**CTS-SS求解GFDL CM模式CNOP及其在**

**ENSO事件最快增长初始误差中的应用**

摘要

这是摘要这是摘要。

**关键词：**连续禁忌搜索，数据降维 CNOP，GFDL CM

**This is the English Name**

ABSTRACT

Abstract

Key words：

目 录

[1 引言 3](#_Toc482380921)

[1.1 研究背景和意义 3](#_Toc482380922)

[1.2 研究现状 4](#_Toc482380923)

[1.3 研究内容 4](#_Toc482380924)

[1.4 本文结构 4](#_Toc482380925)

[2 相关研究内容 5](#_Toc482380926)

[2.1 GFDL CM模式 5](#_Toc482380927)

[2.1.1 大气模式 5](#_Toc482380928)

[2.1.2 海洋模式 6](#_Toc482380929)

[2.2 智能算法求解CNOP 6](#_Toc482380930)

[2.2.1 CNOP定义 6](#_Toc482380931)

[2.2.2 智能算法综述 7](#_Toc482380932)

[2.2.3 特征提取方法 7](#_Toc482380933)

[2.2.4 智能算法求解CNOP流程 9](#_Toc482380934)

[2.3 ENSO事件最快增长初始误差 10](#_Toc482380935)

[参考文献 12](#_Toc482380936)

# 1 引言

## 1.1 研究背景和意义

大气和海洋科学中的数值天气和气候可预报性研究一直是国内外研究的热点问题，目前已经有很多方法建立并被应用于探讨数值天气和气候的可预报性问题。由于初始扰动的选择和发展会对预报结果的准确性和稳定性产生重要影响，因此，如何确定描述气候发展变化的非线性系统的最快增长初始扰动成为上述问题中一个重要的研究领域。

穆穆等人提出了条件非线性最优扰动（Conditional nonlinear optimization perturbation, CNOP）[1-2]的概念，它是满足一定约束条件且在预报时刻具有最大非线性发展的一类扰动。CNOP方法充分考虑了动力系统的非线性特征，衡量了非线性在动力系统从初始时刻演变到预报时刻过程中起到的作用，是线性奇异向量方法在非线性框架下的自然扩展，为研究大气和海洋科学的可预报性问题提供了有力的工具。

ENSO是厄尔尼诺（El Niño）和南方涛动（Southern Oscillation）的合称，是短期气候年际变化的最强信号之一。ENSO现象是引起全球大气环流和水分循环异常的重要原因，全球各地许多灾害性的气候事件都与其相关。科学研究表明，ENSO是导致全球各地破坏性干旱、暴风雨和洪水的罪魁祸首。1997至1998年的厄尔尼诺导致成千上万人死亡，并造成世界直接经济损失达数十亿美元。同时，我国发生的98年长江洪灾、南方雪灾等都有其有关，所以其可预报性研究一直是学术界关注的热点问题，CNOP方法自身的提出就是基于对ENSO事件的可预报性研究。结合CNOP方法，使用数值模式提高ENSO事件的可预见性，是当代短期气候预测研究中的一个重要问题。

## 1.2 研究现状

求解CNOP的过程本质上是获得最优解的过程，而最优解的定义取决于具体求解的问题。在ENSO事件可预报性研究中，最快增长初始误差（Optimally growing initial errors）是指发展最快且导致不容忽视的预报误差的初始误差。作为ENSO事件可预报研究的一个重要内容[3]，研究其发展变化可以帮助我们更好地预测ENSO事件。所以，寻找ENSO事件中产生最快增长初始误差的CNOP就是本研究课题的最优解。当前，伴随方法是求解CNOP最流行的方法，但其完成一次数值实验需要耗费大量的时间，而获得一个有意义的结论有需要通过大量的数值实验[4]，尤其对于复杂的全球海气耦合模式（如本文所采用的GFDL CM），开发相应伴随模式的往往需要耗费数年时间，工作量巨大。而作为计算机领域最优化问题求解问题的一种有力工具，智能算法提供了一种新的求解思路。

