

闫长香, 朱江. 2011. 集合最优插值中的样本选取 [J]. 气候与环境研究, 16 (4): 452–458. Yan Changxiang, Zhu Jiang. 2011. Choice of ensemble members for ensemble optimal interpolation [J]. Climatic and Environmental Research (in Chinese), 16 (4): 452–458.

集合最优插值中的样本选取

闫长香 朱江

中国科学院大气物理研究所, 北京 100029

摘要 背景误差协方差控制了分析向观测调整的幅度以及调整的结构, 所以其对同化分析的质量起着至关重要的作用。对于集合同化方法而言, 样本决定了背景误差协方差的分布。基于 HYCOM 海洋数值模式结果, 针对集合最优插值方法, 探讨了静态样本的选取和更新对背景误差协方差结构分布的影响, 研究表明: 由变量的原始状态数据估计的静态样本会夸大样本相关性, 由扣除季节变化得到的距平数据统计出的静态样本能比较合理地反映背景误差协方差的结构; 在季风控制区, 具有季节变化的样本比静态样本更能适应背景误差协方差随流形分布的特征。同时一系列的同化试验也被实施来进一步调查不同样本对同化分析的影响。

关键词 样本 集合最优插值 背景误差协方差 流依赖

文章编号 1006-9585 (2011) 04-0452-07 **中图分类号** P73, O224 **文献标识码** A

Choice of Ensemble Members for Ensemble Optimal Interpolation

YAN Changxiang and ZHU Jiang

Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029

Abstract The back error covariance controls the magnitude and structure of adjustment to the observations. Therefore, it plays a very important role in the quality of the assimilation. For the ensemble-based data assimilation methods, the ensemble determines the distribution of the background error covariance. This study investigates the impacts of choice and update of static ensemble for Ensemble Optimal Interpolation (EnOI) on the structure of the back-error covariance based on the long-time HYCOM model integration. The result shows that the static ensemble consisting of original model states amplify the ensemble correlation. The ensemble estimated from the model anomalies obtained by subtraction of seasonal signals may reflect a reasonable distribution of the back errors. In the region dominated by the monsoon, the ensemble with season changes describes the flow-dependence of back-error covariance better than the static ensemble. A series of experiments are carried out to validate the effects of different ensembles on the assimilation.

Key words ensemble member, ensemble optimal interpolation, back error covariance, flow dependence

收稿日期 2009-11-18 收到, 2010-12-20 收到修定稿

资助项目 中国科学院方向性项目 KZCX2-YW-143, 国家重点基础研究发展计划 2010CB950401, 国家自然科学基金资助项目 40606008、40437017、40221503、40776011

作者简介 闫长香, 女, 1976 年出生, 副研究员, 主要从事海洋资料同化研究。E-mail: ycxlasg@mail.iap.ac.cn

1 引言

集合同化方法——集合 Kalman 滤波 (Ensemble Kalman Filter, 简称 EnKF) 最初是由 Evensen (1994) 根据 Epstein (1969) 的描述误差统计演变的随机动力预报理论首次提出, 他指出这是一种用蒙特卡罗的短期集合预报方法来估计预报误差协方差的同化方法。它结合了 Kalman 滤波和集合预报的优点, 其背景误差协方差是由一组有限的预报集合来估计, 可以计算具有依赖于流的、随空间、时间变化的、具有复杂结构的预报误差统计量。EnKF 继承了 Kalman 滤波的优点, 避免了扩展 Kalman 滤波的闭合和数值不稳定问题, 已经被广泛应用于大气海洋科学研究 (Houtekamer and Mitchell, 1998; Hamill and Snyder, 2000; Haugen and Evensen, 2002; Koppelman and Rienecker, 2003; Dowell et al., 2004; Leeuwenburgh, 2005)。

基于 EnKF 的思想, Evensen (2003) 提出了一个次优的集合同化方法: 集合最优插值 (EnOI), 它是利用一组有限的、静态的模式状态样本集合来估计背景误差协方差, 也能刻画背景误差的流依赖性。与 EnKF 的实施需要预报多个样本不同, EnOI 只需运行一个样本。这样在计算资源需求方面, EnOI 要远远小于 EnKF, 更易于应用于大气海洋业务化的研究 (Oke et al., 2005; Bertino and Lisæter, 2008)。

背景误差协方差矩阵是决定同化质量的关键因素之一, 所以背景误差模型的建立与构造变得极为重要。对于集合同化方法而言, 其背景误差协方差是由一组有限样本集合来估计的。由于样本的有限性, 必然会带来一些取样误差。Evensen and Van Leeuwen (1996) 指出样本数量比较少, 会导致不满秩的问题。Natvik and Evensen (2003) 调查了样本的数量对同化的影响, 并指出少于 60 的样本集合会降低同化的性能。Houtekamer and Mitchell (1998) 指出当集合样本成员数是 100 时才可以比较准确地描述局地各向同性、斜压相关结构。不同的样本生成方式也会影响背景误差的准确度。对于 EnKF, 初始样本集合通常是通过将扰动叠加在某一给定模式状态而形成

的。欧洲中尺度天气预报中心使用了奇异向量法来产生集合预报的初始条件扰动, 它是关于给定预报轨迹的线性扰动, 这种方法在某些区域会依赖于不同的给定预报轨迹 (Buizza and Palmer, 1995; Molteni et al., 1996; Leutbecher, 2005; Leutbecher and Palmer, 2008)。美国环境预报中心 (NCEP) 采用了 bred 向量法来产生扰动 (Toth and Kalnay, 1993)。Evensen (2003) 和 Wan et al. (2008) 采用了傅里叶变换来生成扰动。这些生成样本方法做不到对样本的预先验证, 评估只能在分析后, 而且评估结果会受分析方案的影响。由于 EnOI 方法在计算需求方面比较经济, 容易实现业务化预报, 在这篇文章里我们主要针对 EnOI 同化方法来探讨样本的选取对于背景误差协方差分布的影响, 以此来预先评估样本集合的优劣。

2 静态样本

EnOI 是利用一组不随时间变化的静态样本集合来统计背景误差协方差信息。样本集合的来源通常是取自多年积分的模式结果。模式结果有多种表现状态, 如变量的原始状态体现的此变量要素的变化, 变量的距平态反映的是此变量偏离平均值的状况。这一节我们主要讨论取样于不同变量状态的样本集合对于刻画背景误差协方差结构的影响。

为了提供样本源, 我们利用混合垂直坐标海洋模式 HYCOM (Bleck, 2002) 在印度洋和西太平洋进行了长时间的积分, 模式水平分辨率约 $1/4^\circ$, 有垂直 22 个分层。HYCOM 海洋模式是由迈阿密等密度面模式 (Bleck and Smith, 1990) 发展而来, 是 3 种垂直混合坐标构成, 在大洋采用的等密度面坐标, 在弱层结的上混合层采用的是 z 坐标, 在浅水区采用的是沿地形的坐标。垂直混合方案采用的是 KPP (K-Profile Parameterization) (Large et al., 1994)。外部强迫场来自欧洲预报中心 (ECMWF) 6 小时一次的海表温度, 湿度和风应力。边界条件以 30 天的时间尺度恢复到气候态。基于模式结果, 我们采取了两种方式来获得静态样本: 第一种是从 20 年 (1986~2005 年) 的 HYCOM 模式月平均数据中选取了 60 个

样本, 60 个样本是从每年 4 个季节中随机选取 3 个组成, 保证了样本集合可以包含季节变化的信息; 第二种是从 20 年 (1986~2005 年) 的 HYCOM 模式月距平数据中以同样的选取方式选取了 60 个样本, 而月距平数据是扣除了季节变化的信息得到的数据。这两种样本集合的最大区别就是来源: 一个是来自月平均态, 一个是来自月距平态。样本集合是否能反映背景误差的流依赖性检验样本集合优劣的一个基本标准, 而误差协方差体现的是一种相关关系, 所以我们就以海表温度的样本相关性来调查这两种样本集合所反映的背景误差的分布特征。

图 1 给出了两种样本的选取对于相关性的影响。很明显, 第一种月平均取法造成的相关性比较大, 点 (1.5°S, 42.5°E) 是位于比较窄的东北向流带上, 其相关较大的区域应该集中在流带上, 但第一种取法在流带外很大海域内仍然存在大于 0.95 的相关关系。而第二种月距平取法所得的样本集合的相关性主要被限制在流轴上, 相对合理地反映出了当地的流系特征。

