

集合背景误差方差中小波阈值去噪方法研究及试验

刘柏年 皇群博 张卫民 任开军 曹小群 赵军

Investigation and experiments of wavelet thresholding in ensemble-based background error variance

Liu Bai-Nian Huang Qun-Bo Zhang Wei-Min Ren Kai-Jun Cao Xiao-Qun Zhao Jun

引用信息 Citation: *Acta Physica Sinica*, 66, 020505 (2017) DOI: 10.7498/aps.66.020505

在线阅读 View online: <http://dx.doi.org/10.7498/aps.66.020505>

当期内容 View table of contents: <http://wulixb.iphy.ac.cn/CN/Y2017/V66/I2>

您可能感兴趣的其他文章

Articles you may be interested in

周期波动性对金融市场稳定性的影响

[Influence of periodic volatility on the stability of financial market](#)

物理学报.2017, 66(4): 040501 <http://dx.doi.org/10.7498/aps.66.040501>

反馈控制棘轮的定向输运效率研究

[Investigation on the directed transport efficiency of feedback-control ratchet](#)

物理学报.2017, 66(1): 010501 <http://dx.doi.org/10.7498/aps.66.010501>

随机参激下 Duffing-Rayleigh 碰撞振动系统的 P-分岔分析

[P-bifurcations of a Duffing-Rayleigh vibroimpact system under stochastic parametric excitation](#)

物理学报.2016, 65(21): 210501 <http://dx.doi.org/10.7498/aps.65.210501>

具有非线性阻尼涨落的线性谐振子的随机共振

[Stochastic resonance of a linear harmonic oscillator with non-linear damping fluctuation](#)

物理学报.2016, 65(6): 060501 <http://dx.doi.org/10.7498/aps.65.060501>

二值噪声激励下欠阻尼周期势系统的随机共振

[Stochastic resonance in periodic potential driven by dichotomous noise](#)

物理学报.2015, 64(24): 240502 <http://dx.doi.org/10.7498/aps.64.240502>

集合背景误差方差中小波阈值去噪方法 研究及试验*

刘柏年^{1)2)†} 皇群博¹⁾²⁾ 张卫民¹⁾ 任开军¹⁾ 曹小群¹⁾ 赵军¹⁾

1)(国防科技大学海洋科学与工程研究院, 长沙 410073)

2)(国防科技大学计算机学院, 长沙 410073)

(2016年3月26日收到; 2016年10月20日收到修改稿)

背景误差方差的集合估计值中带有大量采样噪声, 在应用之前需进行降噪处理. 区别于一般的高斯白噪声, 采样噪声具有空间和尺度相关性, 部分尺度上的噪声能级远大于平均能级. 本文针对背景误差方差中采样噪声的特征, 引入小波阈值去噪方法, 并根据截断余项的小波系数分布特征发展了一种计算代价很小, 能自动修正阈值的算法. 一维理想试验结果表明, 该方法能滤除大量采样噪声, 提高背景误差方差估计值的精度. 相对于原来的小波阈值方法, 修正阈值后减少了因部分尺度上噪声能级过大导致的残差, 去噪后的RMSE减少了13.28%. 将该方法应用在实际的集合资料同化系统中, 结果表明, 小波阈值方法优于谱方法, 阈值修正后能在不影响信号的前提下增大小波去噪强度.

关键词: 小波, 集合资料同化, 背景误差方差, 去噪

PACS: 05.40.-a, 05.40.Ca, 02.60.-x, 43.60.Hj

DOI: 10.7498/aps.66.020505

1 引言

数值天气预报可用一个非线性方程组表示, 初始场的精度在很大程度上决定了预报的准确性. 为了得到高质量的初始场, 变分资料同化需要融合各种观测资料和背景信息^[1]. 其中背景误差协方差矩阵(\mathbf{B}) 在资料同化中起到了误差结构分布、信息传播、权重分配等作用, 因此资料同化的首要任务是精确定义出 \mathbf{B} ^[2]. 在信息量不足的前提下, 采用静态的、均一的和各向同性的 \mathbf{B} 模型能有效节省大量计算资源并简化运算, 但无法反映出大气的时空变化特征^[3,4]. 因此大量研究开始尝试采用不同方法来获取流依赖 \mathbf{B} ^[5-7]. 高性能计算机和大规模并行计算为集合资料同化(EDA) 的发展提供了重要支撑, EDA 通过扰动观测资料、边界条件和预报模式构造一组能够反映 \mathbf{B} 统计信息的低分辨率扰动

集合, 并以此估算流依赖的 \mathbf{B} . EDA 被认为是解决当前确定性资料同化问题的一种有效途径. 国际上几大数值预报业务中心也正在开发和完善EDA 业务系统, 如欧洲中期天气预报中心(ECMWF)^[8]、法国气象局^[9,10]等.

EDA 的关键是选择合适的集合成员个数. 由于EDA 成员是通过随机扰动方法构造的, 利用有限EDA 成员统计背景误差方差时, 估计值中通常包含有随机采样噪声, 估计精度与集合成员数的均方根成正比^[11]. 一方面, 业务时效性要求EDA 在确定性资料同化之前完成所有成员的计算和 \mathbf{B} 的统计, 在这一前提下, 成员个数不能无限扩大; 另一方面, EDA 通过随机扰动方法构造出多个能够反映 \mathbf{B} 统计信息的扰动集合, 集合成员个数越多, 引入的随机采样噪声越小, 统计值就越精确, 因此成员个数不能过少. 综合以上分析, 一种解决方案是扩大计算资源和并行规模并利用FPGA, GPU等

* 国家自然科学基金(批准号: 41375113, 41475094, 41305101, 41605070)资助的课题.

† 通信作者. E-mail: bnliu@nudt.edu.cn

加速计算技术, 增加尽可能多的样本数, 但由于估计精度的低收敛性限制了这一方案的实际效率; 另一种行之有效的方案是发展计算量小的去噪方法, 通过滤除随机采样噪声来提高估计精度.

针对以上需求, Raynaud等^[12]于2008年根据噪声和信号的特征尺度发展了一种基于空间平均的低通滤波方法, 以空间加权的方式滤除具有频率高、尺度小的随机采样噪声. 经过上述方法处理后的10个样本背景误差方差估计值在精度上与50个样本的直接统计结果相当, 但这种处理噪声方法的效能受大气变量、高度层等因素影响, 且平滑的空间尺度需要根据天气形势进行人为调整和优化, 不适合业务化. 2009年, Raynaud等^[10]改进了滤波算法, 基于气候态 \mathbf{B} 矩阵得到了近似噪声能量谱, 发展了一种能自动估算截断波数的谱滤波方法. 其基本思想是根据信号和噪声能量在谱空间的分布特征定位出信噪分离的截断波数, 由于噪声能量主要集中在波数较小的小尺度空间上, 因此构造一个低通滤波器, 滤除大于截断波数的高频小尺度噪声, 对小于截断波数的大尺度噪声, 则采用类似于Wiener的线性处理方法^[13]. 该方案一方面降低了对集合样本数的要求, 减少了计算量, 另一方面又显著提高了集合方差的估算精度^[8,10,12,14]. 文献^[15]等将该方法成功应用到了我国自主开发的集合资料同化系统中, 使10个成员集合估计值的滤波结果好于30个成员集合估计值. 但是, 谱滤波方法也存在一定的局限性. 首先, 噪声能量谱是利用气候态 \mathbf{B} 估算得到的, 精度和正确完全由 \mathbf{B} 和集合成员数决定. 由于对 \mathbf{B} 采用了静态近似处理, 由此统计的随机采样噪声能量谱也是静态的定常量, 不适用于快速发展的天气系统, 如台风和深对流天气; 其次, 相同尺度上作用的滤波系数是相同的, 这等同于对格点空间中相同尺度的平滑处理在全球范围都是一致的, 不能反映出信号的局地特征.

针对上述方法的局限性, 本文在集合背景误差方差中引入了具有谱和空间局地化特性的小波阈值去噪方法^[16]. 根据小波信号特征, 通过迭代算法得到阈值, 避免了谱滤波方法中噪声能量谱的近似处理和静态假设. 在此基础上, 根据集合背景误差方差中采样噪声具有的空间和尺度相关特征, 设计了一种能自动修正阈值的改进算法, 可减少因部分尺度上噪声能级过大导致的残差, 进而改进滤波效

果. 最后在一维理想模型和实际的集合资料同化系统中测试了该方法的鲁棒性.

2 去噪原理

2.1 谱滤波原理

EDA通过扰动观测资料、海表温度和模式物理倾向项来表征观测、边界及模式物理过程的不确定性. 多个表征误差分布特征的扰动初值, 经模式积分得到的预报集合可统计出流依赖的背景误差^[17-19]. 但是EDA作为一种纯粹的随机方法, 有限个成员统计得到的结果中包含了随机采样噪声. 假定背景误差协方差服从高维的高斯分布, 则采样噪声的协方差 $E[G^\epsilon(i)G^\epsilon(j)]$ 和集合误差协方差矩阵 $\mathbf{E}[\tilde{\mathbf{B}}_{ij}]$ 近似满足以下关系^[10,20]:

$$\begin{cases} E[G^\epsilon(i)G^\epsilon(j)] = \frac{2}{N-1}(E[\tilde{\mathbf{B}}_{ij}])^2, \\ G^\epsilon(i) = (\tilde{\mathbf{B}}_{ii} - E[\tilde{\mathbf{B}}_{ii}]), \end{cases} \quad (1)$$

其中 $E[\cdot]$ 表示数学期望, $\tilde{\mathbf{B}}_{ii}$ 是格点 i 的背景误差方差的集合估计值, $G^\epsilon(i)$ 是格点 i 上的采样噪声, N 是集合样本个数. 由此计算得到噪声与背景误差之间的Daley^[21]长度尺度关系满足:

$$\begin{cases} L_{G^\epsilon}^2(i) = \left(-\frac{d^2 c^2}{dr^2} \right)^{-1} \Big|_{i,r=0} = \frac{L_{\epsilon^b}^2(i)}{2}, \\ L_{G^\epsilon}(i) = \frac{L_{\epsilon^b}(i)}{\sqrt{2}}, \end{cases} \quad (2)$$

$L_{G^\epsilon}(i)$, $L_{\epsilon^b}(i)$ 分别为格点 i 上噪声和背景误差长度尺度, c 为相关函数, r 为格点之间的距离. 上式表明, 随机采样噪声的相关长度尺度总小于背景误差方差的长度尺度. 误差方差倾向于在更大长度尺度上变化. 谱滤波方法根据这一特性, 首先利用等式(1)近似估算出随机采样噪声的能量谱 \mathbf{S}^ϵ , 然后将方差的集合估计值 $\tilde{\mathbf{V}}$ 转换到谱空间, 并以 $\tilde{\mathbf{S}}$ (谱滤波的原始输入信号) 表示. 由于随机采样噪声的相关长度尺度总小于背景误差方差的长度尺度 (见等式(2)), 因此, 可利用以下低通滤波器^[10]来削弱输入信号中的随机采样噪声的强度:

$$\rho(n_p) = \left[\cos \left\{ 0.5\pi \frac{\min(n_p, N_{\text{trunc}})}{N_{\text{trunc}}} \right\} \right]^2, \quad (3)$$

其中 $\rho(n)$ 是低通滤波器的滤波系数, n_p 为谱空间中的波数, 截断波数 N_{trunc} 为信号和噪声能量谱开始分离的波数. 谱滤波的处理流程如图1所示.

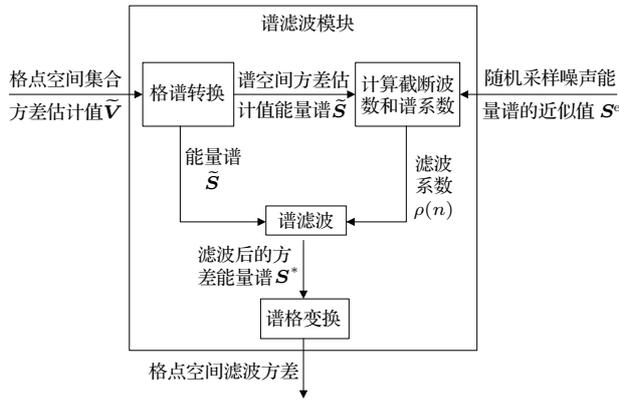


图1 谱滤波模块处理流程

Fig. 1. Processes of the spectral filtering.

谱滤波方法能根据谱空间中信号、噪声能量谱的分布情况自动计算出截断波数并实现滤波, 具有计算成本低、滤波效率高等优点 [8,10,12]. 当信号具有空间缓变、各向同性、短距离相关特性时, 滤波效率更为显著. 目前 ECMWF 和 Mete-France 的业务化集合资料同化系统都采用了这种方法 [8,10,12,14]. 但是该方法也存在一定的局限性: 1) 谱滤波方法采用了确定性同化系统中的简化、静态 B 模型统计噪声能量谱, 并引入了一定的假设和近似, 这种处理方式降低了噪声能量谱的估计精度; 2) 噪声能量谱的统计完全独立于输入信号, 且作为一个时常量应用到后续的滤波过程中, 不能反映背景误差方差及随机采样噪声随天气形势变化的特征; 3) 谱滤波的实质是在相同波数上, 作用一个滤波系数. 在格点空间, 这等同于在同一尺度空间上乘上一个全球相同的加权平均系数. 这种各向同性、均一的滤波方式会忽略信号的局地变化, 一些十分重要的特征信号可能被平滑或弱化.

2.2 小波阈值去噪原理

小波阈值去噪是一种常用的信号处理方法, 它将小于阈值的小波系数置零后, 再重构出信号, 因此阈值的选取是一个十分关键的问题. Donoho 和 Johnstone [16] 将阈值表达为集合成员个数和噪声方差的函数. 在缺乏信号的先验信息条件下, 小波阈值能减少定义在 Holder 和 Besov 空间中的误差, MAD 方法则是将小波的最小尺度系数模的中位数绝对偏差 (MAD) 作为噪声的能级 [20]. 对于高斯白

噪声, 可采用以下迭代算法得到最佳阈值 [22], 具体步骤如下:

1) 将原始信号 X 转换到小波空间 \tilde{X} , 并划分为噪声 \tilde{X}^e 和信号 \tilde{X}^* 两部分, 在循环迭代开始时置 $\tilde{X}^e = \tilde{X}$, 即将所有信号当成噪声;

2) 计算第 k 步的噪声方差 $\sigma_{W,k}^2 = \frac{1}{n} \sum_{i,j} |\tilde{X}_{i,j}^e|^2$

和阈值 $T_k = \sigma_{W,k} \sqrt{2 \ln(n)}$, n 表示信号长度, $\tilde{X}_{i,j}^e$ 表示尺度 j 上第 i 个噪声小波系数;

3) 进行小波阈值截断

$$\tilde{X}_{i,j}^* = \rho_T(\tilde{X}_{i,j}^e) = \begin{cases} \tilde{X}_{i,j}^e, & |\tilde{X}_{i,j}^e| > T, \\ 0, & |\tilde{X}_{i,j}^e| \leq T, \end{cases}$$

并计算出新的噪声 $\tilde{X}_{i,j}^e = (1 - \rho_T)(\tilde{X}_{i,j}^e)$;

4) 重复 2), 3) 步直到满足条件 $|\sigma_{W,k}^2 - \sigma_{W,k+1}^2| \leq \Delta$, 其中 Δ 表示既定的精度, 如 $\Delta = 10^{-5}$.

采用上述迭代方法能快速收敛得到噪声方差 $\sigma_W^2 = \sum_{i,j} |\tilde{X}_{i,j}^e|^2$ 和阈值 $T_A = \sigma_W \sqrt{2 \ln(n)}$, 滤波后可通过 $X' = \sum_{i,j} \tilde{X}_{i,j}^* \psi_{i,j}$ 重构得到物理空间信号.

迭代算法如图 2 所示, 各个方向的白噪声相互正交于圆内. 圆对应最大的噪声, 即阈值. 由于迭代开始时噪声方差非常大, 大部分小波系数都囊括在圆内. 经第一次去噪后, 排除了那些模大于阈值 T_0 的系数, 余下的系数则构成新的噪声, 用于计算下一个阈值 T_1 .

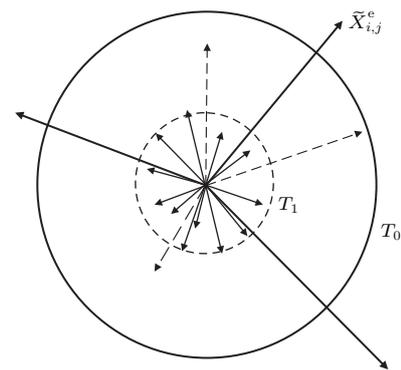


图2 噪声方差和阈值迭代算法示意图 粗实线圆代表 T_0 , 虚线代表 T_1 , 箭头表示小波系数

Fig. 2. Illustration of the recursive algorithm for estimating the noise variance and the threshold. The estimated thresholds T and T_1 are represented by the bold and dashed spheres respectively. The arrows represent a selection of wavelet coefficients $\tilde{X}_{i,j}$.

小波阈值方法等价于用一个适合输入信号局地变化特征的滤波器来估计真实信号. 其事实依据是, 函数 f 在尺度 j 和位置 $x_j(i)$ 的小波转换, 度量了函数 f 在 $x_j(i)$ 附近的变率, $x_j(i)$ 的尺度与 j 成正比. 快速变化的信号能使小尺度上的小波系数变大. 小波阈值方法通过将小于阈值 T 的系数置 0, 构造一个依赖于小波系数的自适应滤波器. 较大的系数 $|\tilde{X}_{i,j}| > T$ 表明函数 f 在小尺度范围内的变化剧烈, 这部分系数予以保留避免平滑掉. 而满足 $|\tilde{X}_{i,j}| \leq T$ 的系数则表示 f 变化较为平缓, 通过置 0 后滤除.

背景误差方差的集合估计值中, 噪声具有空间和尺度相关性, 不能简单地当作高斯白噪声处理. 谱空间中信噪比随波数的变化可以用简单的正弦函数表示^[10], 基于此构造的形如等(2)式的低通滤波器可以有效地减少采样噪声在各个尺度上的绝对量. 在小波空间中为了滤除具有相关性的噪声, 一种处理方法是对每个尺度使用不同的阈值^[23]:

$$T_N(j) = \sigma(j)\sqrt{2\ln(n_j)}, \quad (4)$$

其中 $\sigma(j)$ 表示尺度上的标准偏差, $n_j = 2^j$ 表示尺度 j 上小波系数个数. 由于个别尺度对应的 n_j 过小, 统计得到的 $\sigma(j)$ 具有较大的误差. 另一种处理方法是采用新的全局阈值, 如对迭代阈值 T_A 进行调整:

$$T_P = \alpha \times \sigma_{w'}\sqrt{2\ln(n)}, \quad \alpha \geq 1, \quad (5)$$

通过改变 α 值, 使满足 $\sigma(j) > \sigma_{w'}$ 上的部分过大的噪声小波系数也能被阈值 T_P 置 0. Pannekoucke 等^[24] 给出了初步实验结果验证了该方法的有效性, 但由于噪声具有非高斯性, 理论建模和推导都十分复杂, 目前没有很好的方法估算 α .

阈值 T_P 在调整过程中, 需要综合考虑对立因素: 既要保证部分能级较大的噪声能被阈值置 0, 这要求 T_P 要足够大; 同时 T_P 又不能无限大, 确保信号的小波系数不受 T_P 的影响. 如图 3 所示, 小圆表示高斯白噪声, 椭圆表示具有空间和尺度相关的非高斯噪声, 在某些尺度上噪声能级大于平均能级, 表现为椭圆的两端溢出了小圆范围. 大圆则是以最大噪声能级为半径. 如果简单地将背景误差方差中的采样噪声当作高斯白噪声处理, 采用 $\sigma = \sigma_w$ 计算得到的阈值将明显偏弱(小圆只包含了椭圆的一部分), 大量噪声仍未被滤除; 但如果采

用 $\sigma = \max(\sigma(j))$, 则滤波偏强, 部分信号被滤除, 导致失真. 一种可行的处理方式如下:

首先, 将噪声 \tilde{X}^e 划分为高斯 \tilde{X}^{Ge} 和非高斯项 \tilde{X}^{NGe} , $\tilde{X}^e = \tilde{X}^{Ge} + \tilde{X}^{NGe}$. 小波变换后, 由于非高斯噪声 \tilde{X}^{NGe} 的小波系数个数远小于高斯噪声 \tilde{X}^{Ge} 的小波系数, 因此 \tilde{X}^{Ge} 的噪声统计特征如标准差 $\sigma(\tilde{X}_{i,j}^{Ge})$ 可作为 $\sigma(\tilde{X}_{i,j}^e)$ 的一种近似值. 其次, 仍采用 Azzalini 的方法迭代得到初步的阈值, 记为 T_A , 截断余项满足 $\tilde{R}_{i,j} \in \|X_{i,j}\| \leq T_A$, 该项包含了少数的信号小波系数 $\|X_{i,j}^*\| \leq T_A$ 、非高斯项噪声 \tilde{X}^{NGe} 的小波系数 $\|\tilde{X}_{i,j}^{NGe}\| \leq T_A$ 和绝大部分高斯噪声 \tilde{X}^{Ge} 的小波系数 $\|\tilde{X}_{i,j}^{Ge}\| \leq T_A$. 然后, 将截断余项 \tilde{R} 的标准差 $\sigma_{w'}$ 作为 σ_w 的初猜值, 则位于区间 $[-\sigma_{w'}\sqrt{2\ln n}, \sigma_{w'}\sqrt{2\ln n}]$ 的部分非高斯项噪声小波系数 $\|\tilde{X}_{i,j}^{NGe}\| > T_A$ 看成是服从截断余项分布的小概率事件. 最后, 为了能将少数系数 $\|\tilde{X}_{i,j}^{NGe}\| > T_A$ 包含进来, 在修正阈值 T_A 时采用了以下经验公式

$$T_S = \max(2\sigma_{w'}\sqrt{2\ln(n)}, T_A), \quad (6)$$

其中截断余项的标准差 $\sigma_{w'}$ 计算方法为

$$\sigma_{w'} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_{i,j} - \bar{X}_{i,j})^2},$$

$$\|X_{i,j}\| \leq T_A,$$

其中 $\bar{X}_{i,j}$ 为 $X_{i,j}$ 的数学期望. 根据概率理论, 等式(6)形成的区间 $[-2\sigma_{w'}\sqrt{2\ln n}, 2\sigma_{w'}\sqrt{2\ln n}]$ 能使约 95% 的噪声系数落入到此区间. 一方面尽可能多的把非高斯项噪声小波系数包含进来, 另一方面也避免 T_S 过大将过多的信号小波系数被置 0 导致信号失真. 以下将通过一维集合资料同化系统模型来验证该方法的合理性.

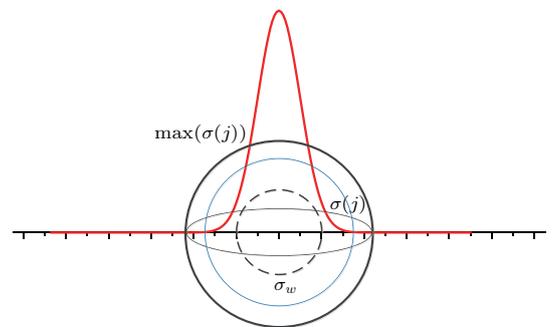


图 3 高斯白噪声和具有空间、尺度相关噪声的示意图
Fig. 3. Graphical representation of white and correlated noises.

3 滤波试验

3.1 试验设置

将赤道纬圈展开为一维等距的 $n = 401$ 个格点, 构造一个简单的 \mathbf{B} 矩阵, 其方差 \mathbf{V}^* 随空间缓慢变化, 并采用均一、各向同性的高斯函数作为相关函数:

$$c(i, i+r) = \exp\left(-\frac{r^2}{2L_{\epsilon^b}^2}\right), \quad (7)$$

其中 i 为纬圈上的格点序号, r 是格点间距, 相关长度尺度 L_{ϵ^b} 设为 300 km. 集合样本数 N 取 50 个. 为能真实模拟采样噪声的特征, 这里采用与实际系统相同的随机方法估计背景误差标准偏差. 首先随机生成 50 个服从 $\mathcal{N}(0, \mathbf{I})$ 的随机矢量 $\alpha_k, k = 1, \dots, N$, 其次将 $\mathbf{B}^{1/2}$ 与每个矢量 α_k 相乘得到样本 ϵ_k^b , 此时 ϵ_k^b 满足 $\mathcal{N}(0, \mathbf{B}^{1/2})$ 分布, 最后计算出 ϵ_k^b 的方差 \mathbf{V} , 即为背景误差方差的集合估计值, 方差计算方法如下:

$$\begin{cases} \alpha_k \sim \mathcal{N}(0, \mathbf{I}), \\ \epsilon_k^b = \mathbf{B}^{1/2} \cdot \alpha_k, \quad k = 1, \dots, N, \end{cases} \quad (8)$$

$$\begin{cases} V_i = \frac{1}{N-1} \sum_{k=1}^N (\epsilon_{k,i}^b - \overline{\epsilon_i^b})^2, \\ \overline{\epsilon_i^b} = \frac{1}{N} \sum_{k=1}^N \epsilon_{k,i}^b. \end{cases} \quad (9)$$

因 Coif5 具有良好的正交和双正交特性, 且在时域和频域都具有良好的紧支撑和消失矩, 试验中选用 Matlab 自带的 Coif5 正交小波模块, 实现信号分解和重构.

3.2 试验结果分析

图 4 给出了预定义的背景误差方差真值, 即 \mathbf{B} 的对角元素, 以及 50 个样本的方差估计值. 在大部分区域, 预定义的方差真值变化较为平缓, 而在第 150, 250 个格点附近出现了陡峭变化. 这种变化可表征风暴、深对流和台风等剧烈天气事件的背景误差. 从图 4 可以看出, 尽管集合平均能清晰反映出真实方差的大尺度特征, 但是存在以信号为中心上下波动的小尺度采样噪声. 在方差偏大的格点附近, 采样噪声也明显偏大.

采用上述方法对集合估计值进行去噪, 得到图 5 中的结果. 对比可以看出, 谱滤波(绿线)结果

最为平滑, 但两个特征信号同时也被大幅度削弱. 其原因在于谱滤波本质上属于一种空间加权平均滤波, 平滑范围由噪声和信号的特征尺度决定. 权重系数是一个全局量, 不能随空间位置变化, 因此谱滤波无法刻画出信号的局地特征. 当特征信号的尺度与噪声相当时, 信号被当作噪声处理, 导致信号失真. 未经修正的小波去噪结果, 如图 5 蓝线所示, 与真值之间的 RMSE(0.77) 小于谱方法(1.44), 但是去噪后信号波动最为剧烈. 从某种意义上说, 这是去噪强度偏弱的一种表象, 即迭代方法得到的阈值 $T_A = 1.42$ 偏小. 主要原因在于尺度 $j = 2, 3$ 上的噪声标准偏差 $\sigma_{j=2} = 6.33, \sigma_{j=3} = 2.95$ 均大

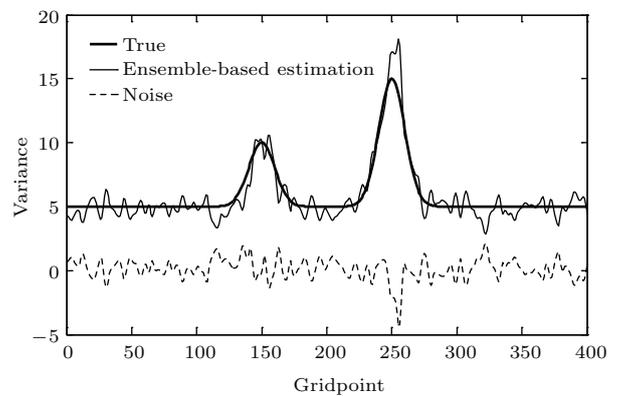


图 4 背景误差方差真值(粗线), 集合估计值(细线)以及噪声(点线)随格点的变化

Fig. 4. True (bold solid) and EDA-based estimated (thin solid) variances and corresponding sampling noise (dashed).

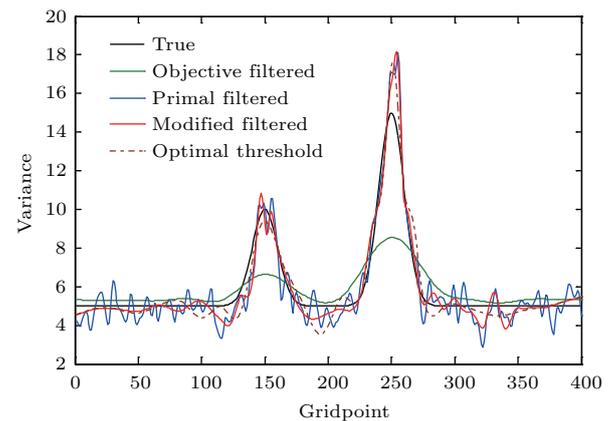


图 5 谱滤波(绿线)、小波滤波(蓝线)、改进后的小波滤波(红线)以及最优阈值小波滤波(点短线)结果的比较, 黑粗线为真值

Fig. 5. Comparison of filtered variances through spectral filtered (green), primal wavelet filtered (blue), modified wavelet filtered (red) and optimal filtered (dotted line), the true variance is denoted by bold solid line.

于 $\sigma_w = 0.41$, 部分噪声的小波系数并没有被置 0. 采用等式 (6) 对阈值进行修正得到 $T_S = 2.65$, 图 5 中红线即为改进后的去噪结果. 可以看出, 去噪后信号的平滑性有了较大的提高, RMSE 也由原来的 0.77 减至 0.65, 接近最优滤波的 0.57. 最优滤波为图 5 中点划线, 是通过手动调整阈值使 RMSE 达到最小得到的. 最优滤波代表了这种去噪方法的最佳效果. 图 6 给出了不同阈值对应的 RMSE 值.

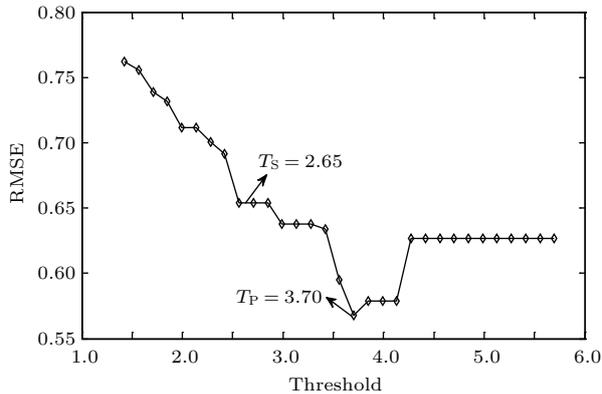


图 6 RMSE 随截断波数的变化 $T_P = 3.7$ 为最优阈值, 滤波后的 RMSE 最小为 0.57 值; $T_S = 2.65$ 为修正后的阈值, 对应 RMSE 为 0.65.

Fig. 6. RMSEs corresponding to different thresholds, $T_P = 3.7$ is the optimal threshold leading to the minimal RMSE value 0.57; $T_S = 2.65$ is obtained by equation (6), its filtered RMSE=0.65.

谱滤波方法也可以通过调整截断波数达到最优化, 结果如图 7 所示. 可以看出, 调整截断波数能够减少谱滤波的误差, 但是 RMSE 值均在 1.25 以上, 总大于小波去噪结果. 这也进一步说明在背景误差方差滤波中, 小波去噪方法要整体优于谱方法. 需要指出的是, 背景误差方差的集合估计值在进入同化系统之前还需要进行降分辨率处理, 如 ECMWF 的业务化集合资料同化系统, 需要将 T399 分辨率的方差估计值进行谱截断, 得到 T65 分辨率的方差. 这种谱截断处理, 会消除小波滤波结果中存在的高频小尺度振荡信息, 而保留大尺度特性, 使滤波结果更加平滑.

图 8 给出了小波空间中方差集合估计值 $\tilde{\mathbf{X}}$, 噪声 $\tilde{\mathbf{X}}^e$ 和信号 $\tilde{\mathbf{X}}^*$ 在区间 $[-5, 5]$ 的概率分布情况. 可以看出, 噪声系数 $\tilde{\mathbf{X}}_{i,j}^e$ 主要集中在以 0 为中心的临域, 且近似为高斯分布. 小波阈值去噪方法采用迭代方法计算出截断阈值 T_A , 然后将落入到区间 $[-T_A, T_A]$ 的小波系数置 0. 对于非相关的高斯白噪声, 大部分小波系数会落入到该区间, T_A

的取值是恰当的 [16]. 但是, 对于非均匀、具有相关性的非高斯噪声, 仍有 13% 的噪声小波系数分布在区间 $[-T_A, T_A]$ 外, 导致部分噪声仍未被滤除, 因此需要修正和放大 T_A . 图 8 中绿线为截断余项 $\tilde{R}_{i,j} \in \|X_{i,j}\| \leq T_A$ 的分布情况, 可以看出即使对于非均匀、具有相关性的非高斯噪声, 其小于 $[-T_A, T_A]$ 的噪声小波系数仍可近似为高斯分布. 由高斯分布的性质可知, $[-2\sigma, 2\sigma]$ 区间的概率为 95%, 由此计算得到的 $T_S = 2.65$ 包含了 90% 的噪声小波系数. 因此其滤波效率高于原有方法.

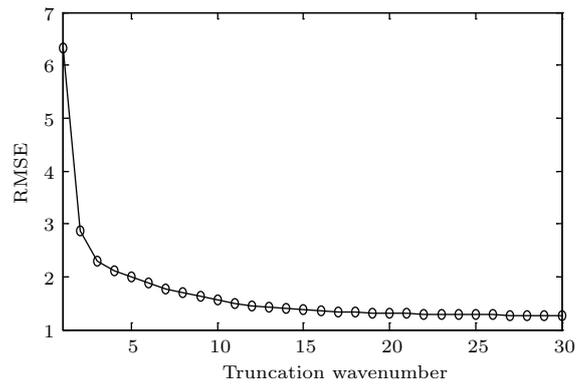


图 7 谱滤波 RMSE 随截断波数的变化

Fig. 7. The variation of RMSE with truncation wave number.

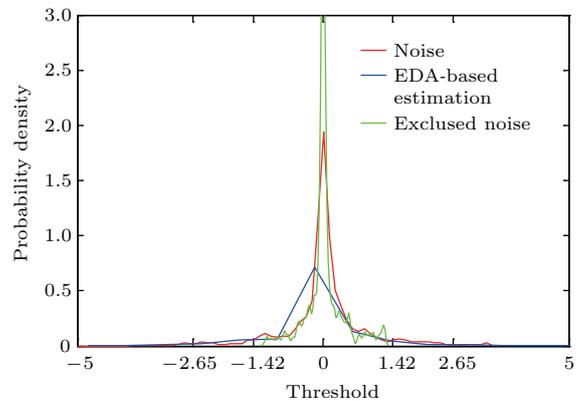


图 8 小波系数的概率密度分布 红、蓝、绿线分别对应噪声, 方差的集合估计值和绝对值小于 T_A 的噪声小波系数
Fig. 8. Probability density distribution of the wavelet coefficients, the red, blue and green line corresponding to noise, EDA-based estimate variance and filtered noise respectively whose coefficients magnitude are small than threshold T_A .

为进一步测试该方法的鲁棒性, 图 9 对比了不同信噪比情况下, 阈值修正前后的滤波效果. 方差集合估计值的精度 (黑线) 正比于集合成员个数的开方. 当集合成员数增加, 背景误差方差估计值中

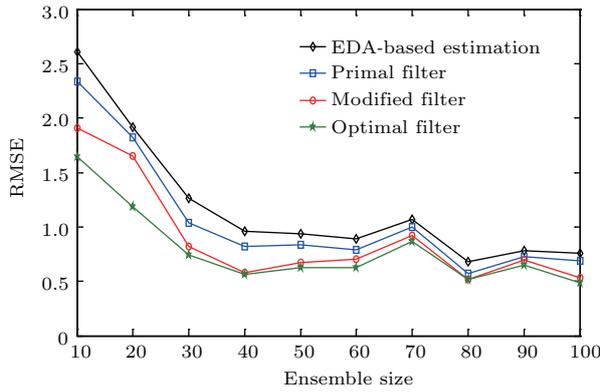


图9 原集合方差估计值(黑色菱形线)和改进前后滤波RMSE随集成员的变化,蓝色方框线为改进前的小波滤波结果,红色圆圈线为改进后的小波滤波结果,绿色星形线为最优滤波结果

Fig. 9. EDA-based variance estimates (black with diamond), filtered results processed by primal wavelet filter (blue with square) and modified wavelet filter (red with cycle) and the optimal filter (green with star).

的信噪比也将增加. 但是这种收敛性非常缓慢, 通过增加样本个数来提高估计值精度的代价是十分昂贵的, 这也凸显了去噪方法的重要性. 可以看出, 在不同信噪比的条件下, 阈值修正后都能显著减少RMSE. 当集成员数 N 取40和100时, 改进幅度已经接近或达到极限值(绿色星形线). 由

这10组试验统计得出, 改进滤波方法后RMSE由原来的1.11减少为0.92, 相对于集合估计值 X 的RMSE(1.40), 滤波效率提升了13.28%.

