

南極海の海水面積の分析

今井 凌

要約. 南極海 (ウェッデル海, インド洋, 太平洋, ロス海, ベリングスハウゼンアムンゼン海) の日時の海水面積値のデータを分析しました. この5つの海間で海水形成に何らかの関係性があるかを見るため, 直近の海水拡大期における海水面積値の日時の差分を取った5次元の時系列データにVARモデルとファクターモデルをフィッティングしました. ここでの解析は仮説検証的ではない探索的なものですが, フィットされたモデルからは, この期間でベリングスハウゼンアムンゼン海の海水面積値の日時変化が他の4つの海のものと同日的にも1日遅れ的にも逆方向であったことが示唆されました. 今回の解析について今後取り組むべき問題点も議論します.

データ. JAXA の JASMES for sea ice で提供されている南半球の日時海水面積値データを使用しました. 海水面積値は衛星画像の海水密度度が一定以上のピクセルの総面積と定義されるようなので (参考: <https://www2.kaiyodai.ac.jp/~koji/SeaIce.html>), これも JAXA の衛星画像データという風に解釈しました. これは南極海の日時海水面積値のデータ (単位は $[\text{km}^2]$, 5日移動平均, 期間は1978年11月1日から2024年9月15日まで (データ取得時)) であって, 南極海はウェッデル海, インド洋, 太平洋, ロス海, ベリングスハウゼンアムンゼン海, の5つの部分に分けられています. これら5つの部分の地理的な位置関係は Figure 1 に表されています.

データの前処理. このデータは明らかに季節変動を表していて, つまり夏には縮小し冬に拡大しています. Figure 2 は例としてウェッデル海の2006年9月15日から2024年9月15日までの海水面積値をプロットしたものです. そこで最も直近の拡大期, 大まかに2024年2月13日から2024年9月5日までがそうであると決めて, この期間のみ分析することになりました. この期間では5つの時系列データのすべてが線形なトレンドを示していましたので (Figure 3), 原系列の1日差分を取りました. 1日差分 $\nabla Y_{it} = Y_{it} - Y_{it-1}$, $i = \text{Weddell, Indian, Pacific, Ross, Bellingshausen Amundsen}$, $t = 2024\text{年}2\text{月}13\text{日}, \dots, 2024\text{年}9\text{月}5\text{日}$ (206のデータ点) のプロットが Figure 4 に示されています.

解析. Figure 4 からは, 時系列 ∇Y_{it} は, ロス海とベリングスハウゼンアムンゼン海については定常に見えますが, ウェッデル海, インド洋, 太平洋については定常かもしれないし定常ではないかもしれないかもしれません. 後ろの3つの海に関しては分散が期間の終わりにかけて徐々に大きくなっているように見えるかもしれないからです. Figure 5 はこれら5つの時系列の標本クロスコリオグラムを示したもので, 各々の周辺自己相関関数 (対角線上にあるプロット) は初めの1日ラグから5日ラグにかけて急激に減少しています. ここから各時系列は定常であると仮定しても大丈夫であろうと判断し, VARモデルとファクターモデルを当てはめることにしました. また, Zhang, Robinson and Yao [4] の手法 (R パッケージ 'HDTSA' の中の `Coint` 関数) をこのデータ行列に適用すると ($\text{lag.k} = 5$), 共和分ランクは5であると推定されました. このことはこれら元の5つの時系列は何か潜在的な弱定常5次元時系列ベクトルの線形変換としてモデル化してもよいことを示唆しており, これはさらに元の5次元の時系列が弱定常であることを示唆しています.

VARモデル. VAR(p) モデルをこの 206×5 のデータ行列に当てはめました. R パッケージ 'vars' の中の VAR 関数を用い, 自己回帰の階数 p は最大ラグは5として BIC によって選択されました. R の出力は Appendix に表示されています. BIC により選択された階数は $p = 1$ で, フィットされた5つのモデルすべてが実用的に0の P 値で統計的に有意となりました. 有意ではない項を外して再

度モデルをフィットした結果は以下のようになります。

$$\begin{cases} \nabla Y_{\text{Weddell}} = 3910.1517 + 0.8407 \nabla Y_{\text{Weddell}-1} + \varepsilon_{\text{Weddell}}, \\ \nabla Y_{\text{Indiant}} = 4092 + 0.6960 \nabla Y_{\text{Indiant}-1} + \varepsilon_{\text{Indiant}}, \\ \nabla Y_{\text{Pacifict}} = 2839.27675 + 0.69725 \nabla Y_{\text{Pacifict}-1} - 0.07592 \nabla Y_{\text{BellingshausenAmundsent}-1} + \varepsilon_{\text{Pacifict}}, \\ \nabla Y_{\text{Rosst}} = 5487.32317 + 0.74746 \nabla Y_{\text{Rosst}-1} - 0.10591 \nabla Y_{\text{BellingshausenAmundsent}-1} + \varepsilon_{\text{Rosst}}, \\ \nabla Y_{\text{BellingshausenAmundsent}} = 0.17238 \nabla Y_{\text{Pacifict}-1} + 0.78714 \nabla Y_{\text{BellingshausenAmundsent}-1} + \varepsilon_{\text{BellingshausenAmundsent}}. \end{cases}$$

モデルの説明力は高く、最も低い調整済み R^2 値は太平洋の 60.11% で最も高いのはウェッデル海の 84.6% です。この推定結果で興味深いのは、ベリングスハウゼンアムンゼン海だけが他の海、すなわち太平洋 (P 値: 0.026531) とロス海 (P 値: 0.0244) に対し負の (しかしそこまで大きくはない) 有意な影響を持っていることです。これはつまり、ある 1 日におけるベリングスハウゼンアムンゼン海の水面積値の減少 (増加) が、次の日の太平洋とロス海における水面積値の増加 (減少) に貢献しているかもしれないことを示唆しています。逆に、この推定結果では太平洋はベリングスハウゼンアムンゼン海に対して正の 1 日遅れの有意な影響を与えていることとなります。また、ベリングスハウゼンアムンゼン海に対して推定された式のみが定数項を持たないことも記しておくべきかもしれません。

ファクターモデル. Lam and Yao [1] の手法 (R パッケージ ‘HDTSA’ 内の `Factors` 関数) を用いてこのデータ行列にファクターモデルを当てはめました ($\text{lag.k} = 5$)。ファクター数は 1 と推定され、 5×1 の因子負荷行列は次のように推定されました。

$$\hat{\mathbf{A}} = [0.8487129 \quad 0.1312050 \quad 0.1402631 \quad 0.4382938 \quad -0.2251585]^\top.$$

この唯一の因子を解釈することは難しいですが (例えば南極周辺の気温や海面温度に関係しているかもしれません)、この推定結果でも興味深いのは、ベリングスハウゼンアムンゼン海に対する係数が他の 4 つの海に比べて符号が逆になっていることです。このモデルの解釈としては、潜在的なレベルで、ベリングスハウゼンアムンゼン海における水面積値は他の海で水面積値が減る (増える) ときに増え (減り) がちであるということになります。VAR モデルではそうした逆方向の影響は 1 日遅れのものとして現れていますが、こちらのファクターモデルでは、潜在的なレベルにおいて同時的な負の相関があるということになります。

ディスカッション. 2024 年の直近の水氷拡大期における、南極海の 5 つの部分の日時水氷面積値データに対し、VAR モデルとファクターモデルを当てはめました。推定結果からは、ベリングスハウゼンアムンゼン海の水氷面積値が、この期間、同日ならびに 1 日遅れのレベルで、他の 4 つの海とは逆方向の日時変動を経験していたことが示唆されました。このような現象が他の年でも見られるかや、対応する水氷縮小期ではどうであるかなどを確認しなければなりません。特に、今回解析した期間の中で、ベリングスハウゼンアムンゼン海の水氷面積値は 2024 年 6 月 11 日から 2024 年 6 月 12 日にかけて急上昇しており (上昇値は 15000 [km²] に迫ります)、その後数日には太平洋とロス海の水氷面積値は減少しています。この変化が解析に影響を与えた可能性はあり、この変動を外れ値として扱うかどうかを注意深く判断する必要があると思われます。

南極の周りには 3 つの主な還流があり (南極周極流, ウェッデル海亜寒帯循環, ロス海亜寒帯循環)、これら還流の水氷面積値への影響を何らかの形でモデルに反映するべきかもしれません。またデータ中で分けられている南極海の 5 つの部分の各面積は互いに等しくないのも、もしかするとその点に関して何か調整をした方がいいかもしれません。

最後に、ここでは季節性の問題から逃げるため直近の水氷拡大期だけに注目しましたが、元のデータセットは 1978 年 11 月 1 日から始まるはるかに潤沢なものです。より重要な問いは、人為的な要因によることが確実視されている近年の急激な地球温暖化が、南極の水氷面積にどのように影響を与えるかということでしょう。こちらの「地球が見える」最新記事 (<https://earth.jaxa.jp/ja/earthview/2024/09/24/8268/index.html>) では北極海の水氷域面積最小値の経年変化が注目されています。このように水氷面積の経年での変化を捉えつつ、かつこの豊かな時系列データが持つより細かい日時的な変動も一挙に扱うためには、時系列の様々な周波数の周期的な変動を自然に扱える手法を適用する必要があるでしょう。そのような手法として Whitcher, Guttorp and Percival

[3] 等に代表されるウェーブレット解析や Matsuda and Komaki [2] の振動子分解モデル等が候補になると思われます。

参考文献

- [1] C. Lam and Q. Yao. Factor modeling for high-dimensional time series: inference for the number of factors. *Ann. Statist.*, 40(2):694–726, 2012.
- [2] T. