 1$ 。进一步写成:

$$[0094] \quad Y = X\alpha + \varepsilon \quad (7)$$

[0095] 式中 $Y = [y_1, y_2, \dots, y_N]^T$, $\alpha = [\alpha_1', \alpha_2', \dots, \alpha_s']^T$, $\varepsilon = [\varepsilon_1, \dots, \varepsilon_N]^T$,

$$X = \begin{bmatrix} x_{1,1} & \dots & x_{1,s} \\ \vdots & \ddots & \vdots \\ x_{N,1} & \dots & x_{N,s} \end{bmatrix} \quad \text{求解的正规方程为:}$$

$$[0096] \quad X^T X \alpha = X^T Y \quad (8)$$

[0097] 对式 (8) 进行中心化和标准化, 得到:

$$[0098] \quad T\alpha' = Y' \quad (9)$$

[0099] 式中 T 的第 i 行、第 j 列的元素是 $t_{i,j}$, $i, j=1, 2, \dots, s$, $Y' = [y_1', y_2', \dots, y_s']^T$ 。对

$$T \text{ 加边, 得到 } T = \begin{bmatrix} t_{1,1} & \dots & t_{1,s} & y_1' \\ \vdots & \ddots & \vdots & \vdots \\ t_{s,1} & \dots & t_{s,s} & y_s' \\ y_1' & \dots & y_s' & y_0 \end{bmatrix}, \quad \text{其中 } y_0 = 1.$$

[0100] 采用逐步回归的方法, 对 T 进行 s 阶削去变换。在每一阶, 比较不同变量削去后的残差变化量 $y_i^2/t_{i,i}$, 若某变量对应的残差变化量是最小值, 则说明该变量对因变量的贡献最小, 应将其削去。对 T 矩阵全部阶数的削去变换完成后, 按照 BIC 准则 (Bayesian Information Criterion, 贝叶斯信息准则), 比较各阶削去后的 BIC 值, 得到定阶模型。对于本发明的研究, 更重要的是, 根据各阶变换削去情况, 也同时得到了回归变量的重要性排序。

[0101] 采用对总样本滚动建模的方式来考察样本中的变点问题。请一并参阅图 3, n_0 是滚动时间间隔, n 为子样本大小, p 为最大延迟步长。变点分析步骤如下:

[0102] 第一步: 设定 n_0, n, p 的大小;

[0103] 第二步: 利用每份子样本建立疏系数混合回归模型, 同时根据残差变化量对回归变量的重要性进行排序;

[0104] 第三步: 以不同子样本对应模型回归变量重要性排序的变化表征回归模型统计特性的变化。若从第 1 期样本建模开始, 变量排序发生变化, 则可估计变点发生的时段为 $(1-2)n_0 + n \sim (1-1)n_0 + n$ 。

[0105] 进一步, 可利用特殊影响因子对异常数据进行分析, 从较长期统计角度, 根据正态

分布均值的假设检验,获得异常数据的产生与特殊影响因子之间的遥相关关系。

[0106] 受风资源/风电时空波动特性影响,风电功率的变化会受到某些影响因子遥相关关系的影响。一些情况下,风电功率序列中异常数据的产生原因是特殊天气过程,例如雨、雪、霾等,作为特殊影响因子。从较长期统计角度,根据正态分布均值的假设检验,来分析异常数据的产生与特殊影响因子之间的关系。

[0107] 将 $\{y_t\}, t=1, \dots, N$ 作为待分析的对象,是一定时间段 Δt 内的异常数据量组成的时间序列。选取可能导致 $\{y_t\}$ 发生变化的特殊影响因子,将其发生的时段作为特殊时段,标记为 $t_1, t_2, \dots, t_n, t_n \leq N$ 。按照和特殊时段相隔不同时间从 $\{y_t\}$ 中抽取子样本 $\{y_{t+i}\}, i=1, \dots, n, 1=0, \pm 1, \pm 2, \dots, \pm L, L$ 为影响因子最大考察时间间隔。对于每一份子样本 $\{y_{t+i}\}$, 计算其均值 $\overline{y_{(i)}}$:

$$[0108] \quad \overline{y_{(i)}} = \frac{1}{n_i} \sum_{i=1}^n y_{t+i} \quad (11)$$

[0109] 式中如果 $y_{t+i} \notin \{y_t\}$ 则 $y_{t+i} = 0$, n_i 为有效样本数量。统计量 U 计算如下:

$$[0110] \quad U = \frac{\overline{y_{(i)}} - \mu}{\sigma / \sqrt{n_i}} \quad (12)$$

[0111] 式中 μ 为 $\{y_t\}$ 的均值, σ 为其标准差。原假设为 $\overline{y_{(i)}} = \mu$, 即 $\{y_t\}$ 不受特殊因子影响, $\overline{y_{(i)}}$ 与 μ 没有显著性区别。但是,如果 $|U|$ 超过了一定置信度水平下的临界值,则表明 $\overline{y_{(i)}}$ 与 μ 存在显著性区别。这种情况下,说明距离特殊因子发生 $1 \cdot \Delta t$ 时间会对异常数据的发生有影响。

