五 问题3建模与求解

5.1问题分析

问题3要求通过机理分析，来建立不同放牧策略对土壤化学性质的影响的模型，并根据建立的模型来预测2022年各小区化学成分的值。由于数据不全，这里仅研究放牧强度对土壤化学性质的影响。且考虑到数据仅提供了各小区化学成分的变化情况，因此我们需要有针对性地分析化学成分变化的内在规律，从时间与放牧强度两方面来进行考虑，从而通过数据预处理方法、可视化手段、灰色关联度分析以及多元回归来进行建模与预测。

5.2 数据预处理

根据附件14，我们通过观察发现以下特征：

1. 各小区均对应一种放牧强度，共有NG（不放牧）、LGI（轻度放牧）、MGI（中度放牧）、HGI（重度放牧）共四种放牧强度。而每种放牧强度共设置了三个小区进行测试。
2. 数据中仅存在2012-2020年偶数年的数据，且2012、2014、2016、2018年中的数据存在多项，而2020年仅存在一项,且说明为均值。
3. 化学成分共有四种，包括SOC（土壤有机碳）、SIC（土壤无机碳）、STC（土壤全碳）、N（全氮），而土壤C/N比为（SOC+SIC）/N得到。

5.2.1均值处理

由于数据中偶数年中各小区的化学成分有多次测数，而2020年为当年均值，因此，这里我们需要同样对偶数年各小区的化学成分取均值，以保证数据结构的稳定性。

5.2.2 异常值处理

达拉依准则又称3准则，在统计学中较为流行，常用于判断数值是否出现较大的异常。在统计学中，我们认为某个数值的取值基本上集中于该数值平均数上下浮动3个单位的范围内。

在正态分布中，表示均值，表示标准差，3准则就是指，数值分布在的概率为99.74%，因此，我们可以认为数值的取值基本集中在这个区间范围，当出现数值不在这个区间的情况时，则认为该数值为异常值，应当予以剔除。

在本文中异常值处理主要是根据拉依达准则（3准则）来剔除，即取空值。



对于均值处理后的数据，我们根据3σ原则，对均值处理后的数据进行检查，检查结果发现，均未出现异常值。我们认为该数据集稳定，认为合理均予以全部保留，以应对接下来的建模。

5.2.3 拉格朗日插值

由于数据中并不存在奇数年的数据，因此本文在进行预测前，先行对其进行插值处理，从而避免出现数据量过少而欠拟合的缺点。

5.2.4 数据标准化

在进行接下来的灰色关联度分析前，我们需要对数据进行标准化，采取Z-score标准化方法对数据进行处理，从而将数据变换为均值为0，方差为1，这样可以使数据具有正态性，对于灰色关联度分析易于计算以及对预测模型提供合适数据。标准化公式如下：



