**1. 数据预处理与特征提取**

1. **对虾运动数据**
   * 以每段 8 min 视频计算平均移动距离（或平均速度），得到 122 条“日期–运动距离”记录。
   * 丢弃完全缺失的日期，或用线性插值/邻近平均法填补（若缺失天数很少，也可直接丢弃）。
2. **对潮水数据**
   * 每天提取两类特征：
     1. **当日潮差** = 当天最大潮高 – 当天最小潮低；
     2. **观测时潮高** = 取最靠近录像时间点（23:59–00:08）那次潮水高度。
   * 结果是同样 122 条“日期–潮差–潮高”记录。

**2. 简单相关分析**

* **Pearson/Spearman 相关**
  + 计算“运动距离” vs. “当日潮差” 的相关系数；
  + 计算“运动距离” vs. “观测时潮高” 的相关系数；
  + 并分别做散点图加回归直线，查看线性关系强度。
* **滞后效应（CCF）**
  + 对上述两个特征做 **交叉相关函数**（Cross‑Correlation Function），可检验前一天或后一天潮差/潮高对今日虾活跃度的滞后影响。

**3. 周期模型：Cosinor 与循环回归**

* **何时用 Cosinor？**
  + Cosinor 分析最适合“在已知生物节律周期（如 12 h 潮汐、24 h 昼夜）下，用余弦/正弦函数拟合时间序列的振幅和相位”Cornelissen (2014)。
  + 优点是能处理不等距采样、并给出节律参数（mesor、振幅、相位差）的显著性检验。
* **对你数据的建议**
  + **潮水侧**：若想深入刻画潮汐本身的 12.42 h 周期，可对每小时潮高做 Cosinor，检验振幅与相位。
  + **虾运动侧**：因你每天只在同一时间（约午夜）有一次 8 min 快照，分辨率不足以捕捉高频（12 h）波动，故直接做 Cosinor 拟合虾运动并不理想。
  + **替代方案——循环线性回归**：
    1. 先用潮水数据找出“观测瞬间距上次低潮已过时间占周期比例”（即潮相 phase）；
    2. 构造两列：$\cos(2\pi\phi)$ 和 $\sin(2\pi\phi)$，在普通线性回归里加以解释：  
       distancei=M+Acos⁡(2πϕi)+Bsin⁡(2πϕi)+εi\text{distance}\_i = M + A\cos(2\pi\phi\_i) + B\sin(2\pi\phi\_i) + \varepsilon\_idistancei​=M+Acos(2πϕi​)+Bsin(2πϕi​)+εi​
    3. 回归系数 $A,B$ 即可检验潮相对虾活跃度的影响（原理与 Cosinor 类似，但问题规模更小、更针对单次快照）。

**4. 时序‐频域分析（可选）**

* 如果未来你拥有更长时间的连续录像（例如全天多时段），可以考虑 **小波相干（Wavelet Coherence）**，在时–频域同时定位两信号（潮汐 vs. 活动）何时、在哪个周期上强相关。

**推荐路线图**

1. **实现每日对齐**：生成 122 条 (Date, ShrimpDist, TideAmp, TideLevel, TidePhase) 数据框。
2. **初步探索**：绘制三张散点图和时间序列走线图，直观对比。
3. **相关检验**：Pearson/Spearman + CCF（找最佳滞后）。
4. **循环回归**：基于潮相构造余弦/正弦项，做线性回归并解读振幅/相位影响。
5. **模型稳健性**：若担心日间差异，可在回归中加入“日期”为随机截距，做广义线性混合模型（GLMM）。

**分析计划建议**

用户想让我们分析其虾类活动与潮汐之间的关系。看到1月到4月的活动较弱，因此可能潮汐与活动的相关性减弱。为了更严谨，我们可以建议使用混合模型、季节性分解法，剔除季节趋势后进行部分相关分析，尝试交叉小波分析、时间序列分解等。可以在高峰季节与低活动季节分别做相关性分析，同时考虑做年度数据合并，使用GLMM模型对季节性和潮汐效应进行建模。

已思考 4 秒

要回答“到底是潮汐驱动盲虾活动”还是“季节效应（1–4 月低、10 月高）”在背后主导，我们需要把潮汐因素和季节因素同时纳入模型，而不是简单地“剔除某几个月再比较”。下面给你一个更系统、科学严谨的思路和可行方案。