3.3 在实际系统中的初步应用

YH4DVAR是我国自主研发的全球四维变分资料同化系统, 能够为全球中期和区域短期数值天气预报模式提供高质量的初值场^[1,2]. 文献^[15]已经基于YH4DVAR初步构建了集合资料同化系统, 提高了预报效果^[25]. 以下采用与文献^[15]中相同的试验平台和设置, 选择2013年8月2日0900 UTC(国际时)第91个模式层上的涡度的10个样本估计值作为去噪对象. 此时2013年第九号台风“飞燕”位于海南省文昌市的东南部, 台风涡旋导致涡度背景误差方差出现局部最大值. 由于实际系统缺乏真值, 图10仅给出了原集合估计值、谱滤波和改进前后小波滤波的结果. 可以看出谱滤波和小波阈值方法都能有效滤除集合估计值中的小尺度噪声, 但是与30个样本估计值(这里未给出结果)相比, 谱滤波严重弱化了台风中心涡度方差的极值(由原来的 8.31×10^{-5} 变为 6.20×10^{-5}), 且中心位置出现小幅度的漂移, 效果低于小波阈值方法, 与

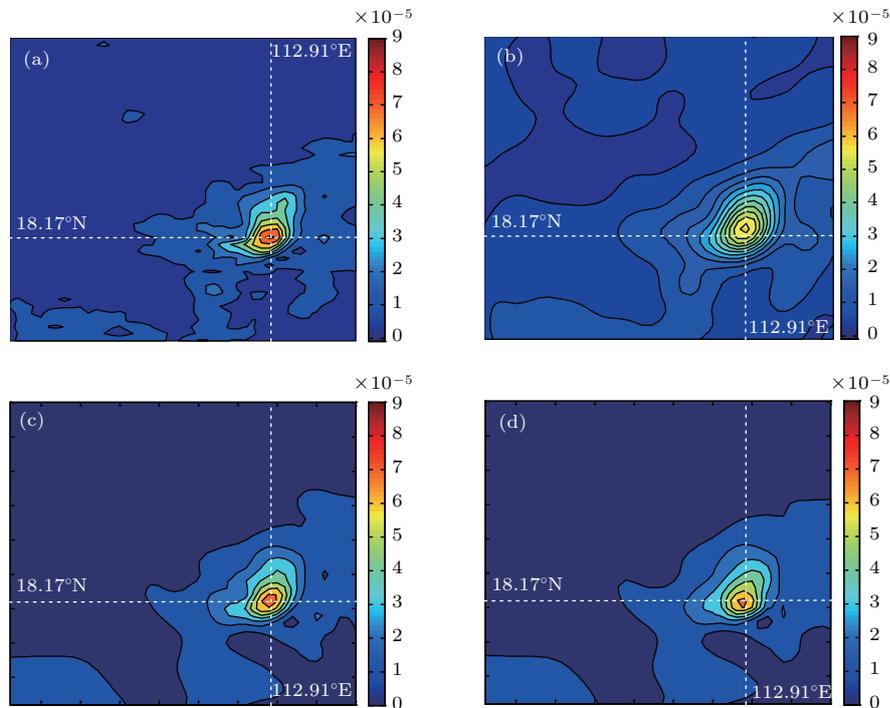


图10 第91个模式层上涡度标准偏差, 时间对应2013年8月2日0900 UTC (a), (b), (c), (d)分别对应10个集合样本的估计值、谱滤波结果、阈值未修正时小波阈值去噪结果和阈值修正后的去噪结果

Fig. 10. Standard deviations of vorticity at ML = 91, corresponding to 2 August 2013 at 0900 UTC (a). Unit: 10^{-5} s^{-1} . Filtered result based on 10 members raw with: (b) spectral filter, (c) raw wavelet filter and (d) modified wavelet filter.

一维理想试验的结论相同. 小波阈值 1.87×10^{-5} 经修正后变为 2.62×10^{-5} , 去噪后台风中心极值分别为 7.65×10^{-5} 和 7.30×10^{-5} , 两者整体结构相似. 阈值修正后滤除了部分小尺度信息, 形状更加平滑, 去噪强度有所增强. 由于缺乏真值和可信的参考值, 该方法的有效性以及对系统的影响还有待进一步分析, 如修正阈值后对预报和分析场的贡献.

4 结 论

集合资料同化中方差滤波能减少因采样导致的随机噪声, 提高背景误差方差集合估计值的精度. 本文针对谱滤波方法不能表征信号在格点空间局地变化的局限, 引入了同时具有谱和空间局地化能力的小波阈值方法, 显著改善了方差的滤波效果. 尤其当方差信号局地变化较大时, 小波滤波方法的优势更加明显. 当随机噪声为均匀的高斯白噪声时, 通过迭代方法计算得到的小波阈值能够消除大部分噪声的小波系数. 但是, 集合资料同化中随机噪声具有尺度相关、非均匀、非高斯等特性, 个别尺度上的噪声能级过高, 导致迭代得到的小波阈值偏小. 对此, 本文根据截断余项的分布特征, 对阈值的计算方法进行了修正和改进. 一维理想试验结果表明, 改进后的小波阈值能进一步减少噪声比重, 提升背景误差方差估计值的精度. 10组试验的统计结果表明, 采用改进的小波阈值方法, 能使背景误差方差的RMSE值减少13.28%, 同时该方法具有很好的鲁棒性.

将该方法应用在了实际的业务系统中, 以台风中心的涡度背景误差方差的集合估计值作为试验对象, 对比了三种去噪方法的效果. 初步结果表明, 小波阈值方法的去噪效果优于谱方法. 虽然阈值修正前后的去噪结果相似, 但修正后能减少更多的小尺度信息. 后续将进一步评估研究该方法对系统整体性能的影响, 以及研究基于球面小波框架的去噪方法.