图 2 给出的是黑潮流轴上某个点 (27°N, 127°E) 的样本相关分布, 在此点附近的流场主要是较强的东北向的黑潮, 有比较明显的流幅, 第一种方法得到的样本除了在流幅上相关性高外, 在远远大于黑潮流幅的海域仍旧与该点存在着非常密切的关系 (相关系数 $r > 0.95$), 这种相关关系显然是不正确的, 这意味着这种方法会带来一些虚假相关。第二种方法反映的样本相关性主要集中在流轴上, 非常好地体现了相关分布随流形变化的特征。

第一种方法获得的样本集合夸大了相关性, 在某些区域甚至错误地刻画了背景误差协方差的结构, 原因可能是第一种方法产生的样本是直接由模式的原始状态构成, 没有剔除季节变化, 季节的相关性占有很大的比重。一般来说, 随着季节的变化, 太阳辐射表现出随纬度变化的特点, 而受太阳辐射影响较大的海表温度在某些区域也呈现出相似的变化, 造成了该区域内海表温度随季节变化的位相比较一致, 从而导致了相关性较大, 这样以来就会掩盖一些中尺度及更小的物理现象, 所以这种样本集合不能很好地反映模式的误差分布。第二种样本集合通过扣除季节变化,

突出了局地区域内的一些现象, 能比较合理地反映背景误差结构随流形的分布。

样本的选取影响了背景误差协方差的分布, 理论上也会影响同化分析的质量。为了进一步调查样本对同化分析的影响, 实施了同化试验。图 3 给出了月平均样本和月距平样本所产生的同化效果, 可以看出, 观测比背景场的温度在该区域普遍偏高, 由于月平均样本所产生的较大相关性遍布了整个区域 (见图 2), 造成了该区域所同化观测的贡献都比较大, 所以较高温度的范围比观测要大, 而且虚假的较大相关也造成了同化的温度比观测要高。而月距平样本由于其反映的合理相关分布, 使得所同化观测对背景场的贡献也比较合理, 从而所产生的同化结果与观测极为相近。

3 静态样本的更新

EnOI 的静态样本意味着在整个同化试验执行的过程中, 样本始终保持不变, 所反映的背景误差协方差结构也始终如一。因为样本反映了背景误差随流形分布的特征, 所以这种保持不变的样本集合同时也意味着流形不随时间变化。然而在某些季风控制海域, 流场是受季风影响的, 流形也会随着季风的改变而改变。上述的分析已经表明由距平数据估计的集合能比较合理地反映背景误差协方差的结构, 下面我们就以这样的样本集合来检查一下静态样本在季风海域是否仍旧有效。

众所周知, 印度洋是受季风影响比较大的海区, 在此区域, 我们从 HYCOM 模式 20 年 (1986~2005 年) 的试验中选取了 60 个月平均距平状态作为静态样本。由于此区域受季风控制, 所以我们也根据季节的变化来选取样本, 即从 20 年模式试验的春季获得了 60 个样本, 组成春季样本, 同样的方法分别获得了夏季、秋季和冬季样本。

图 4 给出了静态样本和冬季样本相关性及流场在赤道印度洋的分布。可以看出, 冬季, 赤道印度洋附近盛行东风, 点 (2°N, 53°E) 位于印度洋东风带内, 此带内的表层流是自东向西的。静态样本反映了背景场误差在此点的相关性范围比较小, 相关的形状接近以经向为长轴的椭圆, 与纬向的流形垂直。冬季样本反映了背景误差协方差的水平扩展比静态样本要大的多, 与西向流的流幅比

