

T/CIS

中国仪器仪表学会团体标准

T/CIS 01001—XXXX

近红外光谱分析技术术语

Terms of near-infrared analysis

（征求意见稿）

在提交反馈意见时，请将您知道的相关专利连同支持性文件一并附上。

XXXX—XX—XX 发布

XXXX—XX—XX 实施

中国仪器仪表学会 发布

前 言

本文件按照GB/T 1.1—2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构部承担识别专利的责任。

本文件由中国仪器仪表学会提出并归口。

本文件起草单位：

本文件主要起草人：

近红外光谱分析技术术语

1 范围

本文件界定了近红外光谱分析技术常用的基本术语和定义。

本文件适用于涉及近红外光谱分析技术领域的教学、科研、生产、应用、贸易和技术交流以及技术文件、书刊的编写。

2 规范性引用文件

本文件没有规范性引用文件。

3 近红外光谱分析技术原理术语

3.1 近红外光谱分析技术基本原理术语

3.1.1

近红外光 near infrared light

波长范围在780 nm到2500 nm（对应波数范围12820 cm^{-1} 到4000 cm^{-1} ）之间的电磁波。

3.1.2

近红外光谱 near infrared spectroscopy

一种利用近红外波段电磁辐射（波长范围780-2500 nm，对应波数范围12820-4000 cm^{-1} ）与物质发生反射、透射、散射等相互作用而产生的吸收光谱信号。

3.1.3

波长 wavelength

λ [ISO]

在周期波传播方向上，相邻两波同相位点间的距离。为了方便起见，通常在波形的极大值或极小值处进行测量。波长示意图见图1。

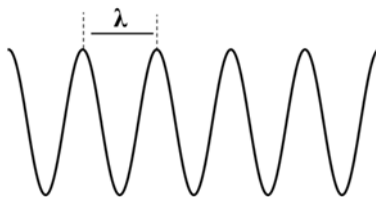


图1 测量波长示意图

注：近红外光谱分析中单位一般为纳米（nm）。

[来源：GB/T 8322-2008, 2.3, 有修改]

3.1.4

波数 wavenumber

σ [ISO]

每厘米中所含波的数目，即等于**波长**（3.1.3）的倒数。

注1：波数的单位为厘米 $^{-1}$ （ cm^{-1} ）。

注2：波数和波长的换算关系为： $\sigma = 10^7/\lambda$ ，其中， σ 为波数，单位是 cm^{-1} ， λ 为波长，单位是nm。

[来源：GB/T 8322-2008, 2.4, 有修改]

3.1.5

频率 frequency

ν

单位时间内电磁辐射振动周数。

注：频率的单位为赫兹（Hz）。

[来源：GB/T 8322-2008, 2.5]

3.1.6

透射比 transmittance

τ

透过的辐射（光）通量与入射辐射（光）通量之比。

注：透射比的计算公式为：

$$\tau = \frac{\Phi_{tr}}{\Phi_0}$$

式中：

Φ_{tr} ——透过的辐射（光）通量，单位：瓦[特]（W）；

Φ_0 ——入射辐射（光）通量，单位：瓦[特]（W）。

[来源：GB/T 8322-2008, 2.9, 有修改]

3.1.7

吸光度 absorbance

A

透射比（3.1.6）倒数的对数（以10为底）。

注：吸光度的计算公式为：

$$A = \lg\left(\frac{1}{\tau}\right)$$

[来源：GB/T 8322-2008, 2.10]

3.1.8

光路长度 optical path length

l

光通过吸收池内物质的入射面和出射面之间的路程。

注：吸收物质的折射率与光路长度的乘积为光程。

[来源：GB/T 14666-2025, 7.4.11, 有修改]

3.1.9

吸光系数 absorptivity

待测物质在单位浓度、单位厚度时的特征吸光度。

注：按照浓度单位不同，分为质量吸收系数（ ε ）和摩尔吸收系数（ a ），计算公式为：

$$\varepsilon = \frac{A}{L \times c}$$

$$a = \frac{A}{L \times \rho}$$

式中：

ε ——摩尔吸光系数，单位为升每厘米摩尔[L/(cm·mol)]；

L ——厚度，单位为厘米（cm）；

c ——吸光物质的量浓度，单位为摩尔每升(mol/L)；

a ——质量吸光系数，单位为升每厘米克[L/(cm·mol)]；

ρ ——质量浓度，单位为克每升(g/L)。

[来源：GB/T 14666-2025, 7.4.10, 有修改]

3.1.10

朗伯-波格定律 Lambert-bouguer's law

一束辐射（光）通量为 Φ_0 的平行单色辐射，垂直入射，通过吸收介质，若该吸收介质的表面是互为平行的平面，且它内部是各向同性的、均匀的、不发光、不散射的，则透射辐射（光）通 Φ_{tr} 随吸收介质的光路长度 l 的增加按指数减少。计算公式为：

$$\Phi_{tr} = \Phi_0 \cdot e^{-kl}$$

式中：

e ——自然对数的底；

k ——线性吸光系数，单位为每（ cm^{-1} ）。

[来源：GB/T 8322-2008, 4.1, 有修改]

3.1.11

比尔定律 Beer's law

一束平行单色辐射[光]，垂直入射，通过一定光路长度的吸收介质，它的透射辐射[光]通量随介质中吸收物质浓度的增加而按指数减少。由下列计算公式表示：

$$\Phi_{\text{tr}} = \Phi_0 e^{-k_m \rho}$$

$$\text{或 } \Phi_{\text{tr}} = \Phi_0 e^{-k_c c}$$

式中：

k_m ——质量线性吸收系数，单位为升每厘米克[$\text{L}/(\text{cm} \cdot \text{mol})$]

k_c ——摩尔线性吸收系数，单位为升每厘米摩尔[$\text{L}/(\text{cm} \cdot \text{mol})$]；

c ——物质的量浓度,单位为摩尔每升(mol/L)。

[来源：GB/T 14666-2025, 7.4.12]

3.1.12

朗伯-比尔定律 Lambert- Beer's law

将朗伯-波格（3.1.10）和比尔定律（3.1.11）两定律合并，由如下公式表示：

$$\Phi_{\text{tr}} = \Phi_0 10^{-a \cdot l \cdot \rho}$$

$$\text{或 } \Phi_{\text{tr}} = \Phi_0 10^{-\varepsilon \cdot l \cdot c}$$

[来源：GB/T 14666-2025, 7.4.14]

3.1.13

吸收 absorption

辐射能与物质作用时所发生的辐射能量减少并使物质内能增加的过程。

[来源：GB/T 8322-2008, 2.12]

3.1.14

散射 scattering

光线通过不均匀介质时，一部分光偏离原来的传播方向的现象。

3.1.15

干涉 interference

相干辐射的叠加使得辐射的振幅产生局部衰减或增强的现象。

3.1.16

衍射 diffraction

辐射通过障碍物边缘时出现的辐射传播方向偏离的现象。

3.1.17

反射 reflection

辐射光束投射到不同介质的分界面时，未改变辐射的单色成分的频率而使之被界面或介质折回的过程。

3.1.18

漫反射 diffuse reflectance

光经过样品表面或表面以下，向四周随机地散射的现象。

3.1.19

透射 transmission

保持波长不变地穿过介质的辐射现象。

[来源：GB/T 8322-2008, 2.8, 有修改]

3.1.20

漫透射 diffuse transmission

光进入含散射物质的样品后，经过多次散射和吸收作用后穿过样品的现象。

3.1.21

色散 dispersion

光在物质中的传播方向因波长不同而产生的变化。

[来源：GB/T 1466-2025, 7.1.36]

3.1.22

电子能级 electron energy level

原子或分子中电子所处的离散量子化能量状态，是分子光谱特别是振动能级和能量分布的基础。

3.1.23

振动能级 vibrational energy level

分子中原子核在平衡位置附近振动时，由量子化效应形成的离散能量状态。

注：其能量主要取决于分子的振动模式、键强度及参与原子的质量，近似可表示为：

$$E_v = h\nu \left(n + \frac{1}{2} \right)$$

式中：

E_v ——振动能级

h ——普朗克常数；

ν ——振动量子数；

n ——旋转量子数。

3.1.24

转动能级 rotational energy level

分子绕其质心旋转时，由量子化效应形成的离散能量状态。

注：其能量依赖于分子的转动惯量与转动量子数，近似可表示为：

$$E_J = J(J+1) \frac{h^2}{8\pi^2 I}$$

式中：

E_J ——转动能级；

J ——转动量子数；

h ——普朗克常；

I ——分子的转动惯量。

3.1.25

分子能级 molecular energy levels

分子内部能量状态的量子化分布。

注1：分子能级包括电子能级、振动能级和转动能级。

注2：分子能级可以表示为 $E = E_e + E_v + E_J$ ，其中 E_e 为电子能级， E_v 为振动能级， E_J 为转动能级。

3.1.26

偶极矩 dipole moment

衡量分子中正负电荷分布不对称程度的物理量，其大小由电荷量与电荷分布间距决定。

注：公式为： $\mu = q \cdot r$ ，其中 q 为电荷量， r 为与正、负电荷重心之间距离。偶极矩的存在是分子能否发生红外跃迁的必要条件之一。

3.1.27

简正振动 normal mode of vibration

多原子分子的一种基本振动形式。

注：多原子分子的所有振动运动都可以视为该分子各简正振动模式的叠加。对于含有 n 个原子的分子，其振动模式的数量为：线性分子是 $3n-5$ 种，非线性分子是 $3n-6$ 种。每个振动模式都具有特征频率。振动模式包括，对称伸缩振动；反对称伸缩振动；弯曲振动等。

3.1.28

基频 fundamental bands

分子吸收光子后从一个能级跃迁到相邻的高一能级产生的吸收峰。

3.1.29

倍频 overtone bands

分子吸收光子后，由基态直接跃迁至第二或更高振动激发态所产生的吸收带。

3.1.30

合频 combination bands

由两个或多个不同基频同时激发所产生的吸收带，其能量等于各参与振动模能量之和。

3.1.31

官能团 functional group

在不同的化合物中表现出相似的化学性质的一个或者一组原子。

注：官能团能够决定有机化合物类别的特征以及物理和化学性质

3.1.32

谱线轮廓 spectral line profile

发射辐射强度随波长变化的曲线图(发射线)或吸收率随波长变化的曲线(吸收线)。

[来源：GB/T 14666-2025 7.1.13]

3.1.33

光谱基线 spectral baseline

光谱信号中的背景信号基准线，对于吸收光谱而言，通常表示为零吸收的理论值。

3.2 近红外光谱分析方法原理术语

3.2.1

吸收光谱 absorption spectrum

物质吸收光子，从低能级跃迁到高能级而产生的光谱。

注：吸收光谱可以表示为 $A(\lambda)$ 。其符号 A 表示吸收光谱的强度，单位通常是无量纲；其符号 λ 表示光的波长，单位为nm。

3.2.2

二维相关光谱 two-dimensional correlation spectroscopy

对两个一维光谱进行互相关性运算，并将结果在二维平面展开的光谱处理方法。

3.2.3

自变量 independent variable

由研究者独立操纵以检验其对因变量产生的影响或响应的变量。在近红外光谱分析中，一般指直接测量的光谱数据，例如由多个样品的光谱构成的矩阵，通常使用 X 表示。

3.2.4

因变量 dependent variable

受到研究者操纵或观察以检验其是否会因自变量变化而产生响应或改变的变量。在近红外光谱分析中，一般为样品的物理或化学性质，或由多个样品的物理或化学性质构成的矩阵，通常使用 Y 表示。

3.2.5

单波长成像 single wavelength imaging

一种用近红外光谱成像中每个像素点光谱在指定波长处的吸光度/透过率值进行图像数据可视化的方法。

3.2.6

总吸收成像 total absorbance imaging

一种用近红外光谱成像中每个像素点光谱在所有波数/波长处的吸光度总和值进行图像数据可视化方法。

3.2.7

峰面积成像 peak area imaging

一种采用近红外光谱成像中各像素光谱在指定吸收峰处的积分面积值进行图像数据可视化的方法。

3.2.8

峰比值成像 band ratio imaging

一种采用近红外光谱成像中每个像素光谱在指定两个吸收峰的峰高值/积分面积值的比值进行图像数据可视化表达方式。

3.2.9

近红外一区 first region of near infrared

波长约为700-1000 nm的近红外光波段区域。

3.2.10

近红外二区 second region of near infrared

波长约为1000–1700 nm的近红外光波段区域。

3.2.11

红移 redshift

吸收峰或发射峰相对于参考状态向较长波长（或较低波数）方向移动的现象。

注：红移通常反映样品分子环境的变化，如氢键增强、分子极性变化、浓度效应或相互作用增强等，可用于表征化学键的弱化或能级间距的减小。

3.2.12

蓝移 blueshift

吸收峰或发射峰相对于参考状态向较短波长（或较高波数）方向移动的现象。

注：蓝移通常反映分子间相互作用减弱、键强度增强或环境极性减小等，可用于表征化学键的增强或能级间距的增大。

3.2.13

非线性光学 nonlinear optics

物质在高强度电磁辐射（如激光辐照）作用下，其极化响应不再与入射光场强度成线性关系，而表现出如二次谐波生成、三次谐波、光学参量放大等效应的物理现象。

3.2.14

光谱表征 spectral characterization

通过获取、处理与解释光谱数据，以分析样品的组成、结构或物理化学性质的过程。

4 近红外光谱分析仪器术语

4.1 近红光谱分析仪器原理术语

4.1.1

傅里叶变换光谱仪 Fourier transform spectrometer

利用傅里叶变换原理获得并记录光谱的仪器，通常由光源、干涉仪、采样附件、检测器和数据处理系统组成。

4.1.2

滤光片型光谱仪 filter-based spectrometer

利用带通滤光片、线性渐变滤光片或者法布里—珀罗干涉滤光器等滤光器件，获得并记录光谱的仪器，通常由光源，滤光器件，光电检测器组成。

4.1.3

光栅分光型光谱仪 grating spectrometer

采用光栅作为分光器件，并通过固定阵列检测器或者单点检测器与转动扫描方式实现光谱测量的光谱仪。通常由光源，光栅，狭缝，检测器组成。

4.1.4

声光可调谐型光谱仪 acousto-optic tunable filter spectrometer

基于声光可调谐滤光器件的光谱仪，通过声波频率调谐实现波长选择性透过，获得并记录光谱的仪器，通常由光源，声光可调滤波器，检测器组成。

4.1.5

阿达玛变换型光谱仪 Hadamard transform spectrometer

基于阿达玛变换原理，利用调制和解调实现光谱测量的光谱仪。

注：其工作原理是利用阿达玛矩阵的正交性，使用可编程光调制器（如光栅、光开关阵列等），根据阿达玛矩阵对输入光信号进行调制，将检测到的调制信号进行阿达玛逆变换，还原出各个波长通道的光谱信息。