**一、合并所有可用年份，构建长时序数据集**

1. **统一格式**
   * 对 2016–17、2017–18、2018–19、2020–21 四段数据都做同样的预处理：
     + 得到每天一条 date、shrimp\_dist、tide\_amp、tide\_level。
     + 增加特征：year=YYYY、month=1–12、day\_of\_year=1–365。
2. **数据总览**
   * 绘制 **跨年叠加的月度平均**曲线，观察多年来每月活动高低：
   * X‑axis: month 1–12
   * Y‑axis: shrimp\_dist (multi‑year平均±SEM)
   * 统计每年每月样本量，确保每组都有足够观测（>10–20 天）。

**二、分解季节与趋势：STL 或 Cosinor**

1. **季节分解（STL）**
   * 对每年每日 shrimp\_dist 做 STL（“Seasonal–Trend Decomposition”）：
     + **Seasonal**：一年周期内（365 d）的波动（你怀疑 1–4 月低、10 月高）。
     + **Trend**：跨年长期变化趋势。
     + **Residual**：去掉季节和趋势后剩下的“噪声”或短期波动。
   * 同样对潮位或潮差做 STL，提取它们的 **Residual**。
2. **Cosinor 模型**
   * 用已知年度周期 （T = 365 d） 构造余弦/正弦回归基：
   * shrimp\_dist ~ β0 + β1·cos(2π·day\_of\_year/365) + β2·sin(2π·day\_of\_year/365) + ε
   * 残差 ε 即去季节化后的活动水平。

**三、在“去季节化”的残差上做潮汐相关**

* 用上述 STL 或 Cosinor 得到的 **shrimp\_residual** 与 **tide\_residual**（或原始 tide\_level/tide\_amp）做：
  1. **Pearson/Spearman 相关**
  2. **CCF**（考察滞后关系）
* 这样就能判定：**在同一去除大季节效应后**，潮汐对盲虾活动的“即时”或“滞后”影响有多大。

**四、混合效应模型（LME/GAM）**

直接把季节和潮汐一起放入同一个回归模型中：

**模型示例**

shrimp\_disti=β0+β1 tide\_leveli+β2 tide\_ampi+f(day\_of\_yeari)+(1∣yeari)+εi\text{shrimp\\_dist}\_{i} = \beta\_0 + \beta\_1\,\text{tide\\_level}\_{i} + \beta\_2\,\text{tide\\_amp}\_{i} + f(\text{day\\_of\\_year}\_{i}) + (1|\text{year}\_i) + \varepsilon\_i

* f(⋅)f(\cdot) 可以是：
  + **正弦/余弦基**（Cosinor），
  + 或 **样条函数**（GAM），
  + 或 **月分类**（把 month 当作 factor）。
* (1∣year)(1|\text{year}) 做 **随机截距**，控制不同年份间的基线差异。

通过检验 β1,β2\beta\_1,\beta\_2 的显著性，就能说“在控制季节效应后，潮位/潮差是否真的影响活动”。

**五、分段或分季节相关**

如果你仍想做“简单相关”，可以：

1. **按季节分段**
   * 低活期：1–4 月
   * 高活期：10–12 月
2. 分别计算这两段的 Pearson/Spearman 相关 & CCF。
3. 比较 ρlow\rho\_{\rm low} vs. ρhigh\rho\_{\rm high}，看潮汐在高活期 vs 低活期中的驱动力差异。

**六、不要随意剔除数据**

* **剔除低活动月** 会低估季节对整体的影响，也会人为放大“潮汐-活动”相关。
* 更科学的做法是**在模型里显式控制季节效应**，然后再检验潮汐效应。

**总结推荐路线**

1. **合并** 所有年份——做 **STL 分解** 或 **Cosinor**。
2. 在**去季节化残差**上做**相关/CCF**，或者
3. 构建**混合效应模型**同时纳入潮汐和季节变量，检验它们的独立贡献。
4. （可选）**分季节**做子样本分析，探索“季节 vs. 潮汐”谁更强。

这样，你既能验证“1–4 月活动减弱”这一季节现象，也能量化“潮汐”在去掉季节趋势后对盲虾活动的真正驱动力。