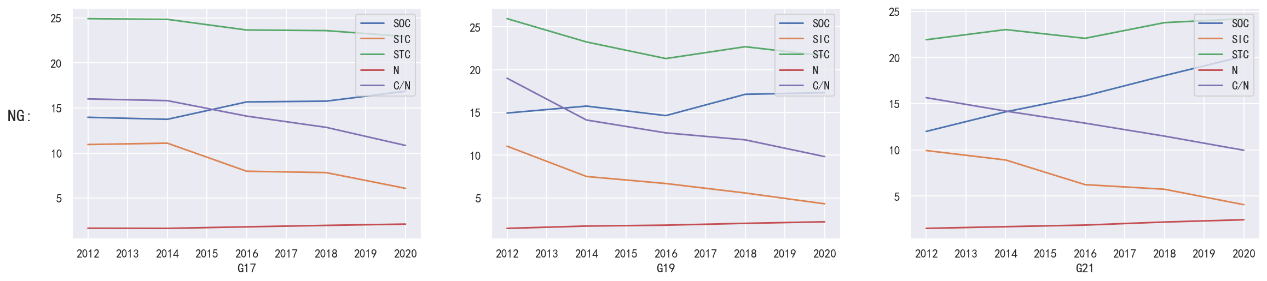
其中为均值，为方差。

5.2 描述性分析

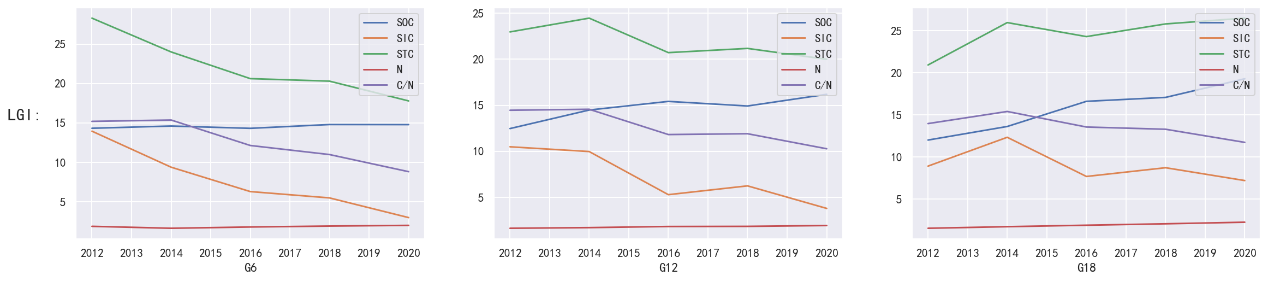
5.2.1 可视化分析

由于各组之间的表现有差异，且各组并未使用同种放牧强度进行测试，因此我们仅能从各组在其对应强度下各化学成分随时间的发展趋势进行分析。

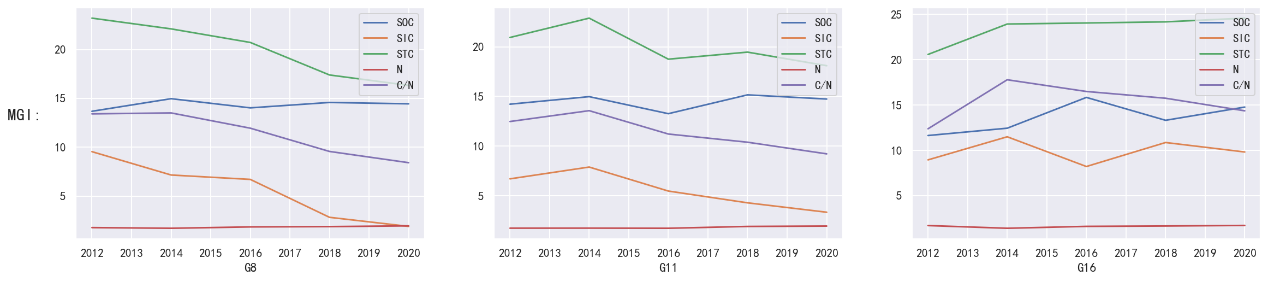
我们通过折线图绘制出各组的发展情况，其中每年化学成分的数值是通过当年的均值得到，但由于数据所对应的年份均为偶数年，因此在奇数年进行插值处理。



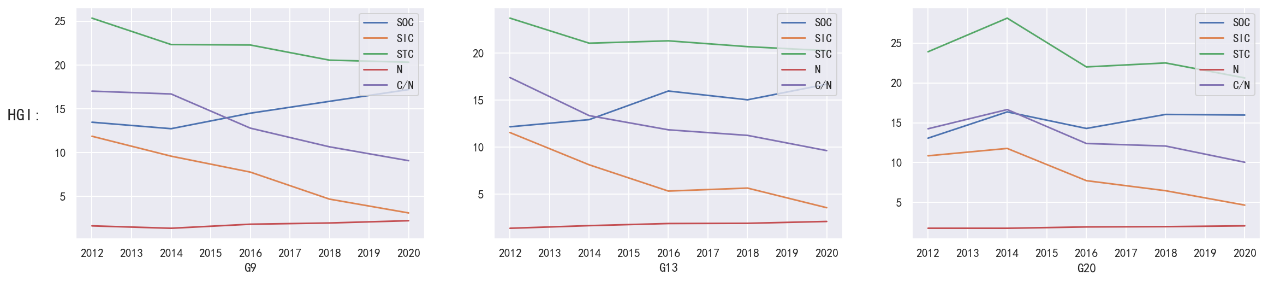
随着时间的发展，在NG强度下，SOC呈上升趋势，SIC呈下降趋势，STC有两组呈下降趋势，N基本呈上升趋势。



在LGI强度下，SOC呈上升趋势，SIC呈下降趋势，STC有两组呈下降趋势，一组呈波上升，N呈缓慢上升趋势。



在MGI强度下，SOC呈波动起伏，SIC有两组呈波动下降趋势，一组呈波动起伏状，STC有两组呈波动下降趋势，一组呈上升趋势，N呈缓慢上升趋势。



在HGI强度下，SOC呈波动上升趋势，SIC呈下降趋势，STC呈波动下降趋势，N呈缓慢上升趋势。

因此，在放牧这个行为上，对SOC，放牧与它呈正相关，能促进SOC的增长。对SIC，放牧与它呈负相关，SIC均表现为逐步下降，这也说明了放牧会抑制SIC的增长。对STC，放牧与它呈负相关，STC基本上随着时间的下降会逐步下降，但随着放牧强度的增加，STC波动上升的趋势不复存在，因此可能放牧强度会使STC下降加快。对N，均表现为缓慢上升，则放牧会在一定程度上刺激N的生长，同时表中明显可以看出，随着放牧强度的增强，N上升的趋势越来越慢，这也可能是放牧强度会抑制N上升的证据。

目前，我们仍无法判断放牧强度对化学性质的影响，因此我们进行进一步的研究与探讨。

* + 1. 时间序列分析
  1. 基于灰色关联度分析的评价模型

针对预处理后得到的数据，由于没有同小区不同放牧强度变化趋势的数据，因此我们转而考虑放牧强度对化学成分变化趋势的影响程度，即考虑不同放牧强度在不同样本下化学成分变化有无关联性。通过建立灰色关联度分析模型，与对照组进行对比，所得到的灰色关联度系数就是指放牧强度是否对化学成分的增长或减小产生了显著的影响。

同时我们还考虑到，在不同的年份，放牧强度的影响也许同样存在差异，因此，我们同样需要考虑随着时间的流逝，放牧强度与时间的灰色关联系数，以考虑放牧强度的影响是否在增大还是在减小。

5.3.1 灰色关联度模型

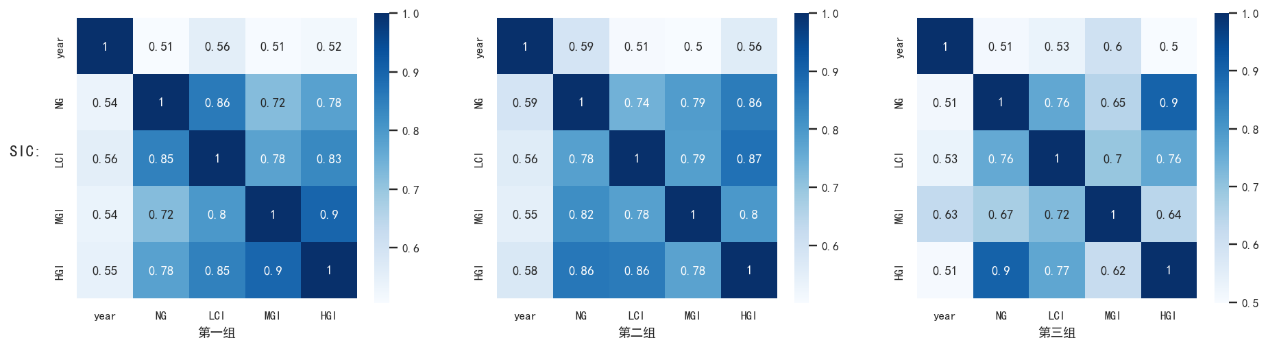
5.3.2 模型建立

由于每种放牧强度均有三个小区进行实验，因此我们分别根据化学成分和放牧强度分成三组，从而得到的灰色关联度系数更为精准，将三组的灰色关联度系数取均值，就可以得到各种放牧强度对化学成分变化趋势的平均影响。

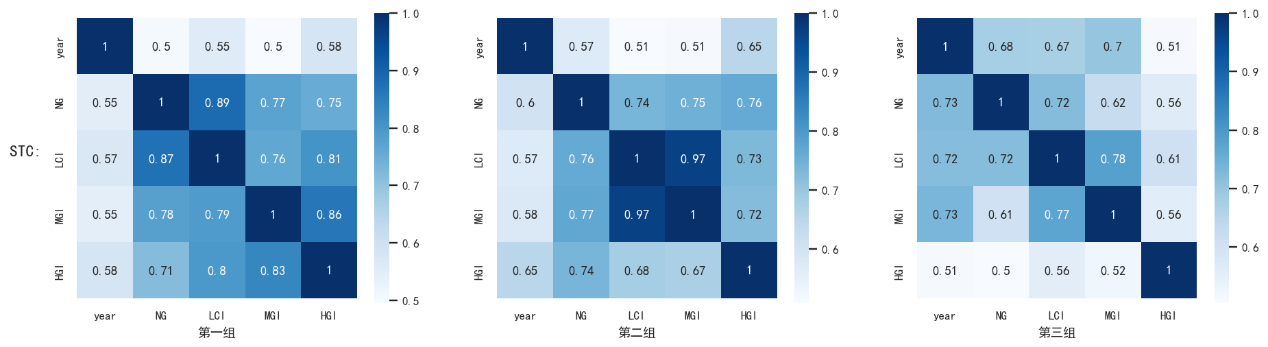
根据灰色关联度系数公式计算，我们得到了各组的灰色关联度系数矩阵，通过绘制热力图如下：