[0112] 实施例一

[0113] 下面结合具体的实际数据,进一步说明本发明应用于风电功率异常数据辨识的效果。算例中使用的数据来源于我国北方某并网大型风电场,风电功率数据和测风塔数据从 SCADA 系统中采集,覆盖时间窗口为 1 年。具体步骤如下:(1) 基于概率变点的不规则突变数据辨识

[0114] 对同时段风电功率序列和风速序列分别进行概率变点分析,辨识序列中的不规则突变,进一步比较二者分析结果,来帮助判断造成功率序列中不规则突变的原因。请一并参阅图 4 及图 5,分别给出了两种典型分析结果。在图 4 中,功率序列的分析结果 $|T_n|$ 在 $i=32$ 达到最大值,风速序列的分析结果在 $i=31$ 达到最大值,且对应的式 (4) 计算结果分别为 2.040 和 2.271,均大于斯米尔诺夫检验临界值 $1_{0.01} = 1.628$,因而判定为变点存在。功率序列变点的发生时刻与风速变点十分相近,根据本发明中所述分析判断原则,该功率变点属于类型一:由天气变化导致的功率突变,为不规则数据。对照功率和风速曲线图相应时刻的变化也可以验证该辨识结果。而图 5 中,功率序列 $|T_n|$ 在 $i=29$ 达到最大值,且统计量计算结果为 2.045,大于 $1_{0.01} = 1.628$,但该功率变点对应时段附近 ($i \pm 2$) 不存在风速变点,因此判断 $\hat{m}_c = m^* + 1 = 30$ 时刻的功率样本属于类型二:错误数据。该结果同样可以在功率和风速曲线

图上得到验证。

[0115] (2) 基于模型变点的异常数据分析

[0116] 以风电功率序列的一阶差分结果作为 $\{y_t\}$, 风速序列一阶差分作为 $\{x_{1,t}\}$, 根据式 (5) 建立混合回归模型, 根据各回归变量重要性排序的变化, 来分析辨识风电功率序列中的模型变点。以图6和图7所示算例说明模型辨识效果。对于图6中截取的风速、风电功率序列, 图7按照不同类型的回归变量 (风速/功率), 给出了每次滚动建模得到的变量重要性排序前两位变化情况, 其中延迟步长的时间间隔是15分钟, 横坐标第 $(l-1)n_0+1 \sim (l-1)n_0+n_0$ 序列点对应的是第 l 期子样本建模排序。可以看出, 从第16、17期建模开始, 风速变量的排序发生了明显变化, 则推知功率序列中模型变点对应的位置在321~352点。在这之前, 变量排序基本符合风速/风电功率在较短延迟步长下的自相关性, 即 $y_{t-1}, x_{1,t-1}$ 分别是功率、风速重要性排序第一的回归变量。变点发生后, y_{t-1} 保持第一的排序, 但 $x_{1,t-1}$ 排序变得较为不规律, 且这种情况在其后多期建模中持续。本算例中, 对比图7中的序列波动情况, 该模型变点代表特殊功率控制时段的开始, 为异常数据类型三所述的情况, 属于不规则数据。

[0117] (3) 特殊天气因子影响分析

[0118] 选取沙尘和降雪作为特殊天气因子来研究其与异常数据发生之间的关联性。所用的异常数据包括先前剔除的明显错误数据, 以及变点分析所得的异常数据。式 (12) 中的统计量 U 计算结果如表1所示。对于沙尘天气因子, 在取式 (11) 中延迟量 l 为0和1时, 统计量 U 大于正态分布限值1.64 (置信度水平0.1), 说明沙尘的发生与当天发生及后1日的异常数据产生之间具有显著的关联性。对于降雪影响因子, 从统计量 U 的计算结果看, 降雪和异常数据产生的关联性可以从发生当天开始一直到发生后2日。

[0119] 表1异常数据发生与天气因子之间的关系 (间隔时段 $\Delta t = 1\text{day}$)

| 天气因子 | 统计量 U | | | | |
|-----------|---------|-------|------|-------|-------|
| | 前 2 日 | 前 1 日 | 当天 | 后 1 日 | 后 2 日 |
| [0120] 沙尘 | -0.1 | 0.2 | 1.94 | 1.73 | 1.36 |
| 雪 | 0.24 | 0.36 | 1.96 | 1.72 | 1.66 |

[0121] (4) 不同异常数据辨识方法获得的风速-功率散点图比较

[0122] 为了进一步论证基于变点分析的异常数据辨识方法的优势, 将其与传统的基于控制图原理的异常数据辨识方法进行比较, 图8给出了使用两种方法前后的风速-功率散点图对比。比较数据窗口为2个月。变点分析结果的图例中, “控制” 对应类型三的异常数据, “突变” 对应类型一的异常数据, “错误” 对应类型二的异常数据。

[0123] 可以看出, 受样本分散性较高的影响, 基于控制图原理的三倍标准差法的异常数据筛选结果不太理想, 特别是高风速区段, 由于限电较多, 控制图方法基本无法对其进行辨识。

[0124] 进一步, 本发明实施例进一步提供一种风电功率异常数据辨识装置, 所述装置包括:

[0125] 异常数据分类模块110, 用于根据异常数据的来源和风电功率序列的不规则变化,

将异常数据进行分类,得到异常数据分类结果;

[0126] 标准建立模块120,用于获取异常数据的分类结果与风速序列的对应关系,建立基于风速序列的异常数据判断标准;

[0127] 第一异常数据分析模块130,用于根据异常数据判断标准,对位于同时段的风电功率序列和风速序列分别进行概率变点分析,得到风电功率序列和风速序列中的概率变点,通过概率变点辨识风电功率序列以及风速序列中的不规则突变,根据该风速序列的不规则突变与风电功率序列中不规则突变的对应关系识别异常数据的类型;

[0128] 第二异常数据分析模块140,用于根据风电功率序列和风速序列建立混合回归模型,根据各回归变量重要性排序的变化,获得风电功率序列中的模型变点,并根据模型变点识别异常数据的类型。

[0129] 在其中一个实施例中,所述第一异常数据分析模块还用于:

[0130] 当获得一个概率变点后,将原风电功率序列根据该概率变点位置拆分为前后两个子序列;

[0131] 分别对每个子序列进行辨识,直到没有新的概率变点被发现,获得风电功率序列中所有的概率变点。

[0132] 在其中一个实施例中,所述第二异常数据分析模块还包括:

[0133] 回归模型获取单元,用于以风电功率序列的一阶差分结果作为因变量数组,风速序列一阶差分作为自变量数组,建立混合回归模型;

[0134] 模型变点获取单元,根据各回归变量重要性排序的变化,分析辨识风电功率序列中的模型变点。在其中一个实施例中,所述风电功率异常数据辨识装置还包括:

[0135] 相关关系建立模块,用于利用特殊影响因子对异常数据进行分析,根据正态分布均值的假设检验,获得异常数据的来源与特殊影响因子之间的相关关系。

[0136] 本发明实施例提供的异常数据辨识方法及装置,具有以下特点:(1)不仅可以辨识异常数据,还可以对其成因进行分类;(2)不论是概率变点方法还是模型变点方法,都考虑了样本点间的时序关系,而不是完全将其作为孤立点看待,因此从图8中可以看出不仅离散分布的异常点可以被辨识,位于样本集中区域的异常点也可以被辨识。(3)特殊影响因子的变点分析,可以帮助判断特殊天气现象与异常数据发生之间的关联性,这为不同类型数据间的遥相关分析提供了一种有效方法,也有助于利用极端天气的预报结果,对异常数据的多发进行预判、预警。(4)由于变点分析方法对数据分布没有特定假设,因此并不局限于风电功率异常数据辨识,也可用于风电功率预测复合数据源中其它类型的数据,如风速、风向、气温等,从而提高了风电概率预测的准确性。

[0137] 以上所述实施例仅表达了本发明的几种实施方式,其描述较为具体和详细,但并不能因此而理解为对本发明专利范围的限制。应当指出的是,对于本领域的普通技术人员来说,在不脱离本发明构思的前提下,还可以做出若干变形和改进,这些都属于本发明的保护范围。因此,本发明的保护范围应以所附权利要求为准。