4.1.6

数字微镜阵列型光谱仪 digital micromirror spectrometer

基于数字微镜阵列与分光元件，采用顺序扫描、间隔扫描，或者编码调制扫描，获得并记录光谱的仪器。

4.1.7

计算光谱仪 computational spectrometer

通过调制光信号和数学算法从少量测量中重建光谱的光谱仪。

注：其工作原理是通过调制和编码的方式，同时测量多个波长的混合信号，然后利用反演算法从混合信号中提取出各个波长的光谱信息。

4.1.8**干涉图 interferogram**

傅里叶变换光谱仪在一个干涉仪扫描周期内，探测器检测随时间变化的干涉信号所得到的图形。

注：横轴为光的光程差，纵轴为干涉光的强度。

4.1.9**切趾函数 apodization function**

傅里叶变换光谱仪中，为了减少因干涉仪的扫描距离有限而产生的复原光谱的畸变，对干涉图进行卷积计算的窗口函数。

4.1.10**充零因子 zero filling factor**

傅里叶变换光谱仪中，为了使光谱更加平滑，在干涉图末尾填充零值的比例。

4.1.11**多光谱成像 multispectral imaging**

利用滤光片或其他色散分光元件与检测器件组合，获取多个波段图像的光谱成像技术，

注：通常波段较少（一般为3-30个），分辨率较低，一般为 $\lambda/10$ 量级。

4.1.12**高光谱成像 hyperspectral imaging**

利用分光原件与检测器组合，获取高光谱分辨率的连续、窄波段的图像的光谱成像技术。

注：波段较多（一般为100-300个），分辨率较高，一般为 $\lambda/100$ 量级。

4.1.13**摆扫式光谱成像 whisk-broom hyperspectral imaging**

通过逐像素扫描和逐波长测量获取高光谱数据的光谱成像方法。

4.1.14**推扫式光谱成像 push-broom hyperspectral imaging**

采用逐行扫描方式，使用多通道探测器同时采集多个波长的光谱信息，获取高光谱数据的光谱成像方法。

4.1.15**凝视式光谱成像 snapshot hyperspectral imaging**

无需扫描，通过单次曝光方式同时获取目标空间信息和光谱信息的光谱成像方法。

4.1.16**化学成像 chemical imaging**

结合光谱学和成像技术，获得样品表面或内部的化学组成分布图像的方法。

4.1.17**智能光谱仪 intelligent spectrometer**

结合光谱分析技术与人工智能、物联网、自动化控制等现代信息技术的新型光谱仪器，能够自主完成光谱采集、数据处理、模型优化和结果输出的自动化分析。

注：智能光谱仪通过智能化手段提升光谱数据的采集速度、分析精度、自主决策能力和应用场景适应性，突破传统光谱仪依赖人工操作和离线分析的局限性。

4.2 近红外光谱分析仪器关键器件术语**4.2.1****氙灯 xenon lamp**

利用氙的放电现象制作的灯。

注：发出的光与太阳光很相似，为连续光谱，紫外线很强。

[来源：GB/T 8322-2008 5.1.2]

4.2.2

钨丝灯 tungsten filament lamp

用加热玻璃泡中的钨丝以产生可见和近红外线的辐射源

[来源: GB/T 8322-2008 5.1.3]

4.2.3

卤钨灯 tungsten halogen lamp

在钨丝灯中加入适量卤素或卤化物。

注: 卤钨灯的发光效率比钨丝灯高、寿命也长。

[来源: GB/T 8322-2008 5.1.4]

4.2.4

发光二极管 light emitting diode

一种能够将电能转化为光的半导体器件。

注: 在近红外光谱技术中, 也可以利用短波长发光二极管激发荧光粉产生近红外光。

4.2.5

可调谐激光器 tunable laser

输出波长可以在一定范围内连续或离散调节的激光器, 一般可采用电流调谐、温度调谐等方式。

4.2.6

滤光片 filter

对特定的光波段选择性地透过, 而对另一些波段的吸收或者截止的光学器件。根据通带特性可以分为高通、低通、带通、带阻。

4.2.7

线性渐变滤光片 linear variable filter

在特定方向上具有线性透射率或波长选择性变化的滤光片。

4.2.8

法布里-珀罗干涉滤光器 Fabry-perot filter

由两块平行的高反射率的反射镜组成的干涉型滤光器件, 通过调节两块反射镜间距选择性透过的特定波长。

4.2.9

声光可调谐滤光器 acousto-optic tunable filter

基于声光效应的可调谐滤光器, 能够通过声波频率的调节实现对入射光的波长选择性滤波。

注: 其工作原理是利用声波在晶体材料中传播时引起材料周期性折射率变换, 对通过晶体的光产生布拉格衍射, 通过压电换能器调节声波频率, 可以改变衍射角, 从而得到输出的单色光。

4.2.10

液晶可调谐滤光器 liquid crystal tunable filter

利用液晶材料的光学特性来实现调谐滤波的光学器件, 通过改变液晶材料的电场或温度选择性透过特定波段的光。

4.2.11

光栅 grating

基于对辐射的衍射作用, 使辐射产生色散的元件。

[来源: GB/T 8322-2008 5.2.3.2]

4.2.12

狭缝 slit

形状狭窄细长的缝孔, 一般用来调整光谱分辨率及进入光电检测器的信号强度。

4.2.13

干涉仪 interferometer

根据光的干涉原理制成的、通过测量光程差形成的干涉条纹而测定其他物理量的光学仪器。

注: 根据原理及组成, 分为迈克尔逊干涉仪、马赫-曾德尔干涉仪、法布里-珀罗干涉仪、菲索干涉仪、塞格纳克干涉仪、泰曼-格林干涉仪等。

[来源: GB/T 14666-2025 7.3.14, 有修改]

4.2.14

光阑 aperture

一种可以控制光束形状或者光通量的光学器件，分为可变光阑和固定光阑。

4.2.15

光电二极管 photodiode

基于半导体PN结的内光电效应，具有二极管结构，吸收光辐射而产生光电流的一种光电探测器件。

4.2.16

光电池 photovoltaic Cell

基于光伏效应，可以将光能直接转换为电能的光电探测器件。

注：当光照射到光电池表面时，光子能量被吸收，激发电子从价带跃迁到导带，形成电子-空穴对，从而在PN结两侧产生电势差。

4.2.17

光敏电阻 photoresistor

基于光电导效应，电阻值会随着光照强度变化而变化的光电探测器件。

注：当光照射到光敏电阻表面时，光子能量激发半导体中的电子跃迁，形成更多的自由电子和空穴，从而导致电阻值下降。

4.2.18

热释电探测器 pyroelectric detector

基于热释电效应，将红外辐射引起的温度变化转化为电信号输出的光电探测器件。

注：其核心材料是具有自发极化特性的热释电材料，如铌酸锶钡、锆钛酸铅、钽酸锂等。

4.2.19

光电倍增管 photomultiplier

内部有电子倍增机构，内增益极高，由光窗、光电阴极、电子聚焦系统、电子倍增系统和阳极5个部分组成。

[来源：GB/T 14666-2025 7.3.29]

4.2.20

线阵探测器 linear array detector

将光电检测单元沿直线规则排列，可以同时检测一维行信号的光电探测器件。

4.2.21

面阵探测器 area detector

将光电检测单元按照二维平面规则排列，可以同时检测二维平面信号的光电探测器件。

4.3 近红光谱分析仪器评价术语

4.3.1

光谱范围 spectral range

仪器能测量的光谱的波长范围。

注：主要取决于辐射源、波长选择器和检测器。

[来源：GB/T 14666-2025 7.1.11]

4.3.2

光谱分辨率 spectral resolution

仪器分辨相邻的两条谱线的能力。以谱线半峰宽来表示，决定了仪器对于光谱细节的识别能力。

4.3.3

半峰宽 full width at half maximum FWHM

光谱吸收峰高一半处的峰宽度。以光谱吸收峰到该点对应基线的计数差值的一半作平行于基线的直线，计算直线与谱峰两侧所形成交点之间的波数/波长差值。半峰宽计算示意图见图1。

注：半峰宽用于衡量光谱分辨率，也可称为半高全宽、半高宽。

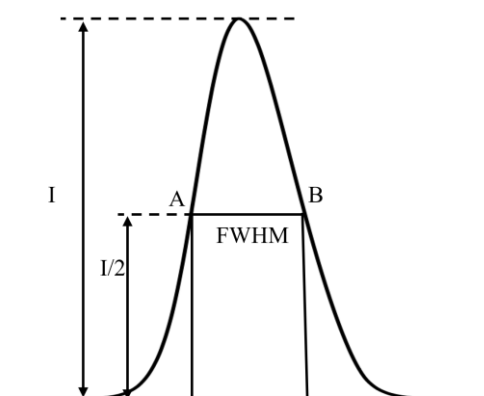


图2 FWHM 计算示意图

4.3.4

光谱采样间隔 spectral sampling interval

仪器测量得到的光谱图中相邻两个光谱采样点之间波数或波长的差值。

4.3.5

光谱检测灵敏度 spectral detection sensitivity

样品中被分析组分单位浓度改变引起检测光谱的变化量。

4.3.6

杂散光 stray light

光谱仪中不遵循仪器设计的光路到达探测器的光。

注：杂散光会干扰光谱测量。

4.3.7

暗光谱 dark spectrum

在无光条件下，光谱仪在给定参数下测量得到的信号，决定了仪器的最低检测限。

4.3.8

参考光谱 reference spectrum

用于校准、比较或作为基准的光谱。它通常代表某一标准物质、光源或理想条件下的光谱响应，用于辅助分析未知样品的光谱特性。

4.3.9

基线漂移 baseline drift

光谱基线随时间朝一定方向缓慢变化的现象

4.3.10

动态范围 dynamic range

光谱仪最大可检信号值与最小可检信号的比值，一般用以10为底的对数表示。

4.3.11

噪声等效功率 noise equivalent power

当光电探测器输出的信号与噪声的均方根相等时所对应的输入光功率。

4.3.12

光谱采集频率 sampling rate

光谱仪单位时间采集的完整光谱的数量。

4.3.13

积分时间 integrating time

曝光时间 exposure time

光电探测器对入射光子进行持续累计的时间。一般情况下，光谱强度与积分时间成正比，但是暗电流也会与积分时间成比例增加。

4.3.14

等吸收点 isoabsorptive point

指在某一波长下，两种或两种以上物质的吸收率相等。

[来源：GB/T 13966-2013 4.45]

4.3.15

100 %线 100% line

仪器连续测量两次背景光谱后，将二者相除得到的透射或反射光谱，理论上该光谱中每一点的透过率或反射率为100 %。

4.3.16

均方根噪声 root mean square noise

在100 %线上选定光谱范围内所有数据点的标准偏差，计算公式如下：

$$\text{RMSN} = \sqrt{\frac{\sum (y_i - \bar{y})^2}{n-1}}$$

式中：

RMSN——均方根噪声；

y_i ——所选光谱范围内第*i*个波数点数据所对应的透过率/反射率数据；

$\bar{y}$ ——所选光谱范围内透过率/反射率数据的平均值；

n ——所选光谱范围内的数据点数。

4.3.17

信噪比 signal-to-noise ratio

在特定的信号水平下，光谱信号与噪声的比值。

注：一般采用*n*次测量背景或者参考物光谱，计算每个波长下*n*次测量的平均值与标准差，用平均值除以标准差得到每个波长下的信噪比SNR(λ)，计算公式如下：

$$\text{SNR}(\lambda) = \frac{\overline{y(\lambda)}}{\sqrt{\frac{1}{n} \sum_{i=1}^n (y(\lambda)_i - \overline{y(\lambda)})^2}}$$

式中：

SNR(λ)—— λ 波长下的信噪比；

$y(\lambda)_i$ ——所选仪器在*i*次测量得到的 λ 波长处光谱强度；

$\overline{y(\lambda)}$ ——所选仪器*n*次测量得到 λ 波长处光谱强度的平均值；

n ——测量次数。

4.3.18

吸光度重复性 absorbance repeatability

在相同条件下，光谱仪器重复测量同一参考物得到的给定波数位置吸光度值的一致程度。

注：连续重复采集*n*次参考物光谱，记录参考物某一特征峰的透过率或反射率值，计算公式如下：

$$Y_T = Y_{\max} - Y_{\min}$$

式中：

Y_T ——仪器的透过率重复性；

$Y_{\max}$ ——参考物某一特征峰处透过率或反射率的最大值；

$Y_{\min}$ ——参考物某一特征峰处透过率或反射率的最小值。

4.3.19

波数准确度 wavenumber accuracy

光谱仪器测量得到的波长值与实际波长值间的一致程度。

注：采集标准物质光谱，连续重复测量3次，记录标准物质特征峰的实际峰位值，计算公式如下，取 $\Delta\nu$ 绝对值的最大值为衡量波数准确度的指标：

$$\Delta\nu = \nu_i - \nu$$

式中：

$\Delta\nu$ ——仪器的测量误差，单位为纳米每厘米 (nm/cm^{-1})；

ν_i ——第*i*次测量标准物质特征峰处的实际峰位值，单位为纳米每厘米 (nm/cm^{-1})；

ν ——标准物质特征峰的标准峰位值，单位为纳米每厘米 (nm/cm^{-1})。

4.3.20

波长重复性 wavenumber repeatability

在相同条件下，光谱仪器重复测量同一参考物得到的给定位置波长值的一致程度

注：采集标准物质光谱，连续重复测量*n*次，记录特征峰的波长值，计算公式如下。

$$\delta_\nu = \nu_{\max} - \nu_{\min}$$

式中：

δ_λ ——仪器的波长重复性；单位为纳米每厘米（nm/cm⁻¹）；

$v_{\max}$ ——特征峰处峰位值的最大值；单位为纳米每厘米（nm/cm⁻¹）；

$v_{\min}$ ——特征峰处峰位值的最小值；单位为纳米每厘米（nm/cm⁻¹）。

4.3.21

稳定性 stability

在正常工作条件下，仪器在规定时间内将吸光度测量值稳定在一个恒定值的能力。

注：采集标准物质吸光度光谱，记录某个特征峰的吸光度值。在一定时间内以固定间隔重复此过程，共记录n个吸光度值，采用吸光度值的相对标准偏差作为衡量稳定性的指标，计算公式如下：

$$RSD = \frac{1}{\bar{A}} \sqrt{\frac{\sum(A_i - \bar{A})^2}{n-1}}$$

式中：

RSD ——吸光度相对标准偏差；

$\bar{A}$ ——n次测量吸光度值的平均值；

A_i ——单次测量得到的吸光度值；

n ——测量次数。

4.4 近红外光谱分析仪器测量附件术语

4.4.1

标准物质 reference material

参考物质

具有足够均匀和稳定的特性，且其特性被证实适用于测量中或标称特性检查中的预期用途的物质。

[来源：GB/T 14666-2025 3.6]

4.4.2

衰减全反射 attenuated total reflection

在折射率不同的两种介质界面处，当光以大于临界角的入射角从折射率高的介质抵达界面时，在全反射过程中被吸收，使反射率小于 100% 的现象。

4.4.3

样品池 sample cell

盛放待测样品并允许光通过以进行测量的专用容器。

注：其核心功能是为样品提供可控的测量环境，确保准确获取光谱信号。

4.4.4

温控样品池 temperature controlled sample cell

能够在特定温度范围内精准调控样品温度的样品池。

4.4.5

多次反射样品池 multiple-reflectance cell

一种用于气体吸收光谱检测的附件，能够使光束在样品池内多次反射通过待测样品，从而达到增加光程，提高探测灵敏度的目的。

4.4.6

积分球 integrating sphere

内表面覆盖无波长选择性的漫反射材料的中空球体，收集不同角度的漫反射或者透射光，实现对光谱的均匀化收集与测量。

4.4.7

光纤探头 fiber probe

利用光纤将光传输至样品，收集样品出射光，并传输至光谱仪的测量附件。

4.4.8

背景光谱 background spectrum

仪器正常工作条件下测量背景得到的光谱。

注：对于透射测样方式，背景光谱通常是在未放置样品时测量得到的；对于反射测样方式，背景光谱通常是将反射背景板放置到样品测量窗口位置或光纤探头接触面上时测量得到的；对于透反射测样方式，背景光谱通常是在

光纤探头与背景反射面间隔一定距离的空气时测量得到的。

5 近红外光谱分析方法术语

5.1 近红外光谱分析方法基本术语

5.1.1

校正 calibration

将一组样品的成分浓度或性质与其光谱进行关联，建立模型的过程。

[来源：GB/T 29858-2013 3.1]

5.1.2

校正模型 calibration model

表达一组样品的成分浓度或性质与其光谱之间关联关系的数学表达式。

[来源：GB/T 29858-2013 3.2]

5.1.3

校正集 calibration set

用来建立校正模型的一组具有代表性的样品，其成分浓度或者性质数据已知。

5.1.4

验证集 validation set

用于验证模型的准确性、重复性和传递性等性能的一组样品，其成分浓度或者性质数据已知。验证集中样品的成分浓度或者性质范围至少覆盖校正集成分浓度或者性质数据的95%，且分布均匀。

5.1.5

预测 prediction

用校正模型和样品的光谱计算样品成分浓度或性质的过程。

[来源：GB/T 29858-2013 3.4]

5.1.6

预测值 predicted values

使用校正模型和样品光谱计算得到的样品成分浓度或性质数据。

[来源：GB/T 29858-2013 3.5]

5.1.7

模型验证 model validation

用验证样品来检验校正模型预测值与参考值之间一致性的过程。

[来源：GB/T 29858-2013 3.10]

5.1.8

交叉验证 cross-validation

通过系统性划分数据集为训练集和验证集，并综合评估校正模型最佳参数及性能的方法。

注：常用方法包括K折交叉验证与留一法交叉验证。

5.1.9

K折交叉验证 K-fold cross validation

将数据集随机划分为K个大小相近的子集，在第i次迭代中，将第i个子集作为测试集，其余K-1个子集合并为训练集，进行K次训练与测试，综合K次结果确定校正模型最佳参数及性能。

5.1.10

留一法交叉验证 leave-one-out cross validation

对校正集中每个样品进行独立验证，假设数据集包含N个样品，在第i次迭代中，将第i个样品作为测试集，其余N-1个样品作为训练集，进行N次训练与测试，综合N次结果确定校正模型最佳参数及性能。

5.1.11

模型传递 calibration model transfer

通过控制制造精密度使不同光谱仪器间保持一致性或通过数学方法对光谱或校正模型进行修正，使得在一台光谱仪器上建立的校正模型能够在其他光谱仪器上有效使用的过程。

注：也称为模型转移

[来源: GB/T 29858-2013 3.19]

5.1.12

主机 master

模型传递中, 将建立校正模型所使用光谱的测量仪器 (或测量条件) 称为主机。

5.1.13

从机 slave

模型传递中, 将待转移光谱的测量仪器 (或测量条件) 称为从机。

5.1.14

定量分析 quantitative analysis

测定试样中各种成分 (如元素、根或官能团等) 的含量而进行的分析。

[来源: GB/T 13996-2013 2.3]

5.1.15

定性分析 qualitative analysis

检测试样中的元素、官能团或混合物组成成分而进行的分析。

注: 定性分析的目的是得到样品的特征或归类信息。

[来源: GB/T 13996-2013 2.2, 有修改]

