HPC Project

Monte Carlo Simulation in Paleo-Climate

刘群 马佳良 胡晨琪

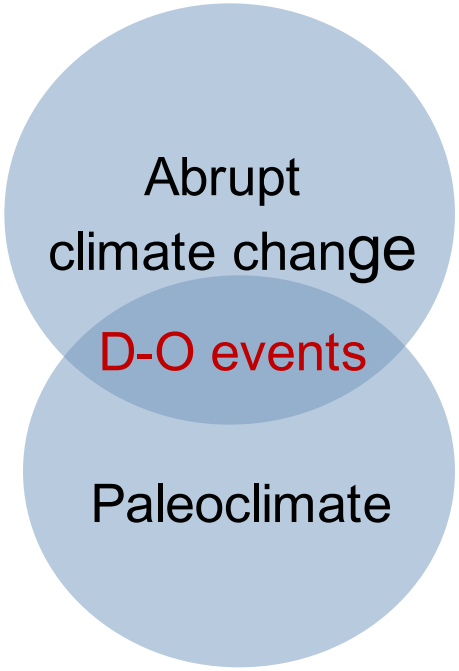
Center for Earth System Science

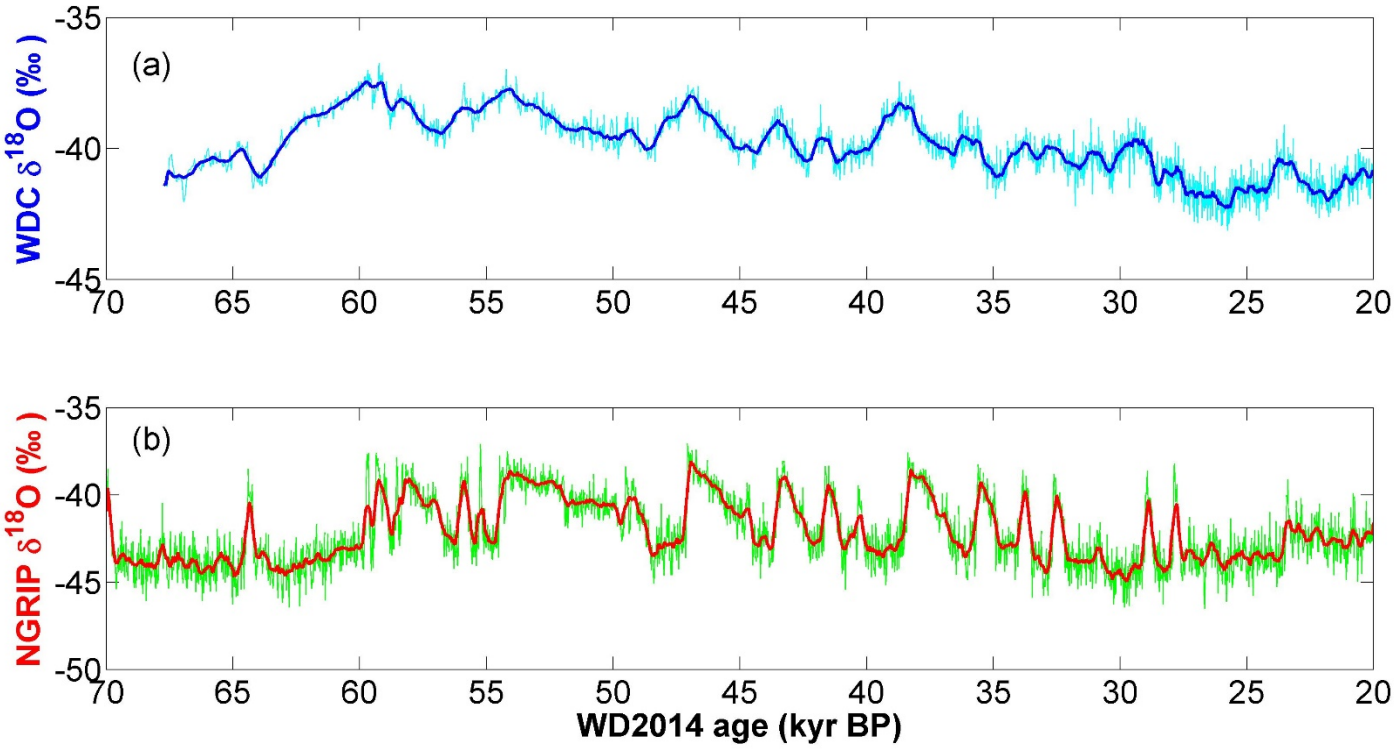
June 19, 2015

Outline

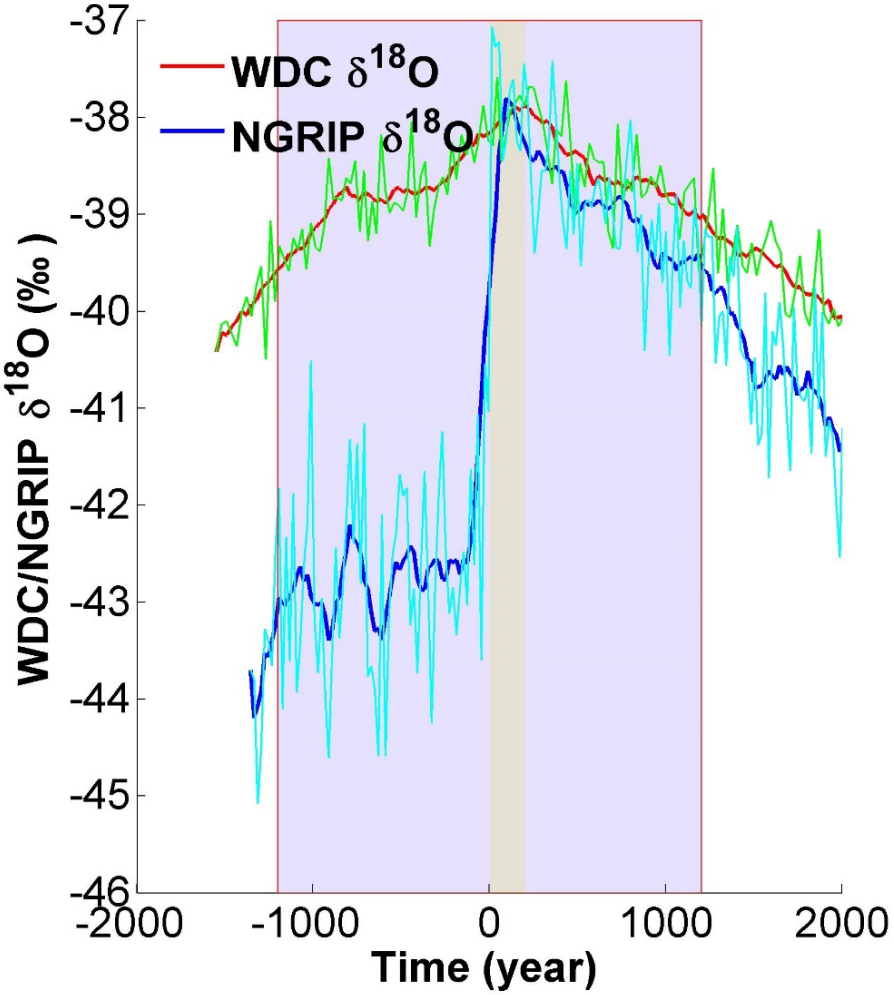
- Background
- Methods
- Results
- Discussion

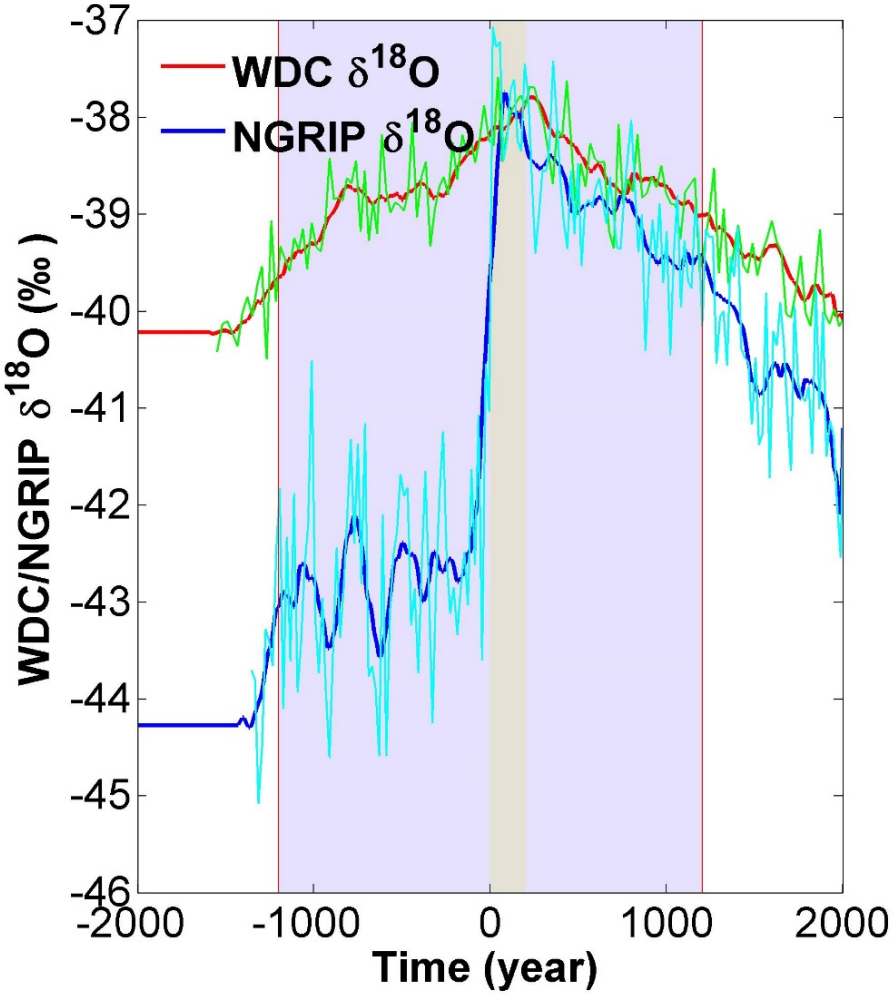
Background


Abrupt
climate change
D-O events
Paleoclimate

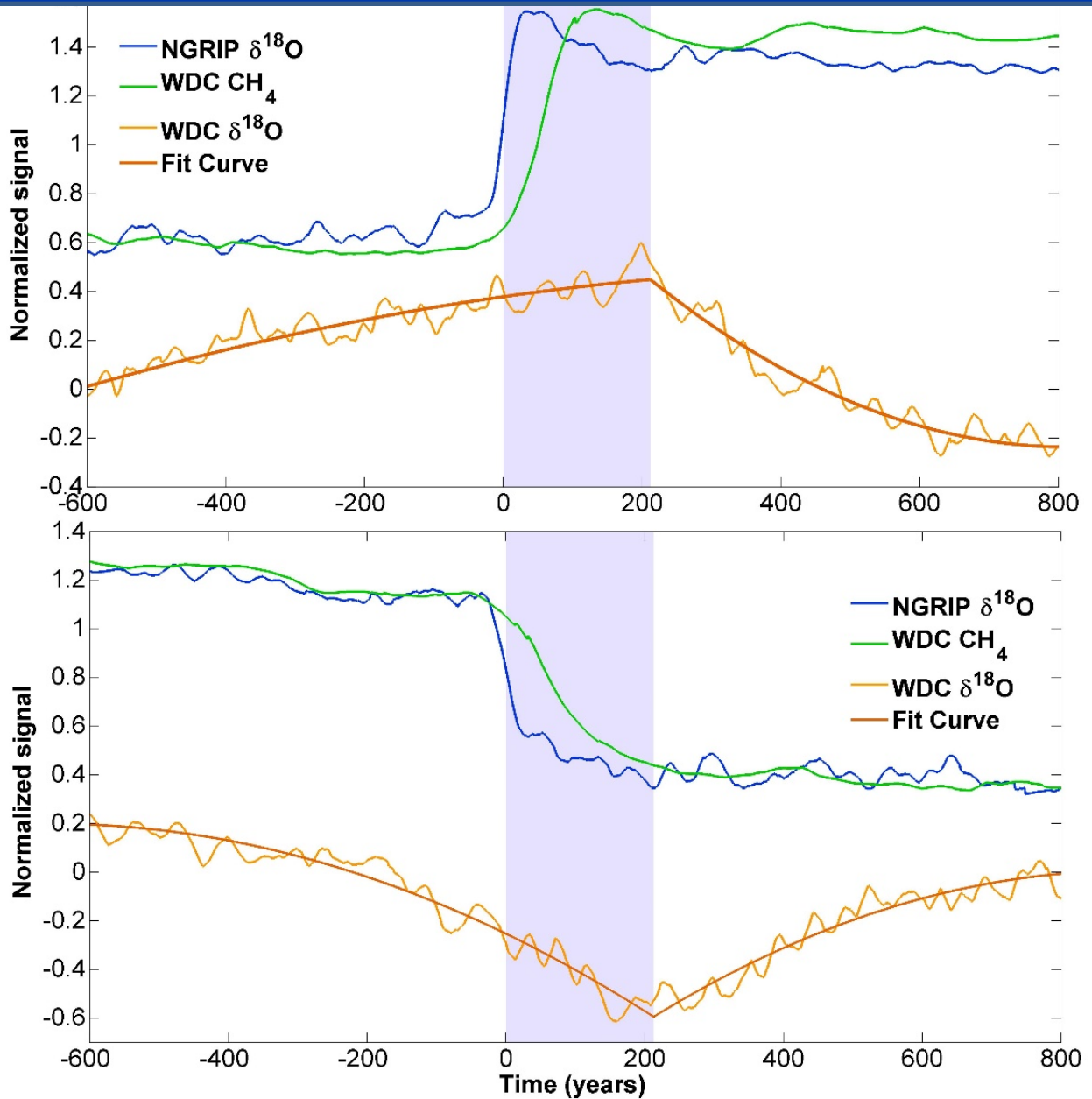

(a)
WDC δ18O (‰)
-35
-40
-45
70
65
60
55
50
45
40
35
30
25
20
(b)
NGRIP δ18O (‰)
-35
-40
-45
-50
WD2014 age (kyr BP)

Methods – Split time series into individual windows





Methods – Stack and Find Breakpoint



Monte Carlo Sensitivity Study

如何进行Monte Carlo 模拟

检验breakpoint时间的可靠性:

1. 在初始值的基础上，加上一些满足高斯分布的随机扰动（分别针对系统误差和非系统误差进行扰动）
2. 对这些数据分别进行求取breakpoint的操作

总共进行800个chronology的模拟，每次模拟重复进行100次，每次的随机扰动都不同。

Main Difficulties in Programming

Original codes are written in MATLB

Change to Fortran:

- Fortran codes more than 900 lines (including some comments)
- Write functions that don't exit in Fortran, including interpolation, polyfit.
- Use high performance computing library, such LAPACK, Intel MKL
- Call the library to calculate the linear equations and to produce Gaussian random number.
- Change serial programs to parallel one using MPI
- Write Makefile to compile and run the program

Change Matlab to Fortran

Fitting problem can be transformed into solving linear equations using Least Square Method.

- We use functions in LAPACK library for then linear equations.
- DGETRF and DGETRI to calculate

$$X^T X A = XTY$$
$$A = (XTX)^{-1} X^T Y$$

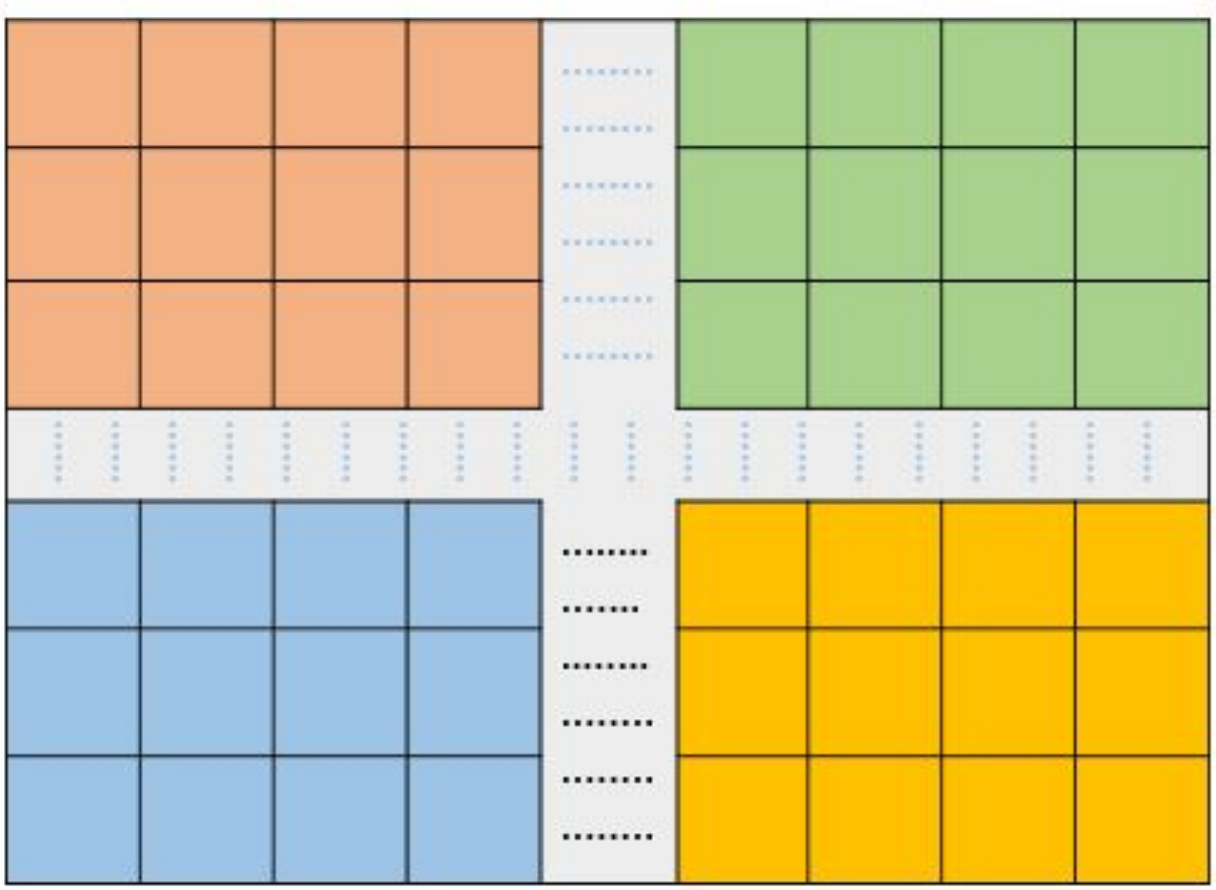
Change Matlab to Fortran

In Monte Carlo simulation, we need to produce random numbers in Gaussian distributions, so we use function in Intel MKL to produce them.

- vdrnggaussian(method, stream, n, r, a, sigma)
- r: return Gaussian random vector
- a: average number
- sigma: standard deviation

How to parallelization

Monte Carlo Simulation parallelization



Each process calculates some MC simulations

(Different colors stand for different processes)

How to parallelization – Parallel IO

串行I/O

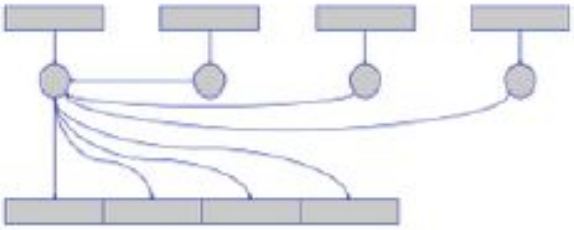
0进程负责分发和收集数据
进行读写操作，将是整个
程序的瓶颈
效率低，可扩展性差

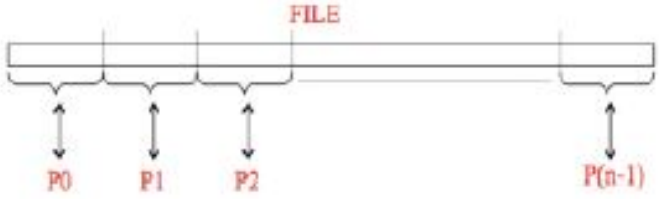
进程独立I/O

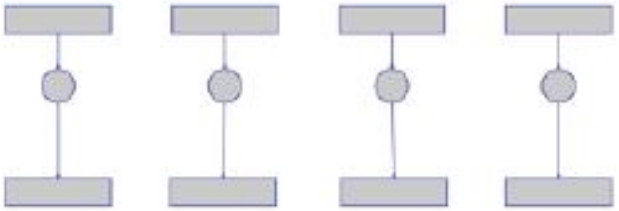
每个进程读写自己的文件
高度并行化
需要管理大量的小文件，
不利于后处理

并行I/O

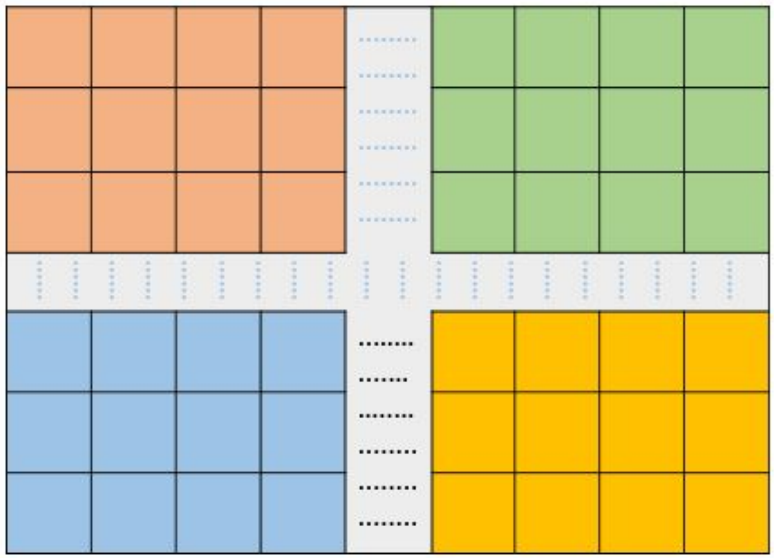
并行读写同一个文件
高度并行
利于文件管理和后处理







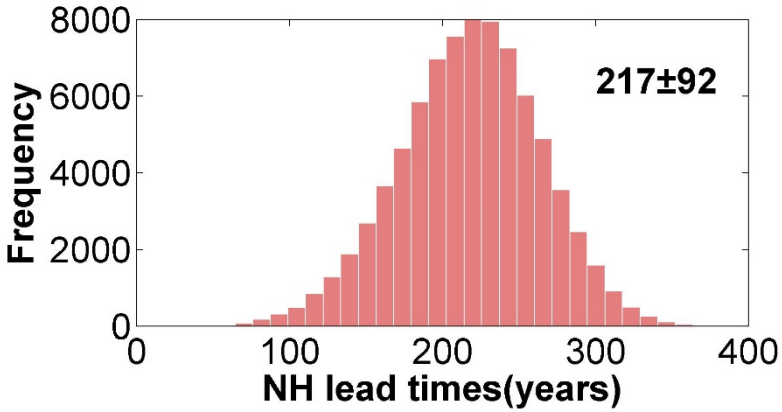