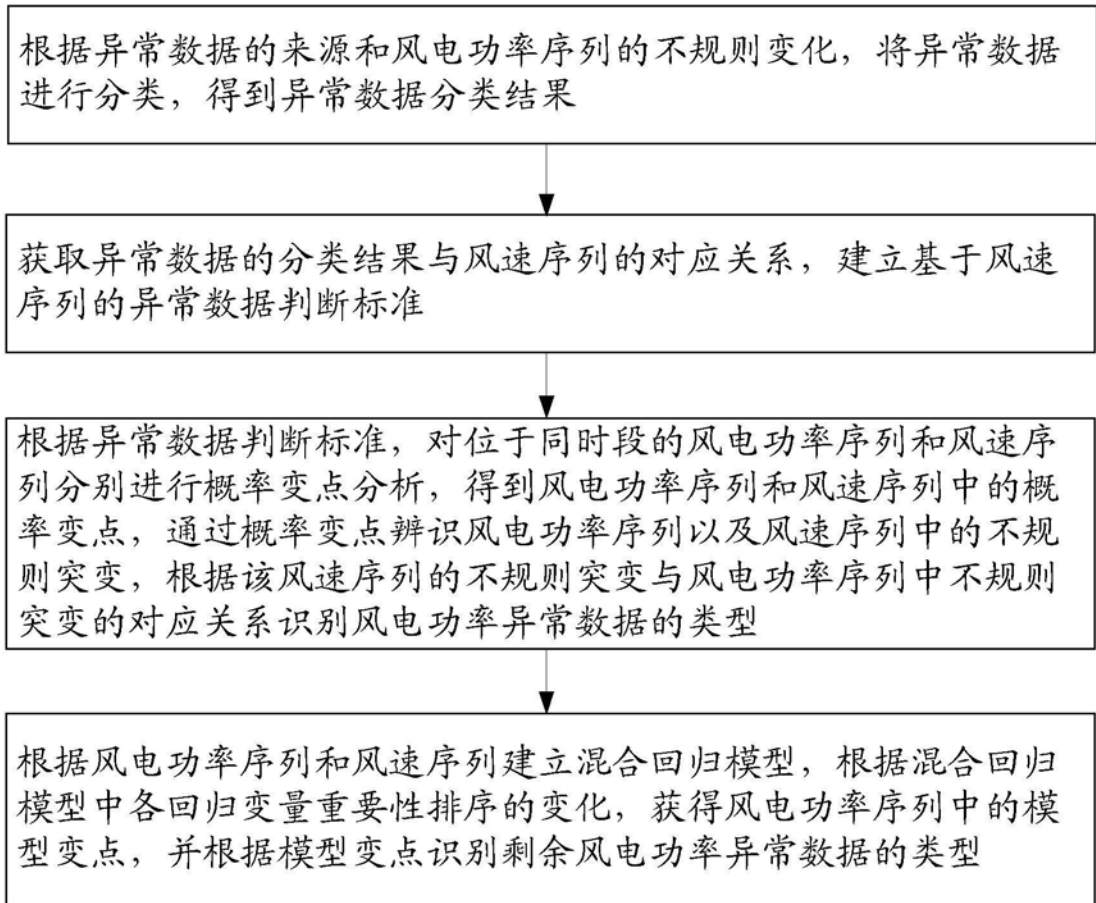


图1

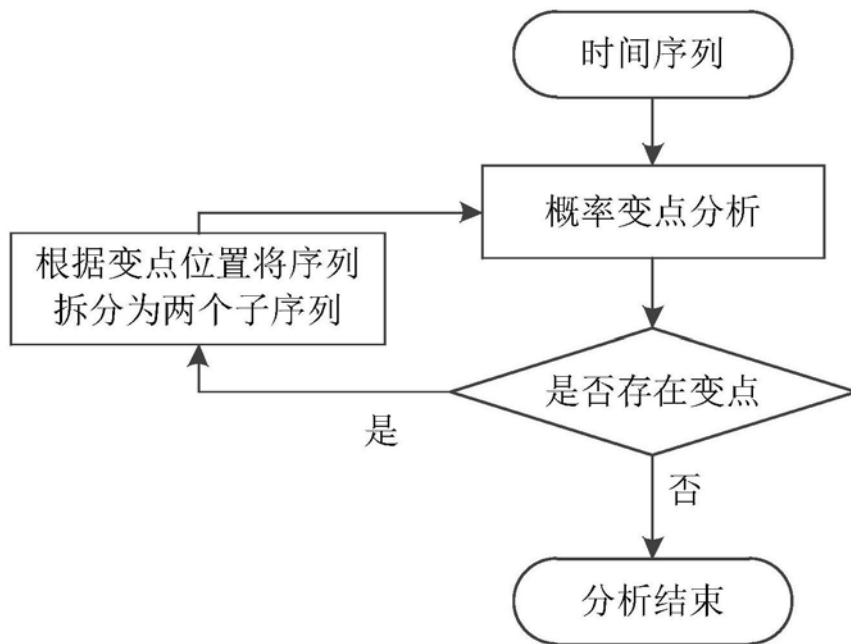


图2

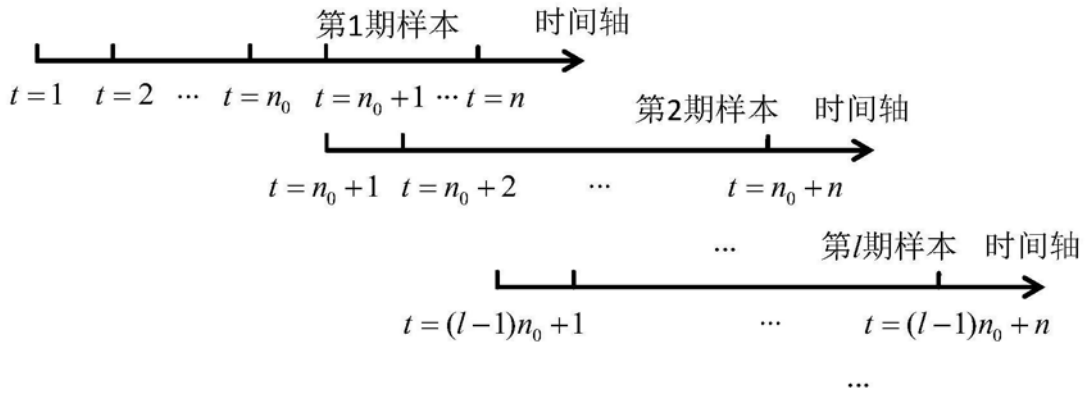


图3

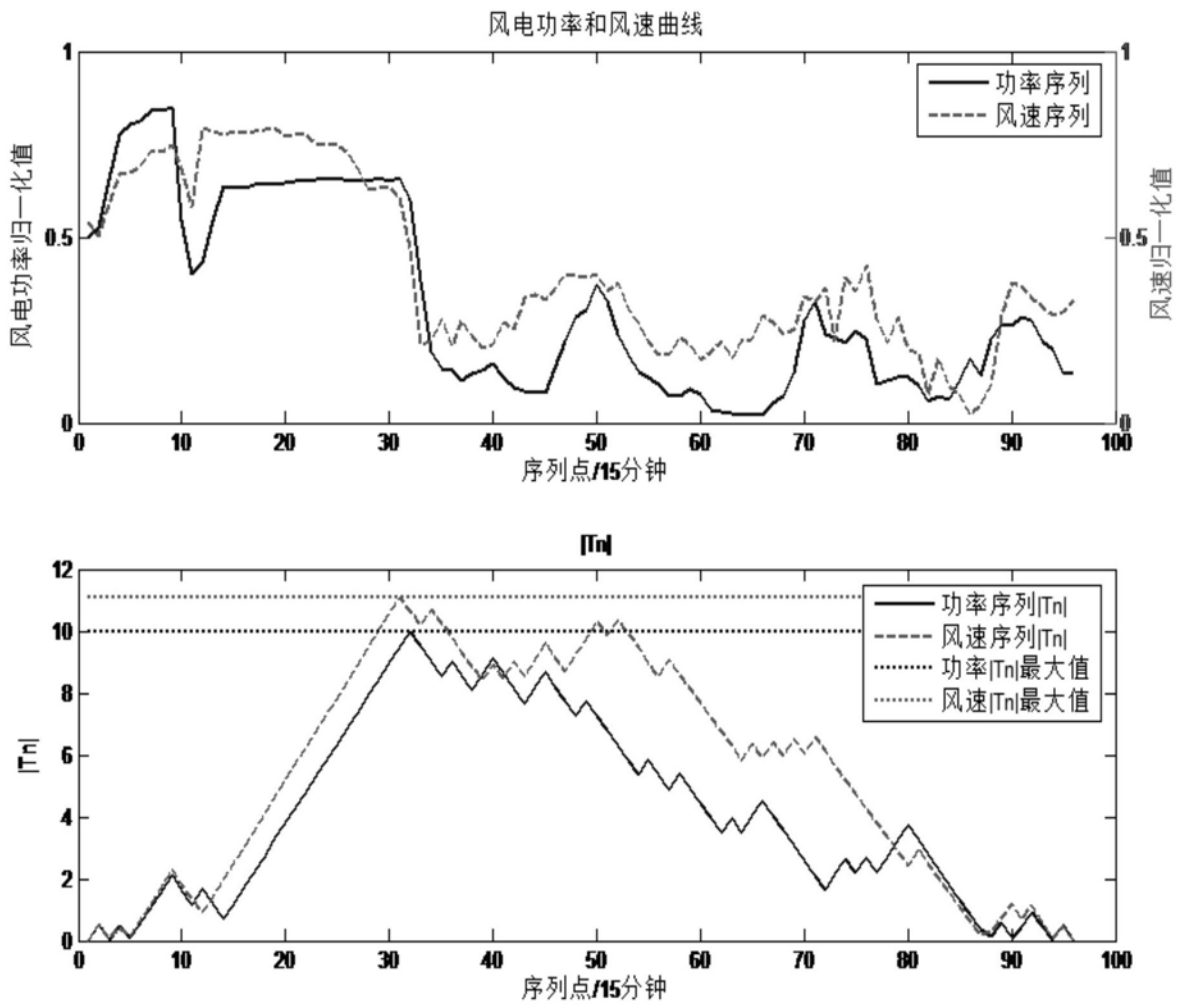