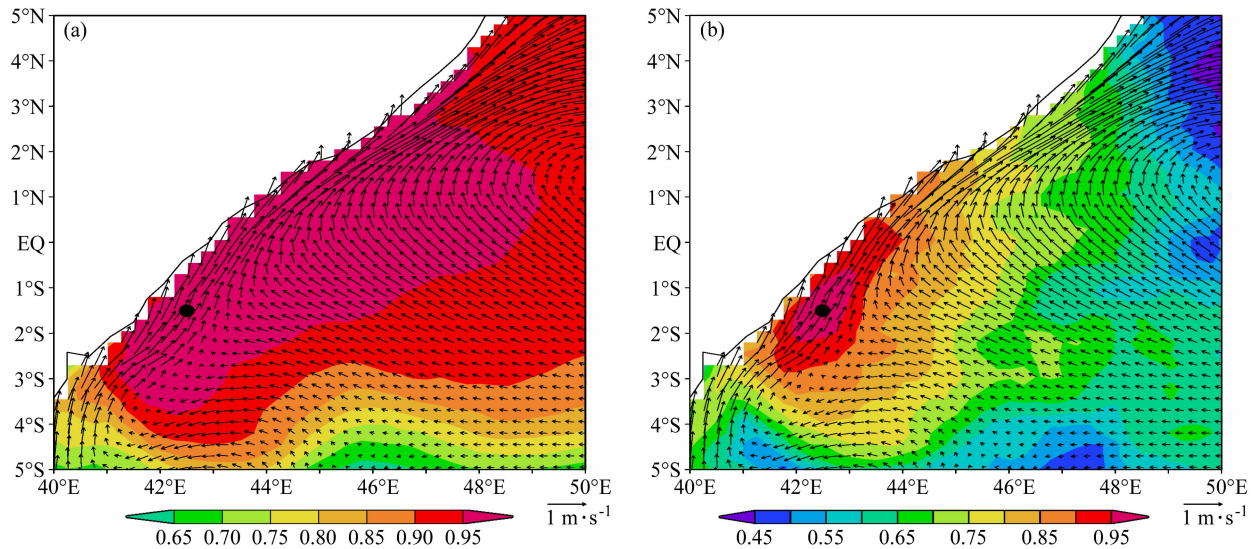


图1 样本海表温度在点(1.5°S, 42.5°E)(以黑色的圆点表示)的相关分布及气候态夏季表层流分布:(a)样本取自月平均数据;(b)样本取自月距平数据

Fig.1 The correlation sea surface temperature from different ensemble at the point (1.5°S, 42.5°E) (denoted by a black dot) and the climatological sea surface current in the summer: (a) Monthly-data-based ensemble; (b) anomaly-data-based ensemble

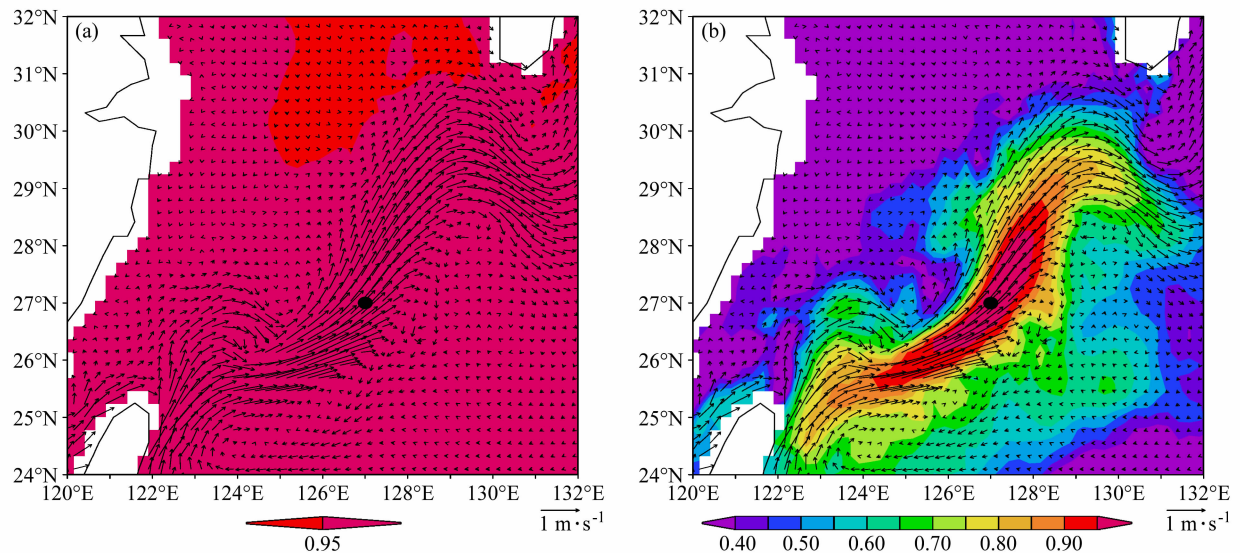


图2 样本海表温度在点(27°N, 127°E)(以黑色的圆点表示)的相关分布及气候态冬季表层流分布:(a)样本取自月平均数据;(b)样本取自月距平数据

Fig.2 The correlation sea surface temperature from different ensemble at the point (27°N, 127°E) (denoted by a black dot) and the climatological sea surface current in the winter: (a) monthly-data-based ensemble; (b) anomaly-data-based ensemble