智能算法是人们受到自然规律的启发而建立的求解问题的一类算法，其目的是解决最优化问题。由于其求解过程不依赖梯度信息，可以用于求解CNOP这类免伴随的问题。根据搜索个体的数据，大致可以分为单粒子智能算法和多粒子智能算法。温仕成[5]比较了多种智能算法在求解CNOP过程中的结果和稳定性，结果表明，无论是单粒子智能算法还是多粒子智能算法，优化智能算法求解CNOP都可以使其能够应用于中等复杂程度以上的数值模式，这为本文的研究提供了理论依据。

## 1.3 研究内容

本文主要研究利用基于Sine映射和分段式搜索的连续禁忌搜索算法（Continuous Tabu Search Algorithm with Sine Maps and Staged Strategy，简写为CTS-SS）[6]求解GFDL CM模式CNOP，并设计数值实验，将其应用于求解ENSO事件中产生最快增长初始误差的CNOP。主要进行的工作有：

（1）研究模式数据的特征提取。GFDL CM作为全球海气耦合模式，直接应用智能算法求解必将导致收敛速度过慢等问题，主要使用Matlab工具，使用主成分分析（Principal component analysis，简写PCA）方法对模式数据的初始值进行降维处理，同时比较不同的PCs对最终结果的准确性和稳定性产生的影响。

（2）设计并调整CTS-SS算法参数。由于CTS-SS仅仅是一种算法思想，在实际使用过程中，我们需要结合具体问题确定其中各项具体参数，包括领域的设置，交换过程的定义和禁忌判断参数的确认等，都需要基于数值实验的分析。尤其是面对GFDL CM这种高复杂度的海气耦合模式，不同初始值的确认会对结果造成非常大的影响。

（3）设计数值实验。实现完整程序代码（基于JAVA语言），将调整过后的CTS-SS算法应用于GFDL CM模式动态求解（每次模式运行结束后即获取结果进行比较），不间断地轮询模式运行状态，将所得结果即使作为输入参数应用于智能算法各个阶段的寻优过程，获得输出后验证结果的正确性和稳定性。

## 1.4 本文结构

根据本文的研究内容，论文的组织结构安排如下：

第一章 引言。主要介绍本文的研究背景和意义，阐述本文的主要研究内容以及具体使用的研究方法和理由。

第二章 GFDL CM模式和智能算法求解CNOP的方法介绍。主要介绍本文选用的GFDL CM模式，描述其各个子模块的关系和选择理由，之后描述求解CNOP的方法，重点介绍智能算法求解CNOP，并分析各自的优缺点。

第三章 GFDL CM模式样本采集和数据降维研究。阐述了GFDL CM模式数据的选择和具体的降维方法及过程，重点描述PCA方法在GFDL CM模式数据的具体应用，同时设计实验以比较不同PCs的选择对实验效率造成的影响。

第四章 CTS-SS智能算法分析与具体应用。本章主要介绍整体实验的设计思路和具体实现，描述了数值实验的具体运行过程，并对实验结果进行分析和讨论。

第五章 结论和展望。对上述工作进行总结，并对今后的研究进行了展望。

# 

# 2 相关研究内容

## 2.1 GFDL CM模式

气候模式是用于提高人们对各种时间尺度（季节，年度，年代等）气候现象的理解和可预测性的重要工具，其预测结果能为人们各类决策提供必要的信息，如水资源管理，农业，交通和城市规划等。其中，GFDL CM是由美国国家海洋大气管理局地球流体动力学实验室(Geophysical Fluid Dynamics Laboratory）开发的一个全球海-气耦合环流模式。其始于20世纪60年代，经过50余年的发展，目前已经在气候变化、 热带气旋等的研究中得到广泛的应用,其已被实验证明在ENSO事件可预报性问题中具有较高的可信度[7],这也是本文选取其作为主要运行模式的原因。

本文所采用的模式版本为GFDL CM2p1，该模式共由大气、海洋、陆地和海冰四个子模式组成，其中最重要的两个子模式为大气模式和海洋模式，下面分别介绍这两个子模式。

### 2.1.1 大气模式

GFDL CM2p1大气模式垂直方向上共24层，其水平分辨率2°\*2.5°。利用时间步长为3小时的大气辐射和时间步长为0.5小时的其他大气物理量，并包括一个日照昼夜循环。该分辨率足以分析产生大中型气旋等天气变化的原因。