5.1.16

正类 (阳性) positive class

对于一个二分类来说, 样品会被分为正类和负类, 正类 (阳性) 为目标类别、或感兴趣类别, 具有正类属性的样本为正例。

5.1.17

负类 (阴性) negative class

对于一个二分类来说, 样品会被分为正类和负类, 与正类不同的另一类为负类。具有负类属性的样本为负例。

5.1.18

真正例 true positive, TP

被校正模型预测为正类的正样本, 真阳性样本。

5.1.19

假正例 false positive, FP

被校正模型预测为正类的负样本, 假阳性样本。

5.1.20

真负例 true negative, TN

被校正模型预测为负类的负样本, 真阴性样本。

5.1.21

假负例 false negative, FN

被校正模型预测为负类的正样本, 假阴性样本。

5.1.22

欠拟合 underfitting

校正模型在训练阶段未能充分提取校正集数据的真实特征或模式的现象。

5.1.23

过拟合 overfitting

校正模型在训练阶段提取了校正集数据中的非代表性特征的现象。

5.2 近红外光谱分析方法预处理术语

5.2.1

光谱预处理 spectral pretreatment

在模型建立前对光谱进行处理的过程, 目的是降噪或增强光谱特征。

注: 常用数据预处理方法包括: 平滑、导数、散射校正、中心化和标准化等。

5.2.2

平滑 smoothing

通过对原始光谱数据进行数学运算，使光谱曲线更加光滑，从而消除光谱中高频噪声，并提高信噪比的光谱预处理方法。

注：常用平滑方法包括移动窗口平均法、Savitzky-Golay卷积法等。

5.2.3**归一化 normalization**

将光谱数据按比例缩放，调整到[0, 1]或[-1, 1]等标准区间内，以消除不同样品或光谱仪器之间的光谱差异的光谱预处理方法。

注：常用的方法有最大最小归一化、矢量归一化等。最大最小归一化计算公式如下：

$$x'_i = \frac{x_i - x_{i,min}}{x_{i,max} - x_{i,min}}$$

式中：

x'_i —— x_i 使用最大最小归一化处理后的光谱数据；

x_i —— n 个样本光谱数据集中的光谱数据， $i = 1, 2, \dots, n$ ；

$x_{i,min}$ —— x_i 中第 i 个样品的最小吸光度；

$x_{i,max}$ —— x_i 中第 i 个样品的最大吸光度。

矢量归一化计算公式如下：

$$x'_i = \frac{x_i - \bar{x}_i}{\sqrt{\sum_{j=1}^m (x_{i,j})^2}}$$

式中：

x'_i —— x_i 使用矢量归一化处理后的光谱数据；

x_i —— n 个样本光谱数据集中的光谱数据， $i = 1, 2, \dots, n$ ；

$\bar{x}_i$ —— x_i 中第 i 个样品的所有 m 个波长点下吸光度的平均值；

$x_{i,j}$ —— x_i 中第 i 个样品的第 j 个波长点的光谱数据， $j = 1, 2, \dots, m$ 。

5.2.4**小波变换 wavelet transform**

将光谱分解成一系列小波基函数的线性组合的光谱处理方法。

注：小波基函数可以由一个母小波函数经过平移和尺度伸缩得到，在光谱分析中主要用于去噪、特征提取、基线校正等任务。

5.2.5**基线校正 baseline correction**

利用数学方法消除基线漂移的光谱预处理方法，用于消除光谱数据中由于仪器漂移或样品的非光学效应引起的基线偏移，以便更准确地获取光谱特征。

5.2.6**去趋势 detrending**

一种基线校正方法，对光谱的吸光度和波长进行多项式拟合，得到趋势线，然后从原始光谱中减去趋势线，从而消除基线漂移的光谱预处理方法。

注：采用二项式拟合的去趋势计算公式如下：

$$\begin{aligned} x_d(i) &= b_0(i) + b_1(i)x(i) + b_2(i)x(i) \\ x'(i) &= x(i) - x_d(i) \end{aligned}$$

式中：

$x_d(i)$ ——光谱数据二项式拟合的趋势线 x_d 中第 i 个波长点的值；

$x(i)$ ——原始光谱数据 x 第 i 个波长点的值；

$x'(i)$ —— 经过去趋势处理后的光谱数据 x' 中第 i 个波长点的值；

b_0, b_1, b_2 ——拟合系数。

5.2.7

导数 derivatives

将光谱吸光度数据对波长（或波数）变量进行求导运算，用来消除光谱数据的基线漂移。

注：常用的光谱求导数方法有数值微分法（如有限差分法）和 Savitzky-Golay 求导法。有限差分法一阶导数计算公式如下：

$$x'_{1st,i} = x_{s,i+g_d} - x_{s,i-g_d}$$

差分法二阶导数计算公式如下：

$$x'_{2nd,i} = x_{s,i-g_d} - 2x_{s,i} + x_{s,i+g_d}$$

式中：

x_i —— n 个样本光谱数据集中的光谱数据， $i = 1, 2, \dots, n$ ；

$x_{s,i}$ —— 第 i 个样本光谱数据平滑处理后的数据， $i = 1, 2, \dots, n$ ；

$x_{i,j}$ —— 第 i 个光谱数据的第 j 个波长点的吸光度， $j = 1, 2, \dots, m$ ；

g_d —— 差分间隔大小；

$x'_{1st,i}$ —— x_i 一阶导数处理后的样本光谱数据；

$x'_{2nd,i}$ —— x_i 二阶导数处理后的样本光谱数据；

5.2.8

多元散射校正 multiplicative scatter correction

用于消除样品颗粒大小分布不均匀或者厚度变化产生的散射影响的光谱预处理方法。

注：假设每条光谱都与平均光谱线性相关，即样品集中每条光谱的任意波长点吸光度与所在样品集平均光谱的吸光度值呈近似线性关系，通过线性回归方法校正原始光谱，计算公式如下：

$$\begin{aligned}\bar{x} &= \sum_{i=1}^n \frac{x_i}{n} \\ x_i &= \alpha_i \bar{x} + \beta_i \\ x'_i &= \frac{x_i - \beta_i}{\alpha_i}\end{aligned}$$

式中：

$\bar{x}$ —— n 个样本光谱数据集的平均光谱；

x_i —— n 个样本光谱数据集中的光谱数据， $i = 1, 2, \dots, n$ ；

α_i —— 将 x_i 与 $\bar{x}$ 进行线性回归的斜率；

β_i —— 将 x_i 与 $\bar{x}$ 进行线性回归的截距；

x'_i —— x_i 使用 MSC 处理后的光谱数据。

5.2.9

标准正态变换 standard normal variate SNV

用于消除光谱基线偏移和强度差异的光谱预处理方法。

注：假设每条光谱各波长点的吸光度值应满足一定的分布（如正态分布），通过将每个样本的光谱标准化为均值为 0、标准差为 1 的分布，从而消除样品间的整体强度差异，保留光谱的形状信息，计算公式如下：

$$\begin{aligned}\bar{x}_i &= \frac{\sum_{j=1}^m x_{i,j}}{m} \\ x'_i &= \frac{x_i - \bar{x}_i}{\sqrt{\frac{\sum_{j=1}^m (x_{i,j} - \bar{x}_i)^2}{m-1}}}\end{aligned}$$

式中：

x'_i —— x_i 经过 SNV 处理后的光谱数据；

x_i —— n 个样本光谱数据集中的光谱数据， $i = 1, 2, \dots, n$ ；

$\bar{x}_i$ —— x_i 中第 i 个样品的所有 m 个波长点下吸光度的平均值；

$x_{i,j}$ —— x_i 中第 i 个样品的第 j 个波长点的吸光度， $j = 1, 2, \dots, m$ 。

5.2.10

均值中心化 mean centering

用于消除光谱全局偏移，突出样品光谱之间的差异的预处理方法，计算公式如下。

$$x'_i = x_i - \bar{x}$$

式中：

x'_i —— x_i 使用均值中心化处理后的光谱数据；

x_i —— n 个样本光谱数据集中的光谱数据， $i = 1, 2, \dots, n$ ；

$\bar{x}$ —— n 个样本光谱数据集的平均光谱，计算公式为 $\bar{x} = \sum_{i=1}^n \frac{x_i}{n}$ 。

5.3 近红外光谱分析方法建模术语

5.3.1

参考方法 reference method

用于测定校正样品和验证样品成分浓度或性质数据的分析方法，通常为标准方法或是推荐认可的方法。

[来源：GB/T 29858-2013 3.8]