较接近,体现的相关性分布接近一个以纬向为长轴的椭圆,与自东向西的流场吻合的比较好。

图5是印度洋西边界流带上的点(1.5°S, 42.5°E)的样本相关分布,在夏季,西边界流的流向是东北方向,流带比较窄,流形是沿着海岸,在流带外,与这点的相关性显然会降低,因为流

带外的水来自自东向西的流,与西边界流不属于同一源头。夏季季节样本在西边界流带东部边缘有一个很明显的相关性梯度,在其东边,相关性明显减小,主要的相关关系集中在狭窄的流带上,很好地反映了背景场误差协方差的流形分布。静态样本的相关性在流带外仍旧存在比较大的相关

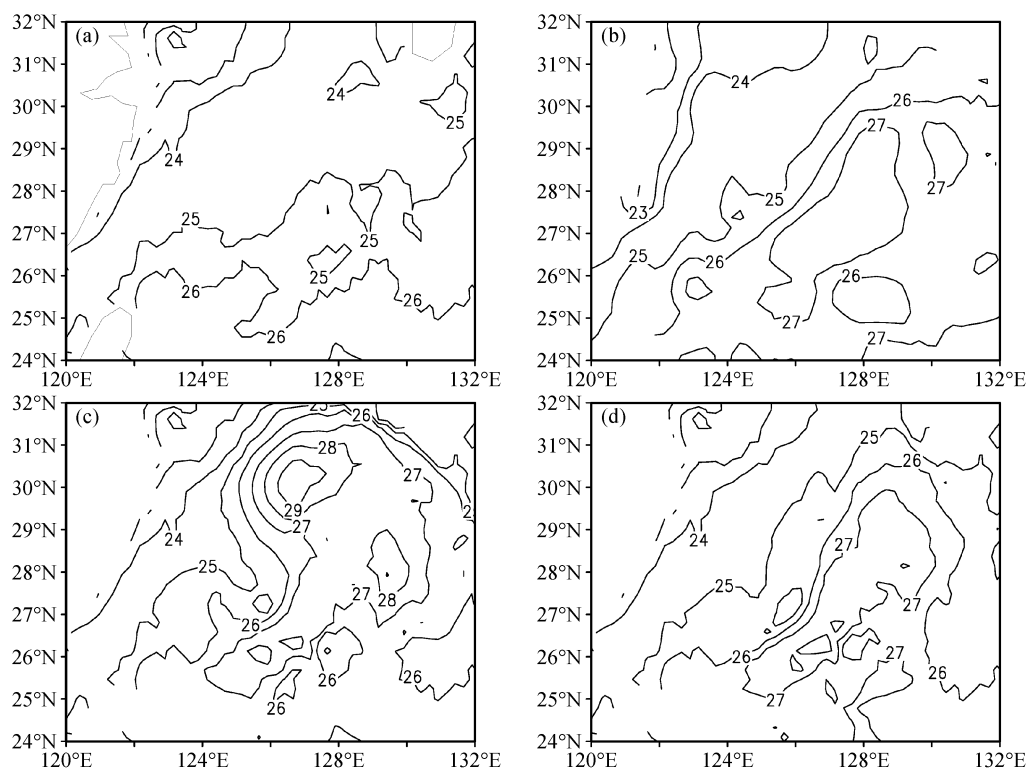


图3 使用不同样本所产生的同化结果(单位:°C)比较:(a)背景场;(b)观测;(c)月平均样本;(d)月距样本

Fig. 3 The assimilated results from different ensembles (unit: °C): (a) Background; (b) observations; (c) monthly-data-based ensemble; (d) anomaly-data-based ensemble