### 2.1.2 海洋模式

GFDL CM2p1海洋模式垂直方向上共50层，其水平分辨率在经度方向为1°，在纬度方向由极地地区的1°变化到赤道的1/3°。对于垂直方向，其最高22层中每层厚度为10米，共220米。其中，一个三极网格贯穿欧亚大陆、北美洲和南极洲，以避免北极上空出现极性滤波。该分辨率下可以比较明显地辨别ENSO事件的发生与否。图2.1为使用模式年数据（101~200），运行GFDL CM2p1模式生成的海表面温度（Sea Surface Temperature，简写为SST)模式生成值减去实际观测值的图像。

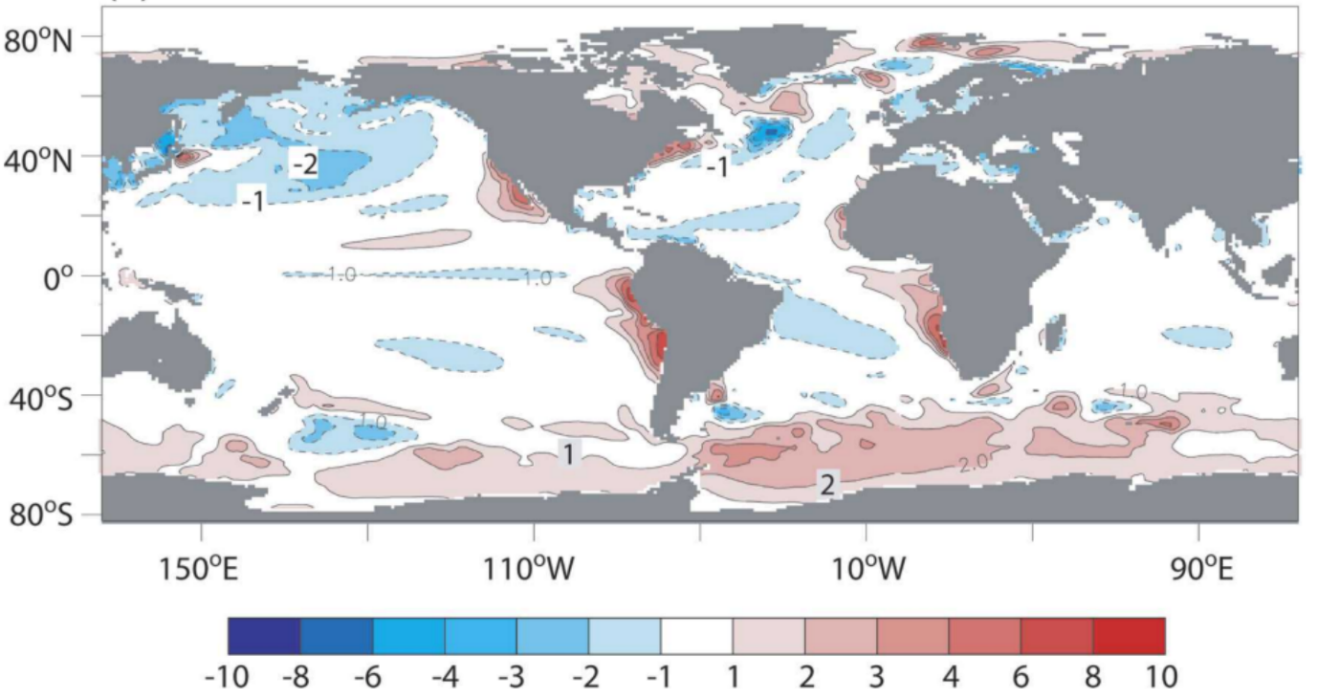


图2.1

## 2.2 智能算法求解CNOP

### 2.2.1 CNOP定义

CNOP是指满足一定的约束条件下，在预报时刻会对预报结果不确定性产生最大影响的一类初始扰动。根据大气和海洋科学可预报性的分类方法，目前CNOP被分为两类：与初始误差有关的CNOP方法，称为CNOP-I方法；与初始误差和模式误差都有关的CNOP方法，称为CNOP-P方法。对于其求解过程而言，首先需要考虑如下初值问题：