5.3.2

参考值 reference values

用参考方法测定校正样品或验证样品的成分浓度或性质结果。

[来源：GB/T 29858-2013 3.9]

5.3.3

异常值 outliers

因含有极端组成，致使其光谱不具代表性，在建模过程中表现杠杆值较大，对模型结果有强烈影响的高杠杆值样品对应的数据点，或因数据测定误差较大或数据录入结果错误等原因，模型预测值与参考值具有显著性差异的样品对应的数据点。

注：在近红外光谱分析中异常值包含光谱数据异常值（即：x-异常值），或参考值异常（即y-异常值）。

[来源：GB/T 29858-2013 3.11]

5.3.4

杠杆值 leverage value

杠杆值是样品点与模型所定义群体空间中心点距离的测量值。

注：在回归分析中，高杠杆值的观测点通常远离数据中心，对模型的影响较大，可能导致模型不稳定或偏差较大。

5.3.5

波长选择 wavelength selection

在近红外光谱中，通过评估不同波长对目标结果（如成分浓度、分类标签）的贡献度或冗余性，筛选关键波长或波段的过程。

注：该过程旨在提高模型效率、减少噪声干扰或增强解释性，常用方法包括统计筛选（如相关系数法）、智能算法（如遗传算法）及模型嵌入特征重要性评估等。

5.3.6

多元线性回归 multivariate linear regression

通过建立回归方程，利用数理统计原理研究因变量与多个自变量之间线性相关关系的一种方法。

5.3.7

主成分回归 principal component regression

在多元线性回归分析中，先对光谱数据矩阵X进行正交分解得到主成分，用主成分代替原有变量与浓度矩阵Y进行回归分析的方法。

5.3.8

有监督模式识别方法 supervised pattern recognition method

在已知样品的监督下进行的模式识别方法

[来源：GB/T 37969-2019 3.10]

5.3.9

无监督模式识别方法 unsupervised pattern recognition method

在没有先验知识情况下进行的模式识别方法

[来源：GB/T 37969-2019 3.11]

5.3.10

聚类分析 cluster analysis

将具有相似信息或数据特征的对象聚在一起的一种数据分析方法。

注：聚在一起的研究对象称为一个类。常用方法有层次聚类法、SIMCA、K-Means 算法等。

5.3.11

线性判别分析 linear discriminant analysis

通过特征的线性组合来区分或描述两类或多类对象。

注：其核心思想是最大化类别之间的差异，同时最小化同一类别内部的差异，从而实现分类或降维的目的，具有更好的样本区分能力。

5.3.12

软独立建模分类 soft independent modelling of class analogy

对校正集中各类样本的光谱进行主成分分析，建立各类的主成分分析校正模型，每一个类的模型可以选取适宜的主成分数。计算待测样本在各类模型的主成分空间的得分及组合残差，通过预设的组合残差距离的阈值判别样本类别。

5.3.13

偏最小二乘判别分析 partial least squares-discriminant analysis

一种基于偏最小二乘回归分析的有监督模式识别分类方法。

注：利用先验分类知识，将类别作为分类变量（因变量）量化，例如本类取值1，非本类取值0的量化方法。然后将光谱变量与分类变量进行PLS校正，建立PLS-DA类模型，最后利用类模型判别待测样品的归属。

5.3.14

模型内插 model interpolation

对成分浓度或性质数据处于校正模型涵盖范围内的样品进行预测的过程。

[来源：GB/T 29858-2013 3.13]

5.3.15

模型外推 model extrapolation

对成分浓度或者性质数据处于校正模型涵盖范围外的样品进行预测的过程。

注：包括两种模型外推：

——待测样品的成分与校正集样品相同，但浓度超过了校正集涵盖的浓度范围；

——待测样品包含校正样品中不存在的成分。

[来源：GB/T 29858-2013 3.14]

5.3.16

机器学习 machine learning

人工智能的一个子领域，通过分析大量数据，从中发现隐藏的规律或模式，通过自动化模型构建，使算法能够适应复杂、动态的任务。机器学习主要分为监督学习、无监督学习和强化学习等范式，涵盖人工神经网络、支持向量机、决策树等多种方法。

5.3.17

深度学习神经网络 deep neural networks

一种基于多层人工神经网络的机器学习方法，属于人工神经网络的扩展。其核心在于通过多层非线性变换自动从数据中提取高层次特征，从而解决复杂任务。

5.3.18

观测变量 observed variable

指在研究或模型中可以直接测量、记录或获取的变量，是数据分析与模型构建的基础。

5.3.19

隐变量 latent variable

指未被直接观测但假设存在的变量，用于解释观测数据背后的隐含结构或生成机制，通过模型推断揭示数据间的复杂关系。

5.3.20

正则化 regularization

机器学习中用于防止模型过拟合的关键技术，通过在损失函数中添加惩罚项来限制模型复杂度，从而提升模型的泛化能力

5.4 近红外光谱模型评价术语

5.4.1

残差 residual

样品的参考值和预测值之间的差。残差计算公式如下：

$$e_i = y_i - \hat{y}_i$$

式中：

- e_i —— 第 i 个样品的残差；
- y_i —— 第 i 个样品的参考值；
- $\hat{y}_i$ —— 第 i 个预测值。

5.4.2

偏差 deviation

样品集中样品残差的平均值，用于衡量模型预测值整体偏离真实值的程度，计算公式如下：

$$\bar{e} = \frac{1}{n} \sum_{i=1}^n e_i$$

式中：

- $\bar{e}$ —— 偏差；
- e_i —— 第 i 个样品的残差；
- n —— 样本数量。

5.4.3

校正标准误差 standard error of calibration

在多元校正中用来评价校正模型预测能力，采用校正样品参考值与预测值计算的标准误差。

注：计算公式如下：

$$SEC = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - 1}}$$

式中：

- SEC —— 校正标准误差；
- y_i —— 校正集第 i 个样品参考值；
- $\hat{y}_i$ —— 校正集第 i 个样品近红外预测值；
- n —— 校正样本数量。

5.4.4

交互验证标准误差 standard error of cross validation

在多元校正中用来评价校正模型预测能力，采用校正样品参考值和使用交互验证的预测值计算的标准误差。

注：计算公式如下：

$$SECV = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-1}}$$

式中：

$SECV$ —— 交互验证标准误差；

y_i —— 校正集第 i 个样品参考值；

$\hat{y}_i$ —— 校正集交互验证所得第 i 个样品近红外预测值

n —— 校正集样本量。

[来源：GB/T 29858-2013 3.16，有修改]

5.4.5

验证标准误差 standard error of prediction

在多元校正中用来评价校正模型预测能力，采用验证样品参考值和预测值计算的标准误差。

注：计算公式如下

$$SEP = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}}$$

式中：

SEP —— 验证标准误差；

y_i —— 验证集第 i 个样品参考值；

$\hat{y}_i$ —— 验证集交互验证所得第 i 个样品近红外预测值

n —— 验证集样本量。

[来源：GB/T 29858-2013 3.17，有修改]