How to parallelization – Parallel IO

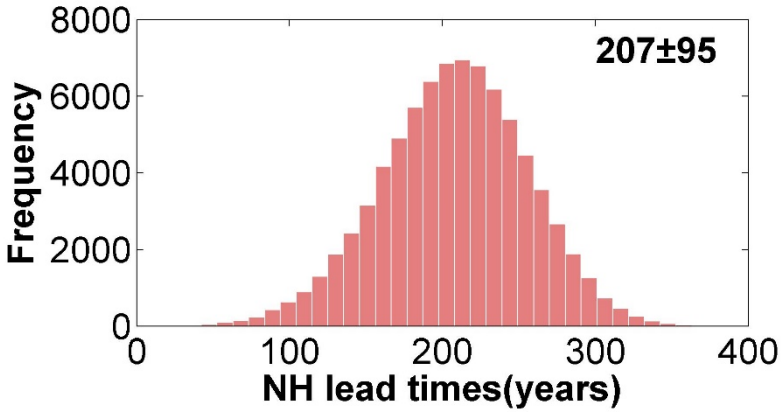


(Different colors stand for different processes)

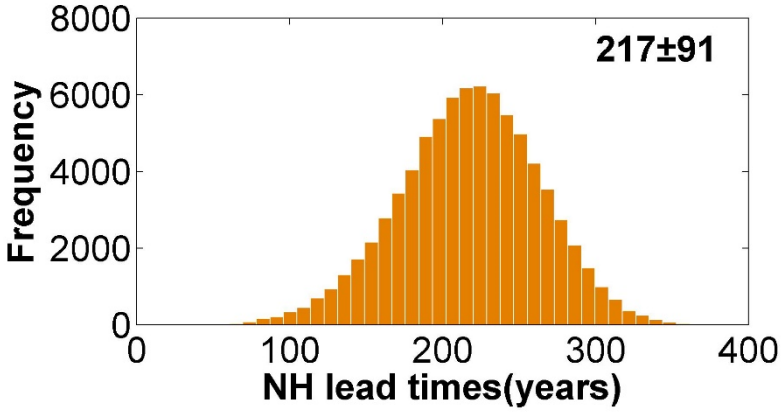
```
call MPI_TYPE_CREATE_SUBARRAY(2, gsize, lsize, starts, &
    MPI_ORDER_FORTRAN,  MPI_DOUBLE_PRECISION, filetype, ierr)
call MPI_TYPE_COMMIT(filetype,  ierr)
call MPI_FILE_OPEN(MPI_COMM_WORLD, "t_MC.dat",  &
     MPI_MODE_CREATE+MPI_MODE_WRONLY, &
     MPI_INFO_NULL, fh1,  ierr)
 call MPI_FILE_SET_VIEW(fh1,  0_MPI_OFFSET_KIND, &
  MPI_DOUBLE_PRECISION, filetype, "native", MPI_INFO_NULL, ierr)
call MPI_FILE_WRITE_ALL(fh1, t_MC, xsize*ysize ,  &
     MPI_DOUBLE_PRECISION, stat, ierr)
call MPI_FILE_CLOSE(fh1,  ierr)
```

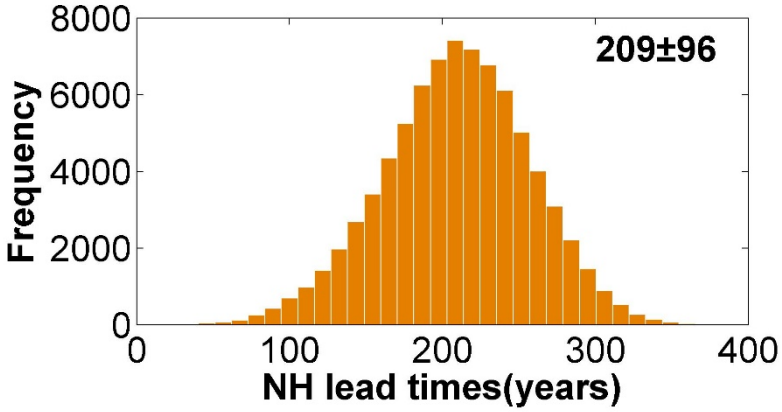
Results – Monte Carlo Sensitivity Study





MATLAB



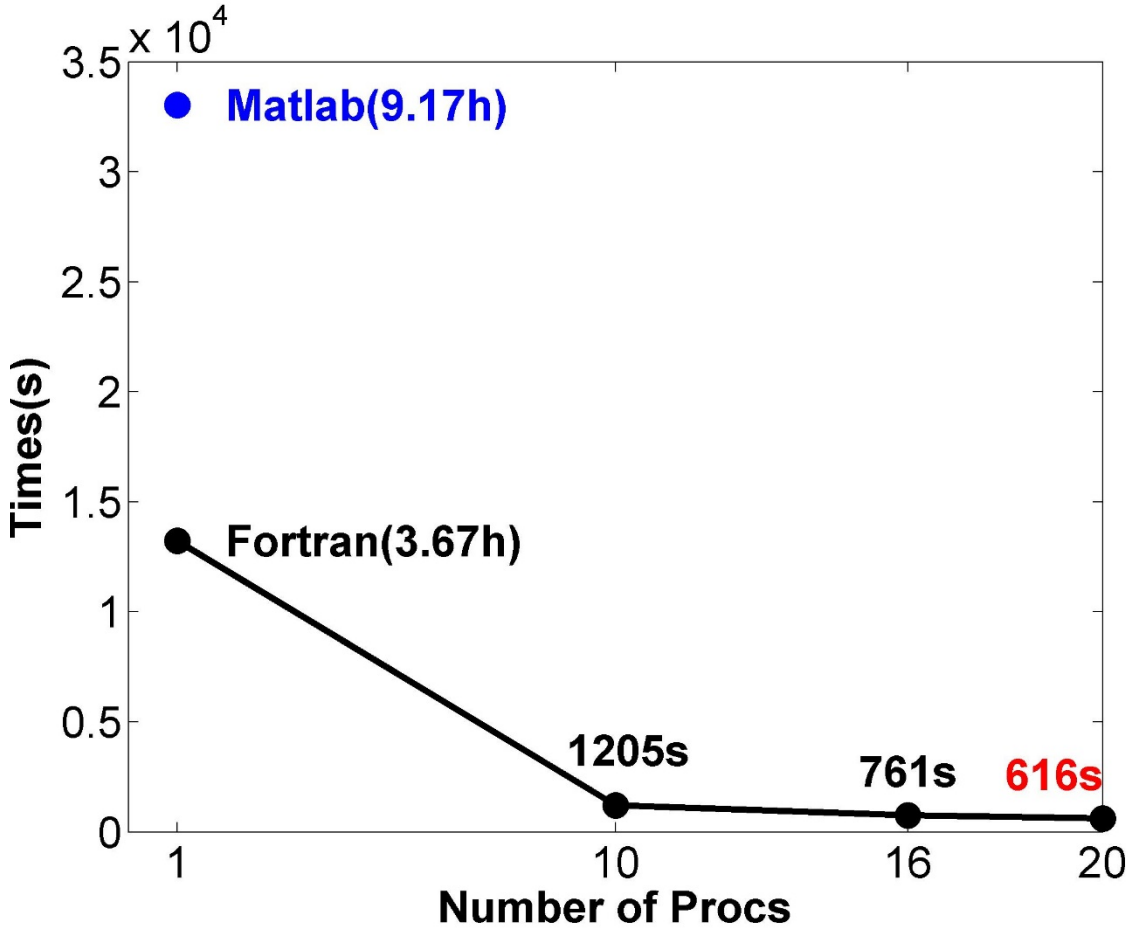


Fortran parallel

warm event

cool event

Runtime



运行效率比MATLAB提高了51.6倍
运行效率比串行Fortran程序提高了20.5倍

Thanks !

Questions & Comments