

环境监测中地下水数据的异常值检测与处理

范子贤

宁夏测衡联合实业有限公司 宁夏 银川 750000

摘要：随着环境保护意识的提升，地下水环境监测成为确保水资源安全的重要手段。在地下水监测过程中，异常值的检测与处理是至关重要的环节，它直接关系到监测数据的准确性和可靠性。本文旨在探讨环境监测中地下水数据异常值的检测方法与管理策略，以期对地下水资源的科学管理与保护提供技术支持。

关键词：地下水监测；异常值检测；数据处理

引言

地下水作为重要的水资源之一，其质量状况直接关系到人类生活与生产活动的可持续发展。因此，对地下水环境进行长期、连续的监测，及时发现并处理数据中的异常值，对于评估地下水质量、预防污染以及指导水资源合理利用具有重要意义。

1 地下水监测的重要性

地下水监测不仅关乎水资源的有效配置和合理使用，还与环境保护、地质安全、社会经济发展以及公众健康息息相关。通过持续监测，可以掌握地下水位的动态变化、水质状况以及补给量等信息，为水资源管理部门提供决策依据。

2 地下水数据异常值的成因

地下水数据异常值的成因复杂多样，每一种成因都可能导致监测数据的失真或偏差。首先，采样误差是一个重要原因。在地下水采样过程中，如果采样器具不洁、采样方法不当或采样点选择不合理，都可能导致采集到的水样与实际地下水状况存在差异，从而产生异常值。其次，仪器故障也是导致数据异常的一个常见因素。监测仪器长时间运行后，可能会出现老化、磨损或校准失准等问题，这些都会直接影响到监测数据的准确性。例如，电导率仪的电极如果受到污染，就会导致电导率数据异常。再者，人为操作失误也是一个不可忽视的原因。操作人员在记录、计算或数据传输过程中，稍有不慎就可能引入错误，如记录错误、计算失误或数据传输错误等，这些都会导致数据出现异常^[1]。最后，自然环境的突变，如地下水位急剧变化、水质突然恶化等，也可能导致监测数据出现异常。这类异常值往往反映了地下水环境的真实变化，但也需要与其他类型的异常值进行区分。这些异常值若不及时发现和um处理，将会混入正常数据中，导致数据分析结果失真，进而影响对地下水环境状况的准确评估，甚至可能误导水资源管理和环

境保护的决策。

3 地下水数据异常值的检测方法

3.1 统计方法

统计方法是基于统计学原理，通过计算和分析数据的统计特征，如均值、标准差、偏度、峰度等，来识别那些与整体数据分布规律明显不符的异常数据点。在应用统计方法进行异常值检测时，首先要对地下水监测数据进行整理，确保数据的完整性和一致性。随后，通过计算数据的均值和标准差，可以了解数据的集中趋势和离散程度。均值反映了数据的平均水平，而标准差则揭示了数据波动的幅度。在得到这些统计量后，可以设定合理的阈值或范围，来判断哪些数据点超出了正常波动的范畴。例如，可以采用“3-sigma”原则，即认为落在均值加减三倍标准差之外的数据为异常值。这种方法简单直观，但在数据分布不均或存在多个峰值时可能效果不佳。除了均值和标准差，还可以利用更复杂的统计量，如偏度和峰度，来刻画数据的分布形态。偏度反映了数据分布的对称性，而峰度则描述了数据分布的尖锐程度。这些统计量的异常值也可能指示出数据的异常情况。此外，还可以利用箱线图（Box Plot）等可视化工具来辅助识别异常值。箱线图通过展示数据的中位数、四分位数和极端值，能够直观地展示出数据的分布情况，从而帮助识别出那些远离主体数据群的异常点。需要注意的是，统计方法虽然强大，但也有其局限性。它对于数据的分布形态有一定的假设，如正态分布等。当实际数据分布与假设不符时，统计方法的准确性可能会受到影响。因此，在应用统计方法进行异常值检测时，应结合数据的实际情况和领域知识，谨慎选择合适的统计量和阈值。