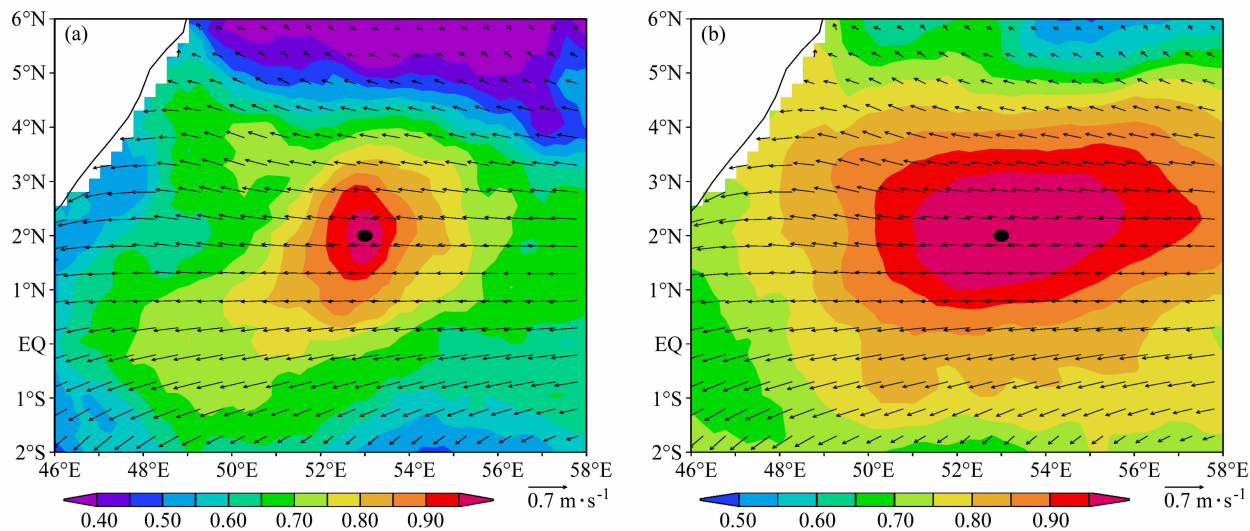


图4 样本海表温度在点(2°N, 53°E)(以黑色的圆点表示)的相关分布及气候态冬季表层流分布:(a)静态样本和冬季表层流;(b)冬季样本和冬季表层流

Fig. 4 The correlation sea surface temperature from different ensembles at the point (2°N, 53°E) (denoted by a black dot) and the climatological sea surface current in the winter; (a) Static ensemble; (b) winter ensemble

性($r > 0.8$),边缘则没有明显的相关性梯度过渡,沿流形分布的特征不显著。

同化试验也被用作检验季节更新样本对同化

的影响,图6给出了同化结果,可以看出在印度洋西边界区域背景场的温度比观测高,两种样本都不同程度地降低了该区域的温度。但是由于静

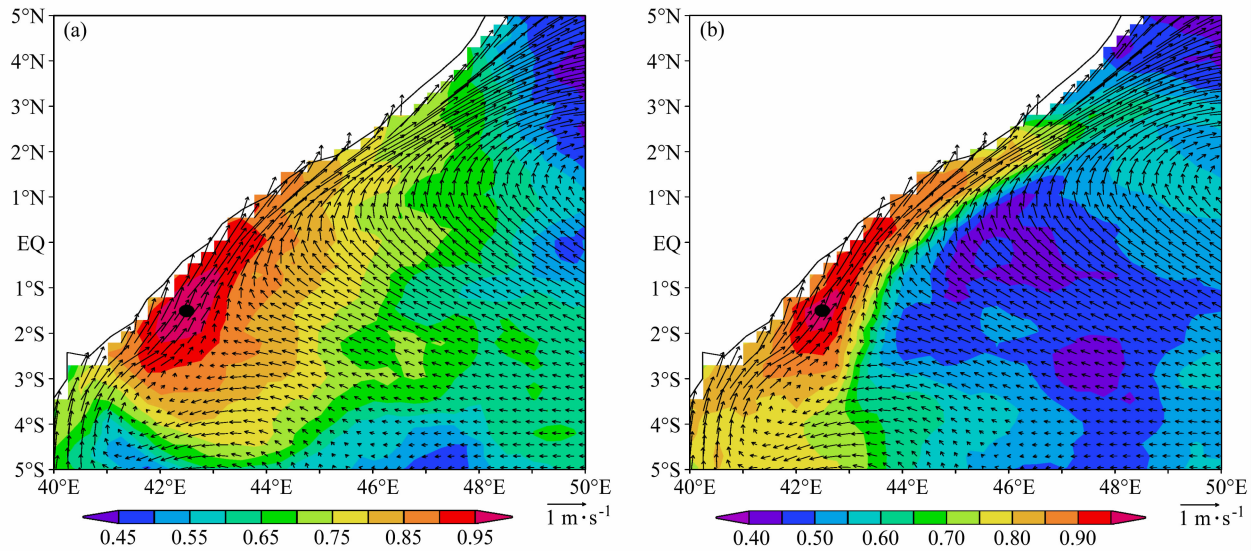


图5 样本海表温度在点 (1.5°S, 42.5°E) (以黑色的圆点表示) 的相关分布及气候态夏季表层流: (a) 静态样本和夏季表层流; (b) 夏季样本和夏季表层流