图4

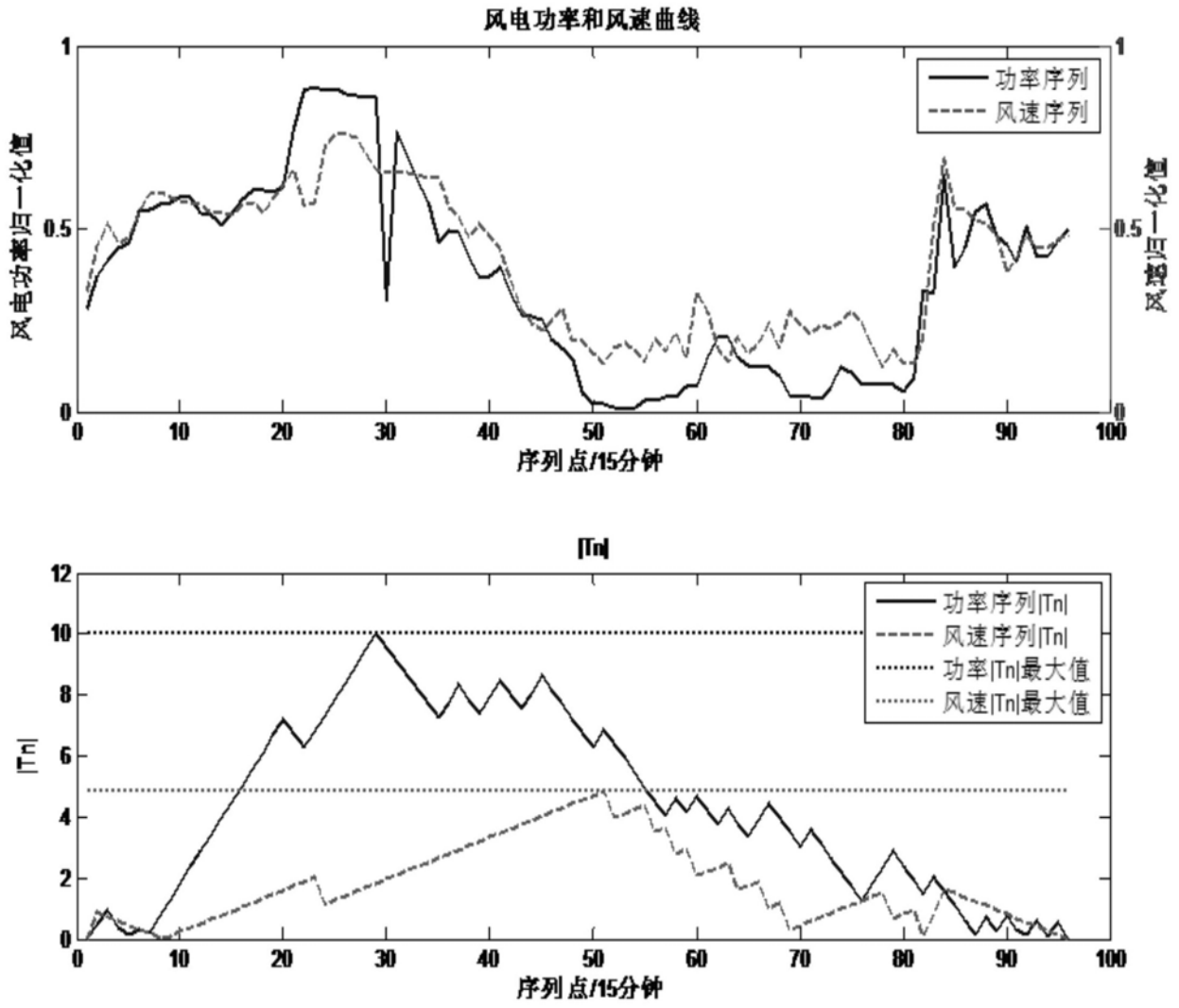


图5

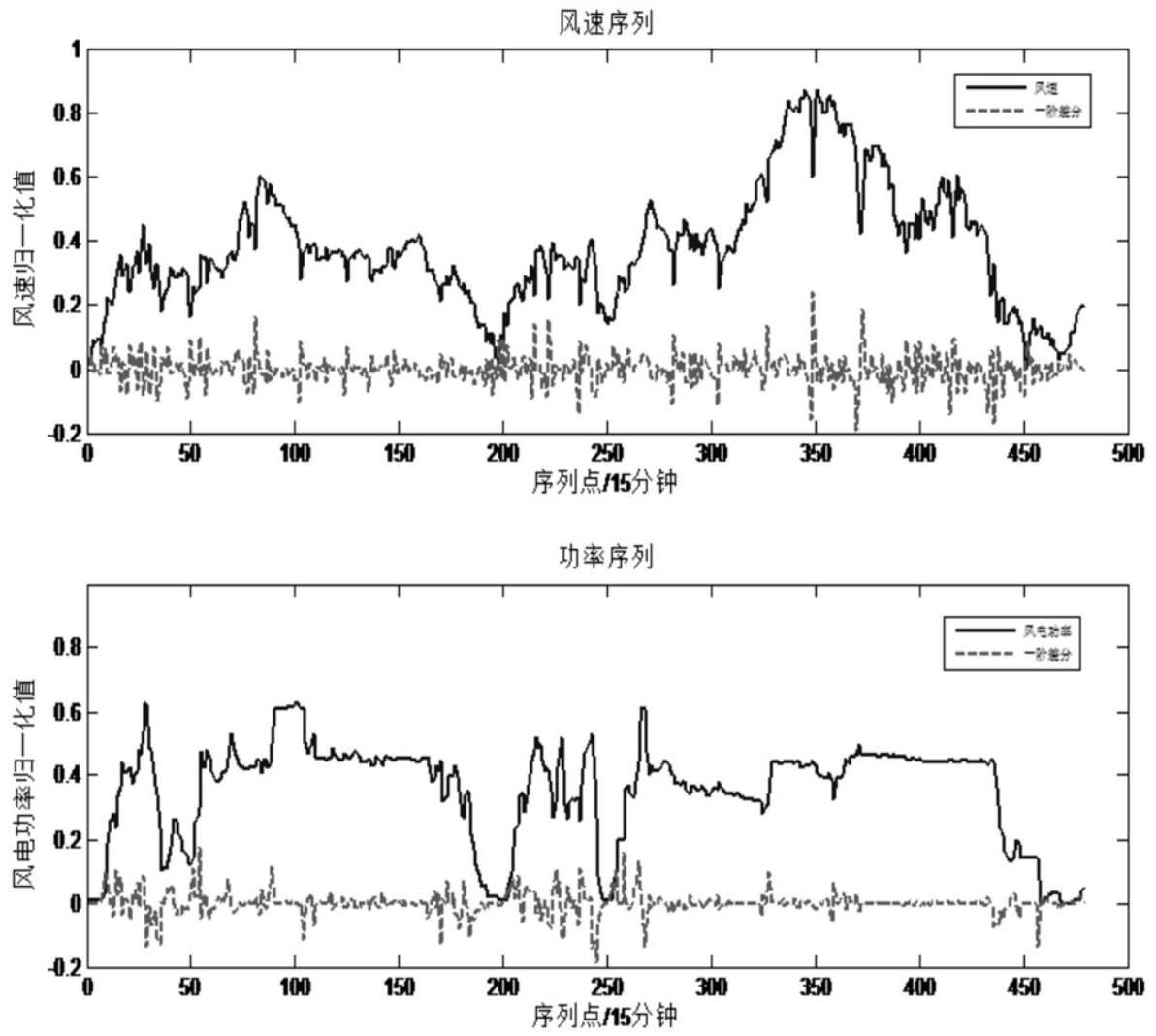


图6

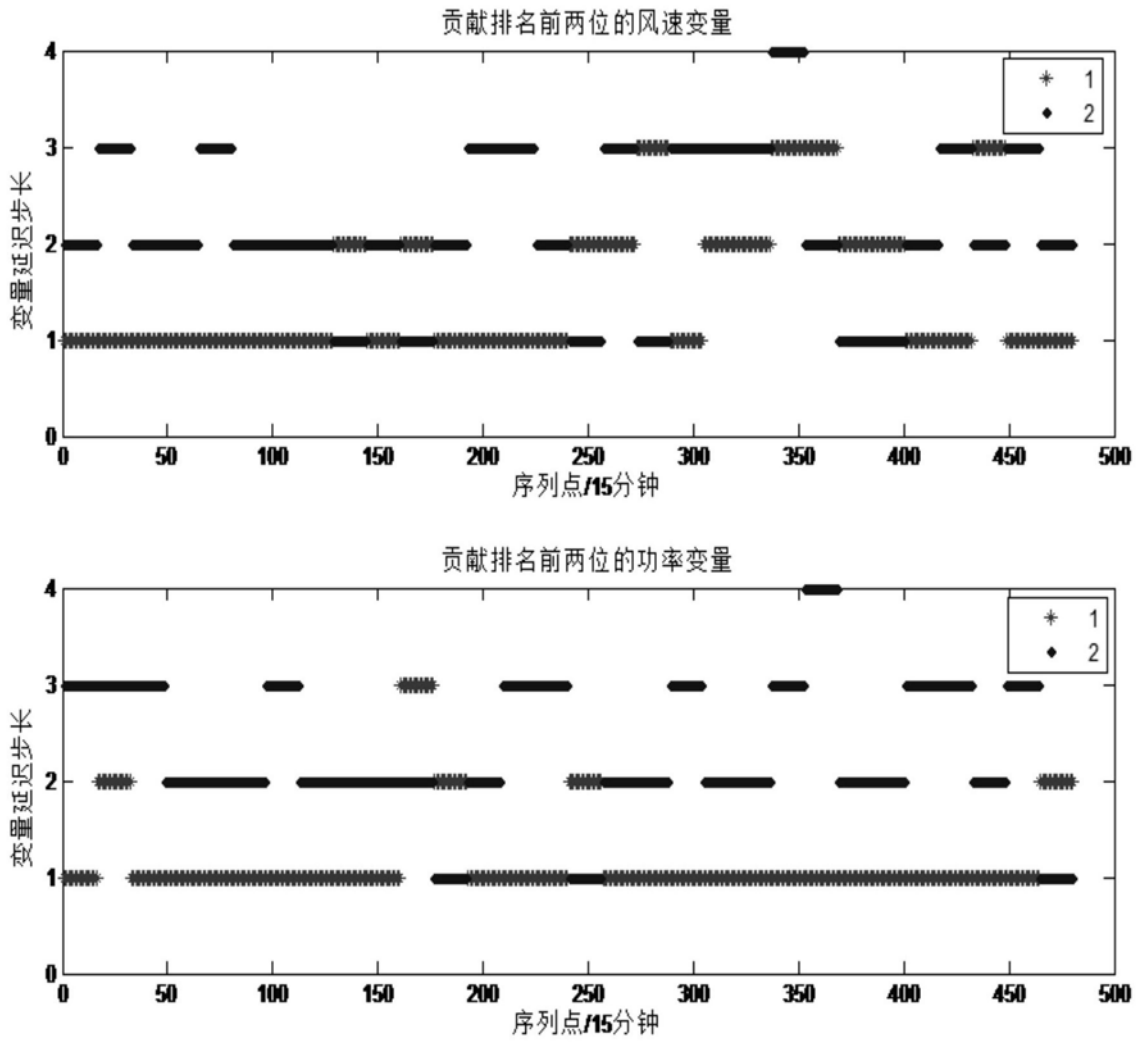


图7