参考文献

- [1] Zhang W M, Cao X Q, Song J Q 2012 *Acta Phys. Sin.* **61** 249202 (in Chinese) [张卫民, 曹小群, 宋君强 2012 物理学报 **61** 249202]
- [2] Wang S C, LI Y, Zhang W M, Zhao J, Cao X Q 2011 *Acta Phys. Sin.* **60** 099203 (in Chinese) [王舒畅, 李毅, 张卫民, 赵军, 曹小群 2011 物理学报 **60** 099203]
- [3] Laroche S, Gauthier G 1998 *Tellus Ser. A* **50** 557
- [4] Derber J, Bouttier F 1999 *Tellus Ser. A* **51** 195
- [5] Buizza R, Houtekamer P L, Pellerin G, Toth Z, Zhu Y J, Wei M Z 2005 *Mon. Weather Rev.* **133** 1076
- [6] Houtekamer P L, Mitchell H L 1998 *Mon. Weather Rev.* **126** 3
- [7] Evensen G 1994 *J. Geophys. Res.* **99** 10143
- [8] Bonavita M, Raynaud L, Isaksen L 2011 *Q. J. R. Meteorol. Soc.* **137** 423
- [9] Berre L, Varella H, Desroziers G 2015 *Q. J. R. Meteorol. Soc.* **141** 2803
- [10] Raynaud L, Berre L, Desroziers G 2009 *Q. J. R. Meteorol. Soc.* **135** 1177
- [11] Pereira M B, Berre L 2006 *Mon. Weather Rev.* **134** 2466
- [12] Raynaud L, Berre L, Desroziers G 2008 *Q. J. R. Meteorol. Soc.* **134** 1003
- [13] Wiener N 1949 *Extrapolation, Interpolation, and Smoothing of Stationary Time Series* (Cambridge: Massachusetts Institute of Technology) pp86–90
- [14] Bonavita M, Isaksen L, Hólm E 2012 *Q. J. R. Meteorol. Soc.* **138** 1540
- [15] Liu B N, Zhang W M, Cao X Q, Zhao Y L, Huang Q B 2015 *China J. Geophys.* **58** 1526 (in Chinese) [刘柏年, 张卫民, 曹小群, 赵延来, 皇群博, 罗雨 2015 地球物理学报 **58** 1526]
- [16] Donoho D L, Johnstone J M 1994 *Biometrical* **81** 425
- [17] Parrish D F, Derber J C 1992 *Mon. Weather Rev.* **120** 1747
- [18] Fisher M 2003 *Proceedings ECMWF Seminar on "Recent Developments in Data Assimilation for Atmosphere and Ocean"* Reading, September 8–12, 2003 p45
- [19] Isaksen L, Fisher M, Berner J 2006 *ECMWF Tech. Memo.* **492**
- [20] Moore S, Wood S, Davies P 1998 *Annals of Statistics* **26** 1
- [21] Daley R 1993 *Atmospheric Data Analysis* 1993 (Cambridge: Cambridge University Press) pp46–50
- [22] Azzalini A, Farge M, Schneider K 2005 *Appl. Comput. Harmon. Anal.* **18** 177
- [23] Yen R N V, Farge M, Schneider K 2012 *Physica D Non-linear Phenomena* **241** 186
- [24] Pannekoucke O, Raynaud L, Farge M 2014 *Q. J. R. Meteorol. Soc.* **140** 316
- [25] Zhang W M, Liu B N, Cao X Q, Zhao Y L, Zhu M B, Zhao W J 2016 *Acta Meteorol Sin.* **74** 410 (in Chinese) [张卫民, 刘柏年, 曹小群, 赵延来, 朱孟斌, 赵文静 2016 气象学报 **74** 410]

Investigation and experiments of wavelet thresholding in ensemble-based background error variance*

Liu Bai-Nian^{1)2)†} Huang Qun-Bo¹⁾²⁾ Zhang Wei-Min¹⁾ Ren Kai-Jun¹⁾
Cao Xiao-Qun¹⁾ Zhao Jun¹⁾

1) (Academy of Ocean Science and Engineering, National University of Defense Technology, Changsha 410073, China)

2) (College of Computer, National University of Defense Technology, Changsha 410073, China)

(Received 26 March 2016; revised manuscript received 20 October 2016)

Abstract

A large amount of sampling noise which exists in the ensemble-based background error variance need be reduced effectively before being applied to operational data assimilation system. Unlike the typical Gaussian white noise, the sampling noise is scaled and space-dependent, thus making its energy level on some scales much larger than the average. Although previous denoising methods such as spectral filtering or wavelet thresholding have been successfully used for denoising Gaussian white noise, they are no longer applicable for dealing with this kind of sampling noise. One can use a different threshold for each scale, but it will bring a big error especially on larger scales. Another modified method is to use a global multiplicative factor, α , to adjust the filtering strength based on the optimization of trade-off between removal of the noise and averaging of the useful signal. However, tuning α is not so easy, especially in real operational numerical weather prediction context. It motivates us to develop a new nearly cost-free filter whose threshold can be automatically calculated.

According to the characteristics of sampling noise in background error variance, a heterogeneous filtering method similar to wavelet threshold technology is employed. The threshold, T_A , determined by iterative algorithm is used to estimate the truncated remainder whose norm is smaller than T_A . The standard deviation of truncated remainder term is regard as first guess of sampling noise. Non-Gaussian term of sampling noise, whose coefficient modulus is above T_A , is regarded as a small probability event. In order to incorporate such a coefficient into the domain of $[-T, T]$, a semi-empirical formula is used to calculate and approach the ideal threshold.

Investigations are first conducted in a one-dimensional (1D) framework: several methods such as spectral filter, primal wavelet filter, optimal wavelet filter, and proposed filter are used to denoise the scale and space-dependent sampling noise in variance estimations. Their validity and performance are compared and examined with different ensemble sizes. Results show that the proposed method can efficiently filter out a large amount of sampling noise efficiently and improve the estimation accuracy of background error variance. Compared with previous filters, the modified threshold can help to reduce some residual noise arising from the scales where the noise level is above the average level, and the filtering performance of the new method is improved by almost 13.28%. Application to the real ensemble four-dimensional variational data assimilation system is then considered. Results show that the wavelet threshold method outperforms the spectral method. Modified threshold can enhance denosing without deforming the signal of interest.

A new nearly cost-free filter is proposed to reduce the scale and space-dependent sampling noise in ensemble-based background error variance. It is able to remove most of the sampling noises, while extracting the signal of interest. Compared with those of primal wavelet filter and spectral filter, the performance and efficiency of proposed method are improved in 1D framework and real data assimilation system experiments. Further work should focus on the sphere wavelets, which is appropriate for analysing and reconstructing the signals on the sphere in global spectral models.

Keywords: wavelet, ensemble data assimilation, background error variance, filter

PACS: 05.40.-a, 05.40.Ca, 02.60.-x, 43.60.Hj

DOI: 10.7498/aps.66.020505

* Project supported by the National Natural Science Foundation of China (Grant Nos. 41375113, 41475094, 41305101, 41605070).

† Corresponding author. E-mail: bnliu@nudt.edu.cn