5.4.6

预测残差平方和 predictive residual error sum of square

在多元校正中用来选择建模最佳主成分数或潜变量数，采用校正样品参考值和使用交互验证的预测值计算的残差平方和。

[来源：GB/T 29858-2013 3.18]

5.4.7

相关系数 correlation coefficient

衡量两个变量之间线性或非线性的关联程度的统计量，其取值范围通常为 $[-1, 1]$ 。绝对值越接近1表示相关性越强，正负号表示相关方向。

注：常见的Pearson相关系数通常用 r 表示，其计算公式如下：

$$r = \frac{\sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2} \sqrt{\sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2}}$$

式中：

r —— 相关系数

y_i —— 验证集第 i 个样品参考值；

$\hat{y}_i$ —— 验证集第 i 个样品预测值；

$\bar{y}$ —— 验证集样品参考值的平均值；

$\bar{\hat{y}}$ —— 验证集样品近红外预测值的平均值；

n —— 验证集样本量。

5.4.8

欧式距离 Euclidean distance

在光谱分析中，用来衡量两条光谱数据之间的相似程度或差异程度的指标。计算公式如下：

$$D_E(x, y) = \sqrt{\sum_{i=1}^n (x(i) - y(i))^2}$$

式中：

$D_E(x, y)$ —— 两条光谱的欧氏距离

x, y —— 两条光谱数据；

$x(i), y(i)$ —— 两条光谱中第*i*个波长点的吸光度值；

n —— 光谱数据的波长点数。

5.4.9

马氏距离 Mahalanobis distance

一种考虑了各个特征之间相关性的距离度量方式，用来衡量异常样品、评估模型预测值的不确定性。

计算公式如下：

$$D_M(x) = \sqrt{(x - \mu)^T S^{-1} (x - \mu)}$$

式中：

$D_M(x)$ —— 样品光谱与平均光谱的马氏距离

x —— 待测样本光谱；

μ —— 样品集平均光谱；

S^{-1} —— 样品集协方差矩阵的逆矩阵。

注：杠杆值和马氏距离都是用于检测数据异常点和影响点的统计量，但侧重不同：杠杆值主要用于回归分析，衡量单个样本对模型拟合的影响程度；而马氏距离则适用于多元数据，衡量样本与数据整体分布的偏离程度。

5.4.10

决定系数 determination coefficient

校正模型可解释的变异占总变异的比重，用于评价模型拟合程度的指标，是相关系数 r 的平方。计

算公式如下：

$$R^2 = 1 - \frac{\sum_{i=1}^n \frac{(\hat{y}_i - y_i)^2}{n}}{\sum_{i=1}^n \frac{(y_i - \bar{y})^2}{n}}$$

式中：

y_i —— 验证集第 *i* 个样品参考值；

$\hat{y}_i$ —— 验证集第 *i* 个样品预测值；

$\bar{y}$ —— 验证集样品参考值的平均值；

n —— 验证集样本数量。

5.4.11

相对分析误差 relative performance deviation RPD

验证样品参考值标准差与预测值标准差计算的比值，在多元校正中用来评价校正模型预测能力，计

算公式如下：

$$RPD = \frac{\sqrt{\sum_{i=1}^n \frac{(y_i - \bar{y})^2}{n}}}{\sqrt{\sum_{i=1}^n \frac{(\hat{y}_i - y_i)^2}{n}}}$$

式中：

y_i —— 验证集第 *i* 个样品参考值；

$\hat{y}_i$ —— 验证集第 *i* 个样品预测值；

$\bar{y}$ —— 验证集样品参考值的平均值；

n —— 验证集样本量。

5.4.12

准确率 accuracy rate ACC

在二分类问题中，判别模型在测试集上判别出的类别与目标类别（真正类或真负类）一致的样本数量与测试集总样本数量的比值，计算公式如下：

$$ACC = \frac{TP + TN}{TP + TN + FP + FN}$$

式中：

TP —— 真正例数量；

TN —— 真负例数量；

FP —— 假正例数量；

FN —— 假负例数量。

5.4.13

错误率 error rate ER

在二分类问题中，判别模型在测试集上判别出的类别与目标类别（真正类或真负类）不一致的样本数量与测试集总样本数量的比值，计算公式如下：

$$ER = \frac{FP + FN}{TP + TN + FP + FN}$$

式中：

TP —— 真正例数量；

TN —— 真负例数量；

FP —— 假正例数量；

FN —— 假负例数量。

5.4.14

真正率 true positive rate TPR

灵敏度 sensitivity

召回率 recall

在二分类问题中，被判别模型判别为真的样本数量与实际真样本总数之间的比值，计算公式如下：

$$TPR = \frac{TP}{TP + FN}$$

式中：

TP —— 真正例数量；

FN —— 假负例数量。

5.4.15

假正率 false positive rate FPR

误报率 false alarm rate FAR

在二分类问题中，被判别模型错误地判定为真的样本数量与实际假样本总数之间的比值，计算公式如下：

$$FPR = \frac{FP}{TN + FP}$$

式中：

TN —— 真负例数量；

FP —— 假正例数量。

5.4.16

真负率 true negative rate TNR

特异度 specificity

在二分类问题中，被判别模型判别为假的样本数量与实际假样本总数之间的比值，计算公式如下：

$$TNR = \frac{TN}{TN + FP}$$

式中：

TN —— 真负例数量；

FP —— 假正例数量。

5.4.17

假负率 false negative rate FNR

漏诊率 missed diagnosis rate

在二分类问题中，被判别模型错误地判定为假的样本数量与实际真样本总数之间的比值，计算公式如下：

$$FNR = \frac{FN}{TP + FN}$$

式中：

FN —— 假负例数量；

TP —— 真正例数量。

5.4.18

精确度 precision

真阳性样品在所有预测为阳性样本中的比例，计算公式如下：

$$P = \frac{TP}{TP + FP}$$

式中：

P —— 精确度

TP —— 真正例数量；

FP —— 假正例数量。

5.4.19

F 度量 F-measure F1

结合精确度和查全率的调和平均值，是评估分类模型的性能的综合指标，计算公式如下：

$$F1 = \frac{P \times TPR}{P + TPR} = \frac{TP + TP}{TP + FP + FP + FN} = \frac{2TP}{2TP + FP + FN}$$

式中：

TP —— 真正例数量；

TN —— 真负例数量；

FP —— 假正例数量；

FN —— 假负例数量。

5.4.20

受试者工作特征曲线 receiver operating characteristic ROC

根据不同阈值的分类器输出绘制的曲线，可考察指标量在大范围取值之间的变化，并全局化地考察模型的优势和不足。

注：ROC曲线又可分为灵敏度-特异性ROC曲线和精确度-召回率ROC曲线。

5.4.21

灵敏度-特异性 ROC 曲线 sensitivity-specificity ROC

纵坐标表征灵敏度（真正率），横坐标表示为特异性（负真率）的ROC曲线，曲线上的点表示不同负真率阈值上的真正率比例。

5.4.22

精确度-召回率 ROC 曲线 precision-recall ROC

纵坐标表征精确度真正率，横坐标表示为召回率的ROC曲线，曲线上的点表示不同精确度阈值上的召回率比例。