3.2 物理化学方法

物理化学方法的核心是根据地下水中溶解的物质、离子、分子等的物理化学性质，来判断所获得的数据是

否在合理的范围内。首先,地下水中各种化学成分都有其特定的物理化学属性,如溶解度、电离常数等。通过测定这些成分的含量,可以推算出水样的某些物理性质,如电导率、pH值等。当实测值与理论推算值出现较大偏差时,就可能意味着数据存在异常。其次,通过对比历史数据,可以发现当前数据是否与历史趋势相符。例如,某一监测点的地下水中铁离子含量历来都较低,如果突然出现大幅上升,那么这个数据点很可能就是异常值。再者,相邻监测点的数据对比也是一个有效的方法。由于地下水体的连续性,相邻监测点的数据通常具有一定的相关性。如果某一监测点的数据与其他相邻点的数据存在显著差异,且无合理的地质或环境因素可以解释这种差异,那么该数据点可能被视为异常。最后,还可以利用特定的化学试剂或仪器对水样进行定性或定量分析,进一步验证数据的合理性。例如,通过滴定法测定水样中某种离子的浓度,与自动监测仪器的读数进行对比,从而判断数据是否准确。

3.3 时间序列分析

时间序列分析侧重于分析数据随时间变化的规律和趋势。地下水监测数据通常具有时间序列特性,即数据是按照时间顺序采集的,因此,利用时间序列分析可以深入探索数据的动态特征。在进行时间序列分析时,首先要对数据进行预处理,包括数据的清洗、插补和平滑处理,以确保数据的质量和连续性。随后,可以利用各种时间序列模型,如自回归模型(AR)、移动平均模型(MA)、自回归移动平均模型(ARMA)或自回归整合移动平均模型(ARIMA)等,来拟合历史数据并预测未来趋势。在模型拟合的基础上,通过分析残差(实际观测值与模型预测值之差),可以识别出那些与模型预测结果显著偏离的数据点。这些偏离点很可能就是异常值,因为它们不符合正常的的数据变化趋势。此外,时间序列分析还可以通过计算数据的变化率或增长率来识别异常^[2]。例如,如果某个时间段内数据的变化率突然增大或减小,且这种变化不能由已知的地质、气象或人为因素解释,那么这些数据点就可能被视为异常。除了上述方法,还可以利用时间序列的周期性、季节性等特征来检测异常。地下水位和水质数据往往受到季节变化、降雨等因素的影响,表现出一定的周期性。通过分析这些周期性特征,可以识别出那些不符合周期性变化规律的数据点。

4 地下水数据异常值的处理策略

4.1 剔除法

在地下水数据处理过程中,剔除法是一种直接且有

效的异常值处理策略。当某个数据点明显偏离正常范围,且这种偏离无法由已知的地质、环境或监测方法等因素合理解释时,该数据点很可能被视为异常值。在这种情况下,采取剔除法是一种可行的处理方式。然而,使用剔除法并非简单的删除操作,而是需要严谨的判断和记录。首先,必须确保被剔除的数据点确实是异常值,这需要基于前面提到的异常值检测方法,如统计方法、物理化学方法或时间序列分析等,进行综合判断。只有经过这些方法的验证,并确认数据点确实存在异常,才能进行剔除。其次,在剔除异常值的同时,必须详细记录剔除的原因。这些原因可能包括采样误差、仪器故障、记录错误等。记录剔除原因的目的在于不仅是为了保持数据的透明性和可追溯性,还有助于后续对数据质量的评估和改进。通过分析剔除原因,可以找出监测过程中可能存在的问题,进而优化监测方法和流程,提高数据质量。最后,需要注意的是,剔除法虽然简单直接,但也可能导致信息损失。因此,在使用剔除法时,应谨慎评估其对整体数据分析结果的影响。如果异常值对分析结果产生显著影响,或者异常值的出现具有某种规律性,那么仅仅剔除这些值可能不是最佳的处理方式。