 （2.1）

其中*N*是非线性算子，*U*用来表示海表面温度异常、斜温层深度距平和表层流速距平，*U0*表示*U*的初始向量。方程（2.1）在*τ*时刻的解可以表示为：

 （2.2）

其中，假设*Mτ*表示从初始时刻到τ时刻的非线性传播算子,*u*0表示初值为*U0*的基态的初始扰动，将这一初始扰动叠加于状态向量*U0*，可以得到：

 （2.3）

其中，代表了初始扰动*u*0 的非线性发展。

CNOP表示这样的一类初始扰动，其满足目标函数在给定初始约束条件（约定，具体约束条件取决于具体的问题）时取到最大值，即：

 （2.4）

其中，

 （2.5）

*U0*代表基态的初态场，*u*0表示初始扰动。为了方便问题求解，我们通常把该最大值问题转化为求解如下的最小值问题：

 （2.6）

### 2.2.2 智能算法综述

智能算法是指人们受到自然（生物界）规律的启发，模仿其求解问题的一类算法，也被称为现代启发式（Modern heuristics）方法或计算智能（Computational intelligence）。这类算法可以上溯到最优化技术，也就是研究在众多解决方案中什么样的方案是最优的，或者怎样找出最优的解决方案。

现实世界中的许多工程问题或管理问题都可以归结为带约束的最优化问题，其中约束和最优的定义都取决于需要解决的具体问题。总体说来，最优化问题可以分为函数优化问题和组合优化问题两大类，其中，函数优化问题是指求解对象为一定区间内的连续变量，而组合优化问题是指求解对象为解空间中的离散对象。

对于经典的数值方法（如解析法，数值法）而言，其求解上述两类最优化问题过程中最大的困难在于目标函数通常存在诸多局部最优解，而我们的数值方法通常是在局部区域内求得最优解，而不是获得整体解空间上的最优解[8]。尤其是对于组合优化问题，其计算复杂度很高，属于非确定性多项式困难问题（non-deterministic polynomial-time hardproblem，NP-hard），除了通过枚举法穷尽一部分解空间外，没有更好的解法。同时随着问题规模的增大，解空间会呈现出指数级甚至阶乘级增长，这个时候希望求得准确的最优解实际上已经不可能。而智能算法便是相对于最优算法提出的，一个问题最优算法可以求得该问题的最优解，而智能算法则可以认为是一种基于直观或经验构造的算法，通过可以承受的花费（时间复杂度和空间复杂度等）来获得待解决优化问题的一个可行解，该可行解与最优解的偏移程度不一定可以预先估计，但从工程实际的角度来看，在一定误差范围内的近似解都是可以接受的。同时，由于数学模型（比如本文采用的CNOP）本身就是对实际问题的简化，或多或少会忽略一些因素，加之数据采集和参数估计准确性等原因，可能使得最优算法所求得的解与智能算法获得的解产生较大误差。

由于智能算法速度快，操作相对简单等特点，其已经得到了很快的发展。无论于工程中的许多问题，比如通讯网络的结构优化，管道重组等，还是计算机科学中的许多问题，如旅行商问题、0-1背包问题、聚类问题等都获得了广泛的应用，并都取得了一定的研究成果。但目前还没有将智能算法运用于求解GFDL CM模式CNOP的工作，这也是本文的可行性和创新性。

### 2.2.3 特征提取方法

对于中等程度以上的模式来说（比如本文采用的GFDL CM模式），其问题的数据规模一般在106  以上，而智能算法一般只能直接应用于上述的小规模问题。当问题规模较小时，智能算法能够很快的求出全局最优解。然而，随着问题规模的增大，智能算法需要增加个体规模来获得对全局系统的搜索，这会导致收敛速度的锐减，甚至造成不收敛的情况，这个时候就需要对数据进行降维处理。将数据从原始的千万级别的高维空间转化为一个数十维的低维空间，之后再在该低维空间上应用智能算法求解CNOP。特征提取方法求解高维度问题的有效性已被文献[9]所证明，同时，已经有将PCA作为特征提取方法，运用于智能算法求解CNOP，并取得理想的结果的工作[10]。下面以PCA为例，介绍特征提取方法的主要原理：