Fig. 5 The correlation sea surface temperature from different ensembles at the point (1.5°S, 42.5°E) (denoted by a black dot) and the climatological sea surface current in the summer: (a) Static ensemble; (b) summer ensemble

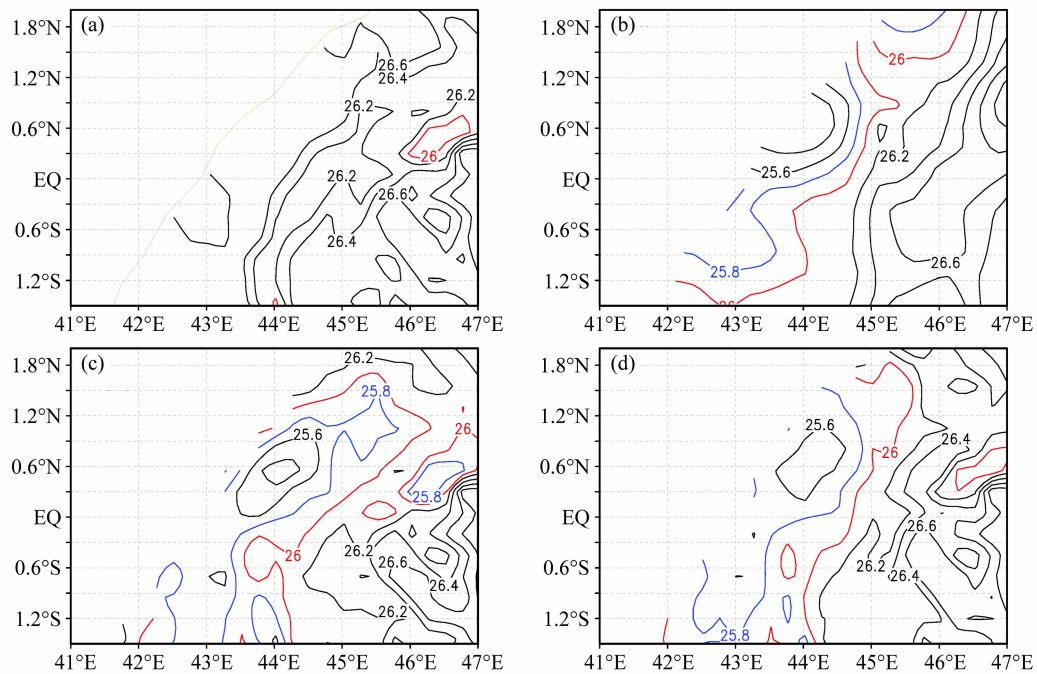


图6 不同样本的同化结果 (单位: °C) 比较: (a) 背景场; (b) 观测; (c) 静态样本; (d) 季节更新样本

Fig. 6 The assimilated results from different ensembles (unit: °C): (a) Background; (b) observations; (c) static ensemble; (d) seasonal ensemble

态样本在该区域的相关性比较大 (图 5), 同化结果也反映出较低温度的水向东延伸范围比较广, 25.8 °C 等温线逼近了 46°E, 而观测只是在 45°E 以内, 26 °C 等温线范围也超出了 46°E, 还可以看

出这种冷水的范围一直向东扩展, 与观测具有较大的差异。而季节更新样本由于其所反映的背景误差协方差的合理结构控制了这种冷水的东扩, 产生了比较合理的效果。

为了提高集合最优插值同化系统的性能,仅使用静态样本是无法实现的,保持样本集合随季节更新更能合理地统计出背景误差协方差矩阵。

4 总结和结论

样本通过影响背景误差协方差的分布决定了最终同化分析的效果,本文针对集合最优插值方法的样本集合进行了同化前的预先评估,发现直接由模式状态构成的样本由于受季节变化信号的影响夸大了相关性的范围,带来了一些虚假相关。扣除季节变化后重新生成的样本可以较好地刻画背景误差协方差结构。同化试验也被实施来调查样本对分析的影响,结果进一步证明了模式状态构成样本的虚假相关也带来了温度的虚假增量,造成了与观测的巨大差异,扣除季节变化后的样本产生了与观测较为接近的效果。

由于集合最优插值样本不随时间变化的静态性,造成在某些区域不能合理地反映背景误差的流依赖分布特征。本文针对季风控制海域——印度洋检查了静态样本所反映的相关特征,发现静态样本体现不出随季节变化的流场分布,当样本随季节更新时,就可以估计出随流形分布的背景误差协方差结构,从而可以产生合理的同化结果,这在同化试验中也得到了验证。

参考文献 (References)

- Bertino L, Lisæter K A. 2008. The TOPAZ monitoring and prediction system for the Atlantic and Arctic Oceans [J]. *Journal of Operational Oceanography*, 2: 15–18.
- Bleck R. 2002. An oceanic general circulation model framed in hybrid isopycnal-cartesian coordinates [J]. *Ocean Modelling*, 4: 55–88.
- Bleck R, Smith L T. 1990. A wind-driven isopycnal coordinate model of the north and equatorial Atlantic Ocean 1. Model development and supporting experiments [J]. *J. Geophys. Res.*, 95 (C3): 3273–3285.
- Buizza R, Palmer T N. 1995. The singular-vector structure of the atmospheric global circulation [J]. *J. Atmos. Sci.*, 52: 1434–1456.
- Dowell D C, Zhang F, Wicker L J, et al. 2004. Wind and thermodynamic retrievals in the 17 May 1981 Arcadia, Oklahoma supercell: Ensemble Kalman filter experiments [J]. *Mon. Wea. Rev.*, 132: 1982–2005.
- Epstein E S. 1969. Stochastic dynamic prediction [J]. *Tellus (A)*, 21: 739–759.
- Evensen G. 1994. Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics [J]. *J. Geophys. Res.*, 99 (C5): 10143–10162.
- Evensen G, Van Leeuwen P J. 1996. Assimilation of Geosat altimeter data for the Agulhas current using the ensemble Kalman filter with a quasigeostrophic model [J]. *Mon. Wea. Rev.*, 124: 85–96.
- Evensen G. 2003. The ensemble Kalman filter: Theoretical formulation and practical implementation [J]. *Ocean Dynamics*, 53: 343–367.
- Hamill T M, Snyder C. 2000. A hybrid ensemble Kalman filter/3d-variational analysis scheme [J]. *Mon. Wea. Rev.*, 128: 2905–2919.
- Haugen V E, Evensen G. 2002. Assimilation of SLA and SST data into an OGCM for the Indian ocean [J]. *Ocean Dynamics*, 52: 133–151.
- Houtekamer P L, Mitchell H L. 1998. Data assimilation using an ensemble Kalman filter technique [J]. *Mon. Wea. Rev.*, 126: 796–811.
- Keppenne C L, Rienecker M M. 2003. Assimilation of temperature into an isopycnal ocean general circulation model using a parallel ensemble Kalman filter [J]. *Journal of Marine Systems*, 40: 363–380.
- Large W G, McWilliams J C, Doney S C. 1994. Oceanic vertical mixing: A review and a model with a nonlocal boundary layer parameterization [J]. *Reviews of Geophysics*, 32: 363–403.
- Leeuwenburgh O. 2005. Assimilation of along-track altimeter data in the tropical Pacific region of a global OGCM ensemble [J]. *Quart. J. Roy. Meteor. Soc.*, 131: 2455–2472.
- Leutbecher M. 2005. On ensemble prediction using singular vectors started from forecasts [J]. *Mon. Wea. Rev.*, 133: 3038–3046.
- Leutbecher M, Palmer T N. 2008. Ensemble forecasting [J]. *Journal of Computational Physics*, 227: 3515–3539.
- Molteni F, Buizza R, Palmer T N, et al. 1996. The ECMWF ensemble prediction system: methodology and validation [J]. *Quar. J. Roy. Meteor. Soc.*, 122: 73–119.
- Natvik L J, Evensen G. 2003. Assimilation of ocean colour data into a biochemical model of the North Atlantic: Part 1. Data assimilation experiments [J]. *Journal of Marine Systems*, 40–41: 127–153.
- Oke P R, Chiller A S, Griffin D A, et al. 2005. Ensemble data assimilation for an eddy-resolving ocean model of the Australian Region [J]. *Quart. J. Roy. Meteor. Soc.*, 131: 3301–3311.
- Toth Z, Kalnay E. 1993. Ensemble forecasting at NMC: The generation of perturbations [J]. *Bull. Amer. Meteor. Soc.*, 74: 2317–2330.
- Wan L, Zhu J, Laurent B, et al. 2008. Initial ensemble generation and validation for ocean data assimilation using HYCOM in the Pacific [J]. *Ocean Dynamics*, 58: 81–99.