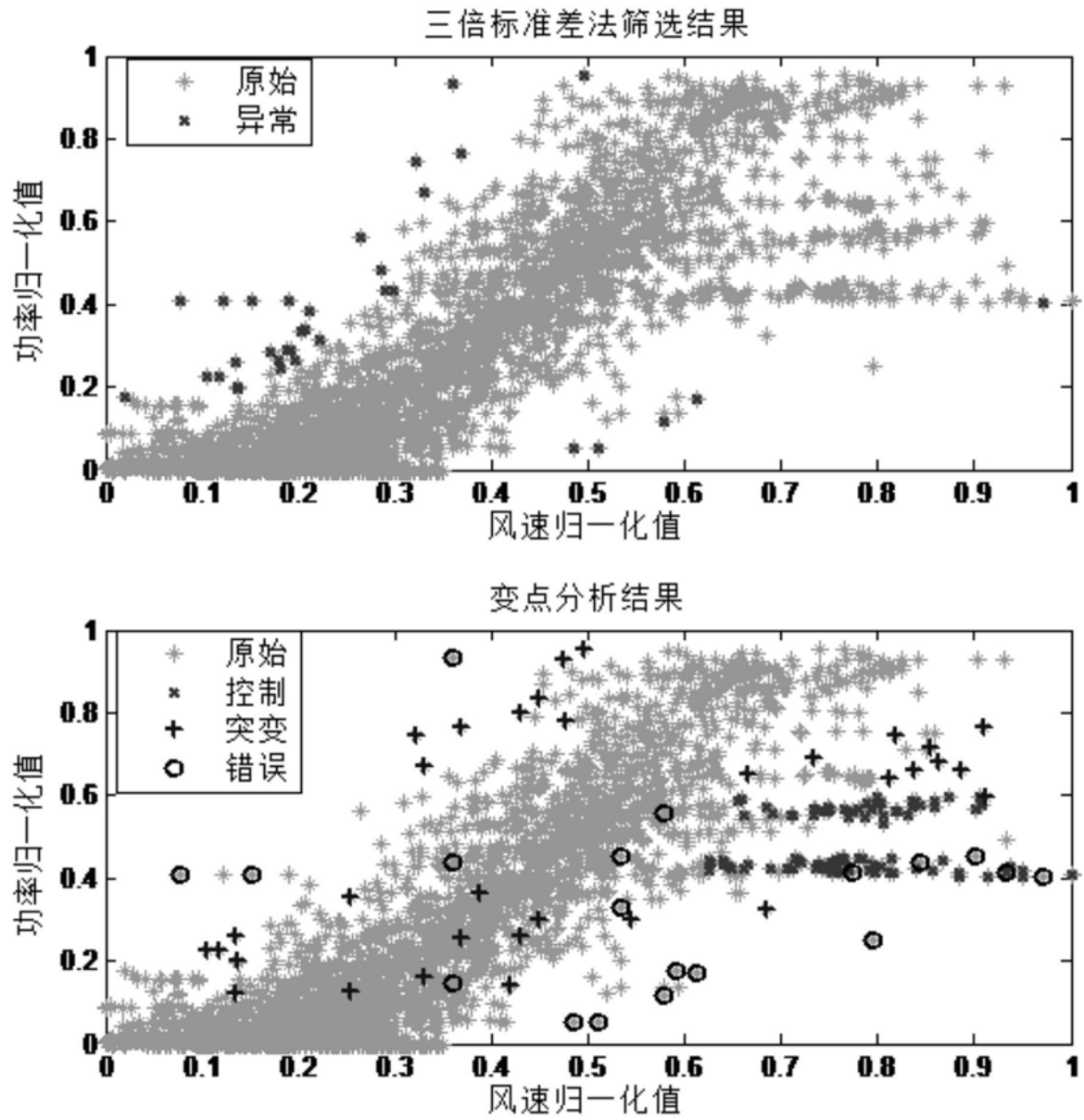


图8

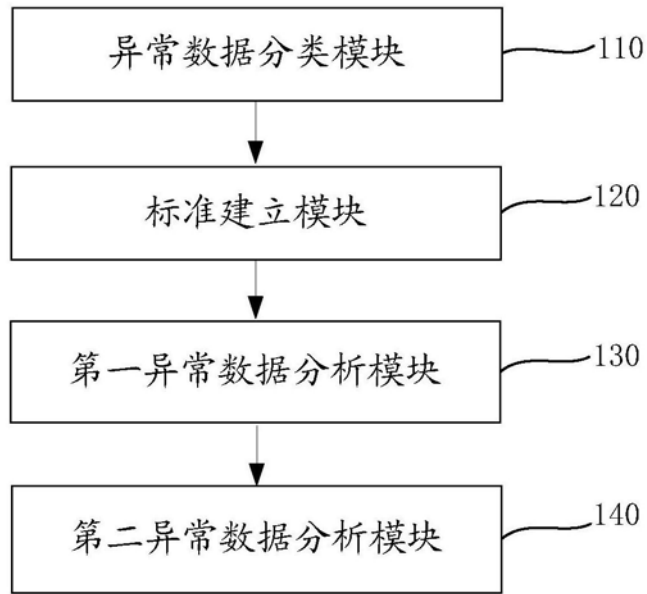


图9