在SOC热力图中，我们可以明显发现随着放牧强度的增加，year与放牧强度的关联度越来越低，但在HGI出现了小幅度上升，这也许是由于在高强度放牧下，SOC不得不超量恢复以达到平衡。而对对照组NG，结果并未出现明显由于放牧强度所导致的SOC发展趋势的不同。



在SIC热力图中，我们可以发现，year与放牧强度之间并未存在显著的关联度。而对对照组NG，我们可以明显看出，放牧强度的增大，其关联度同样在增大，这说明了放牧强度的增大推动了SIC的变化趋势，而从前文，我们可以推断，放牧强度的增大在一定程度促进了SIC的下降。



在STC热力图中，我们发现year与放牧强度之间同样并不存在显著的关联度。而放牧强度在不同组中同样具有不同的关联度。



在N热力图中，很明显year与放牧强度的增大出现了同步的趋势，也就是说随着year的增大，放牧强度对化学成分变化的影响没有出现太大的改变。

我们将放牧强度对各化学成分的关联度取平均值，将放牧强度与year的关联度取平均值，得到结果如下表：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 放牧强度 | SOC | SIC | STC | N |
| NG | 1.00 | 1.00 | 1.00 | 1.00 |
| LGI | 0.66 | 0.80 | 0.78 | 0.76 |
| MGI | 0.62 | 0.74 | 0.72 | 0.65 |
| HGI | 0.66 | 0.85 | 0.65 | 0.73 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| YEAR | SOC | SIC | STC | N |
| NG | 0.80 | 0.54 | 0.58 | 0.85 |
| LGI | 0.74 | 0.53 | 0.58 | 0.84 |
| MGI | 0.63 | 0.54 | 0.57 | 0.66 |
| HGI | 0.76 | 0.53 | 0.58 | 0.73 |

根据灰色关联度，我们对各放牧强度与化学性之间的影响赋予权值。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 放牧强度 | SOC | SIC | STC | N |
| NG | 0.34 | 0.30 | 0.32 | 0.32 |
| LGI | 0.22 | 0.24 | 0.25 | 0.24 |
| MGI | 0.21 | 0.22 | 0.23 | 0.21 |
| HGI | 0.22 | 0.25 | 0.21 | 0.23 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 放牧强度 | SOC | SIC | STC | N |
| NG | 0.27 | 0.25 | 0.25 | 0.28 |
| LGI | 0.25 | 0.25 | 0.25 | 0.27 |
| MGI | 0.21 | 0.25 | 0.25 | 0.21 |
| HGI | 0.26 | 0.25 | 0.25 | 0.24 |

对每一个化学性质，既包含了放牧强度对其造成的直接影响，也包含了年份对放牧强度再对其造成的间接影响，因此我们建立公式如下：

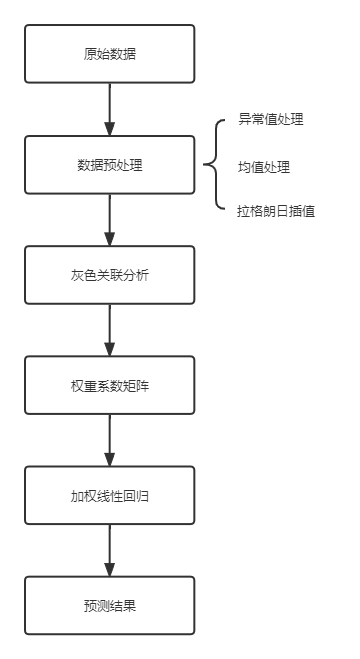


其中c为化学成分的数值，不仅与前一年的化学成分有关，还与放牧强度有关，而s为放牧强度，a为在s强度下，化学成分与放牧强度所对应的权重系数，y为年份与放牧强度之间的关系，b为在s强度下所对应的年份与放牧强度的权重系数。

* 1. 基于加权多元线性回归的预测模型

根据灰色关联度分析所得到权重系数矩阵，我们建立加权多元线性回归模型进行预测，预测结果如下:

* 1. 问题3小结



本文从机理分析的角度，通过灰色关联分析得到权重系数，建立加权线性回归模型从而进行预测，模型效果较好。