假设由*n*个历史样本*x1，x2，...xn，xi∈Rm（i=1，2...n）*组成一个数据样本矩阵。首先对该矩阵进行无量纲化和中心化两步预处理。无量纲化是指将有量纲的物理量（比如SST等）通过数值方法来避免量纲不统一引起的数值计算的问题，其具体操作如下式：

 （2.7）

其中*a*是一个正常数。

中心化是指每个样本减去样本均值，如下式：

 （2.8）

这样样本数据就变成了*x’i*（*i*=1, 2...*n*），新生成的样本矩阵为*X*’=（*x’*1, *x’2*...*x’n*）。经过数据预处理，我们使用PCA方法对数据进行降维，其目的是为了求得这样一组特征向量*β*i，使得*x’i（i=1，2...n）*在其方向的投影最大，用公式表述如下：

 (2.9)

经过验证可知，和分别为样本矩阵的协方差矩阵*M*的特征值和对应的特征向量。

 （2.10）

其中是一个的矩阵，*M*是一个的矩阵。通常情况下，，因此，若直接对矩阵*M*进行特征分解，计算量非常大（在Matlab等软件上运行甚至会因运算量过大无法求得结果）。但是，矩阵是一个，若对它进行特征分解，则相对容易得多。其特征分解过程为：

 （2.11）

和分别为矩阵的特征值和特征向量。对上式两端分别乘以可得：

 （2.12）

是就协方差矩阵*M*对应特征值的特征向量，因此，通过这种方式间接地求出了特征基向量：

 （2.13）

因为样本数据已进行了中心化处理，所以其协方差矩阵的秩不超过（*n*为样本个数）。

### 2.2.4 智能算法求解CNOP流程

基于对智能算法理论和可行性的研究，结合具体的CNOP问题求解，我们可以梳理出智能算法求解CNOP流程，如下图所示：

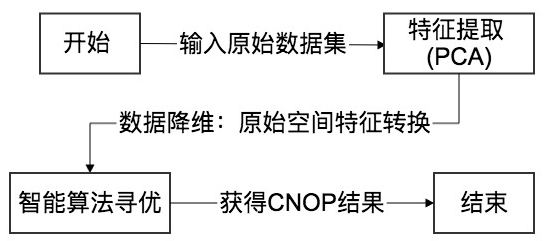


图2.2

根据流程图，具体阐述智能算法求解CNOP流程：首先，输入原始数据集，对数据进行上述无量纲化和中心化预处理步骤，得到样本矩阵*X’*，之后通过公式（2.13）得到特征向量*β*1、*β*2、... *β*k，将其组成主成分矩阵*C*，即*C*=[*β*1, *β*2, ... *β*k]，之后将初始扰动投影至该主成分空间，即*u0=C·ω*，在这里*ω*为初始扰动*u0*投影后的坐标。这样一来，公式（2.6）可以转换为：

 (2.14)

此时的自变量从*u0*换成了*ω*，由于主成分矩阵C的关系，*ω*的维度为k维而不再是原先*u0*的m维，且k << m。事实上，公式（2.14）的约束条件||C·*ω*||≤*δ*，可以被进一步如下简化：

 (2.15)

此时，CNOP问题的适应度函数就可以借助公式（2.15）转化为：

 (2.16)

公式（2.16）就是在提取主特征、完成数据降维后交由智能算法计算的适应度函数，即最优化问题中度量。初始化个体即初始化*ω*。

需要注意的是，特征空间相当于是个超球体，所以在特征空间中应用智能算法更新*ω*，往往会使得*ω*的位置移动到超球体之外，因此需要用如下投影公式，将被投射到边界之外的点在重新投影回特征空间：

 (2.17)