4.2 插值法

插值法在地下水数据处理中占据着重要的地位,尤其是当数据集中存在缺失或异常的数据点时。插值法的主要目的是根据已知数据点的信息,推算出缺失或异常数据点的可能值,从而填补数据空隙,确保数据的连续性和完整性。在地下水监测领域,常用的插值方法有多种,如线性插值、多项式插值、样条插值等。线性插值是最简单直接的方法,它假设相邻两个已知数据点之间的变化是线性的,通过这两点的连线来确定缺失点的值。这种方法适用于数据变化较为平缓的情况。多项式插值则通过构造一个多项式函数来拟合已知数据点,并据此推算缺失点的值。多项式的阶数可以根据数据的复杂性和需要进行调整。多项式插值在处理具有一定曲率变化的数据时表现较好。样条插值是一种更为复杂的插值方法,它通过构造一系列的分段多项式来逼近整体数据的变化趋势。样条插值在保持数据连续性的同时,还能够较好地保留数据的局部特征^[3]。在选择插值方法时,需要综合考虑数据的特性、插值的精度要求以及计算的复杂性等因素。此外,插值后的数据应进行验证,以确保其合理性和准确性。这可以通过对比插值前后的数据统计特征、绘制插值后的数据曲线图与其他相关数据进行对比等方法来实现。需要注意的是,虽然插值法可以有效地填补缺失或异常的数据点,但它并不能完全恢

复原始数据的所有信息。因此，在插值过程中应尽可能保持谨慎，避免过度拟合或失真。

4.3 平均法

在地下水数据处理过程中，平均法是一种常用且有效的技术手段，主要用于减少数据中的随机误差，提高数据的稳定性和可靠性。这种方法的核心思想是在一定时间范围内，对多个数据点进行平均处理，从而得到一个更具代表性的数值。平均法的应用非常灵活，可以根据实际需求选择不同的平均方式。例如，简单平均法就是将某一时间段内的所有数据相加后除以数据点的数量，这种方法适用于数据波动较小、分布均匀的情况。而当数据在不同时间段内呈现出不同的特点时，可以采用加权平均法，根据数据的重要性或可信度赋予不同的权重，再进行平均处理。通过平均法处理后的数据，能够更准确地反映地下水状况的实际变化趋势。因为随机误差往往会在平均过程中被削弱或抵消，从而使得处理后的数据更加平滑、稳定。这对于后续的数据分析、模型建立和预测都至关重要。此外，平均法还可以与其他数据处理方法相结合，如滤波技术，以进一步提高数据质量。例如，可以先利用平均法对数据进行初步处理，再应用滤波技术去除剩余的噪声和异常值。需要注意的是，虽然平均法能够有效减少随机误差，但也可能掩盖一些重要的瞬时变化或异常事件。因此，在应用平均法时，需要根据研究目的和数据特性进行权衡，确保处理后的数据既能满足分析需求，又不失真实性。

4.4 回归分析法

回归分析法在地下水数据处理中是一种高级且精确的方法。它通过建立数学模型，探究自变量与因变量之间的关系，进而对异常值进行预测和替换，使得处理后的数据更加接近真实情况。在具体应用中，回归分析法首先通过收集大量的地下水监测数据，确定影响地下水指标（如水位、水质等）的关键因素作为自变量，如

降雨量、温度、土壤类型等。然后，利用统计学方法，建立一个能够描述这些自变量与地下水指标之间关系的数学模型，即回归方程。这个回归方程可以是一个线性方程，也可以是一个非线性方程，具体形式取决于数据的实际分布情况。一旦回归方程建立完毕，就可以用它来预测在给定自变量条件下的地下水指标值。当遇到异常值时，可以利用这个回归方程来预测该异常值在正常情况下应该取的值，并用这个预测值来替换原始的异常值。这种方法不仅可以有效地处理异常值，还可以揭示出自变量与因变量之间的内在联系，为地下水资源的科学管理和决策提供支持^[4]。需要注意的是，回归分析法的准确性高度依赖于所收集数据的数量和质量，以及所选自变量与因变量之间关系的紧密程度。因此，在应用回归分析法时，应确保数据的准确性和完整性，并谨慎选择自变量，以避免模型的误用和误导。

结语

地下水数据异常值的检测与处理是地下水环境监测的重要环节。通过科学合理的检测方法和处理策略，可以提高监测数据的准确性和可靠性，为地下水资源的保护和管理提供有力支持。未来，随着技术的发展和方法的创新，地下水数据异常值的检测与处理将更加精准和高效。

参考文献

- [1]景永志,艾自东,田相臣.环境监测中异常数据识别与修复[J].环境工程技术学报,2024,14(03):1098-1104.
- [2]陈珂,文生仓,丁梅梅.环境监测中异常值的统计分析与实例[J].青海环境,2019,29(04):187-190.
- [3]李静.环境监测质量保证措施分析[J].中外企业家,2019(34):203.
- [4]吕建猛.提高环境监测数据质量的对策研究[J].河南建材,2019(06):108-109.