公式(2.17)即为投影公式，式中*δ*即为特征空间半径。

## 2.3 ENSO事件最快增长初始误差

在文献[3]中，陈磊已经指出ENSO预报的不确定性来自于气候平均态的年循环、ENSO事件本身和初始误差场的结构这三个因素。在本文中，我们采用完美模式假设，即仅考虑初始误差场的结构这一因素。

如前文所述，最快增长初始误差是指发展最快且导致不容忽视的预报误差的一类初始误差。

陈磊文献[11]指出，最快增长初始误差可分为两类：一类为局部条件非线性最优扰动型初始误差，称为type-1型最快增长初始误差，其主要表现为扰动在热带东太平洋从次表层到表层符号一致的自西向东倾斜结构；另一类为全局条件非线性最优扰动型初始误差，称为type-2型最快增长初始误差，其表现为扰动在热带太平洋次表层时的偶极子结构。

对于本文所使用的GFDL CM模式而言，由于其属于全球海气耦合模式，误差发展是全球发展型的，所以我们主要基于type-2型最快增长初始误差设计数值实验。根据美国海洋与大气局（NOAA）的定义—如果Niño3.4区域（170°W~120°W，5°S~5°S）的SST值与1971年到2000年的气候态平均之差连续三个月大于或等于0.5°C，那么就可以认为发生了一次ENSO事件。基于ENSO事件的基本定义，我们将在具体实验部分寻找到导致ENSO事件的初始误差，并根据初始误差的发展，依据CTS-SS智能算法寻找到最快增长初始误差。

# 3 GFDL CM模式样本数据集分析和降维研究

## 3.1 GFDL CM模式样本数据集分析

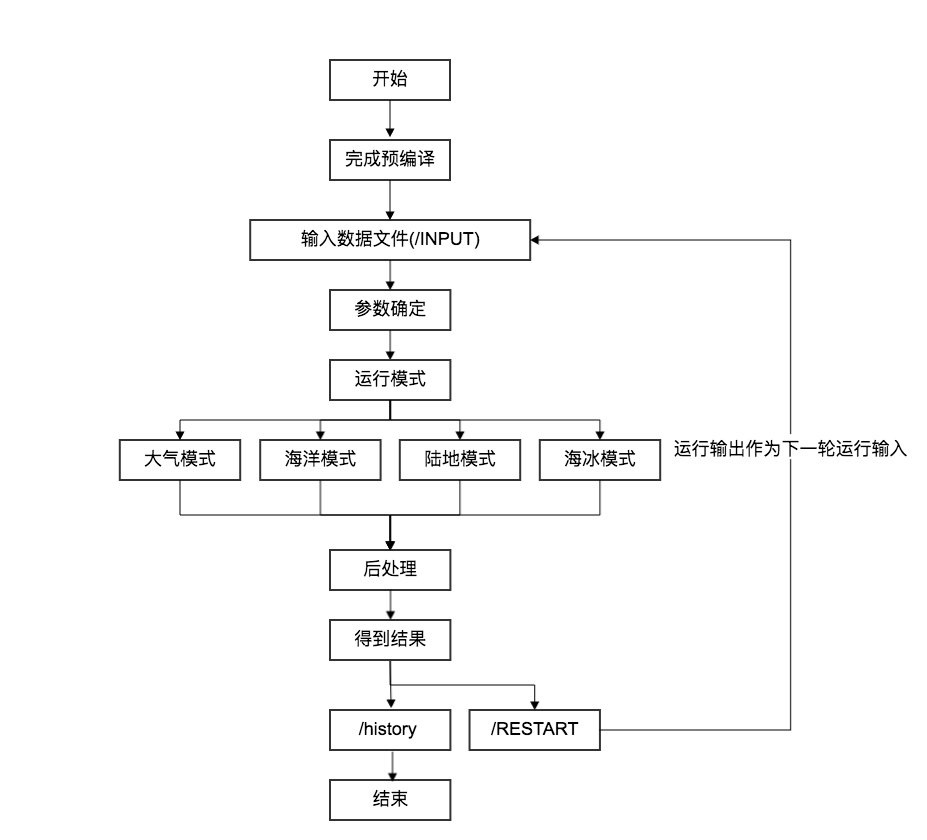
对于本文所选用的GFDL CM2p1模式来说，其主程序采用Fortran语言编写，输入文件独立于主程序，全部存储在/INPUT文件夹，其运行流程如3.1所示：

图3.1 GFDL CM2p1运行流程图

模式结果会存储在/history文件夹下，其包含atmos\_month.nc和ocean\_month.nc两个文件，分别用于存储大气和海洋的月度相关数据。对于ENSO事件来说，我们仅需考虑海洋数据。对应输入数据中与海洋相关的文件结构如3.2所示：

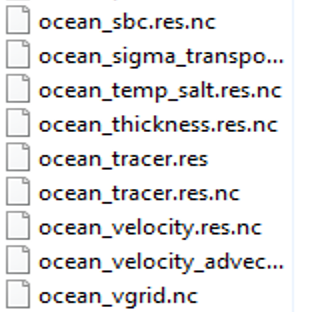
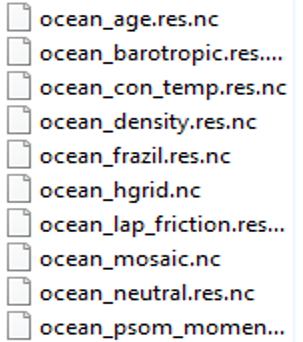


图3.2 海洋模式相关输入文件

作为全球海气耦合模式，理论上来说，在上图中任何海洋相关输入文件中加入扰动，均会对实验结果产生影响。但实验结果表明，对于一些与SST直接关系不强的因素来说（如风向，海风速率等），在其上即使加入较大初始扰动，对最终结果的生成也不会有较大影响（甚至忽略不计）。故本文仅考虑在与SST直接相关的ocean\_temp\_salt.res.nc文件中加入初始扰动，取其中的第一层（海表层），根据模式的分辨率，其初始维度为360\*200\*1（xaxis\_1\* yaxis\_1\* zaxis\_1）。

## 3.2 基于PCA提取主特征

第二章已经提过，在进行智能算法寻优前，需要对模式数据进行降维处理。这里结合GFDL CM模式，分析进行寻优步骤前，需要进行降维预处理的原因：对于

## 3.3 数值实验与分析

# 参考文献

[1] Mu M., Duan W. S. & Wang B., Conditional nonlinear optimal perturbation and its applications. Nonlinear Processes in Geophysics, 2003, 10: 493-501

[2] 穆穆，段晚锁.ENSO 可预报性研究的一个新方法:条件非线性最优扰动.科学通报, 2003,48(7):747-749

[3] 陈磊. ENSO最优前期征兆和最快增长初始误差的相似性及其在识别ENSO目标观测敏感区中的应用[博士学位论文].北京：中国科学院大学, 2015

[4]徐辉. Zebiak-Cane ENSO 预报模式的可预报性问题研究[博士学位论文].北京:中国科学院大气物理研究所,2006

[5] 温仕成. 求解CNOP的算法优化及其应用[博士学位论文].上海：同济大学, 2015

[6]Yuan S., Qian Y., Mu B. (2015) Paralleled Continuous Tabu Search Algorithm with Sine Maps and Staged Strategy for Solving CNOP. In: Wang G., Zomaya A., Martinez Perez G., Li K. (eds) Algorithms and Architectures for Parallel Processing. Lecture Notes in Computer Science, vol 9530. Springer, Cham

[7]孟佳佳，杨宇星. GFDL模式对太平洋海表面温度的年际和年代际变率的模拟评估.海洋科学, 2016

[8]胡一波. 求解约束化问题的几种智能算法[博士学位论文].陕西：西安电子科技大学, 2009

[9]Chen L, Duan W S, Xu H. A SVD-based ensemble projection algorithm for calculating con-

ditional nonlinear optimal perturbation . Science China: earth sciences, on press

[10] Mu B, Zhang L. L.: PCAGA: Principal Component Analysis Based Genetic Algorithm for Solving

Conditional Nonlinear Optimal Perturbation. Submitted to International Joint Conference on Neural

Networks (2015).

[11]穆穆，王强. 条件非线性最优扰动法在大气与海洋目标观测研究中的应用.气象学报，2014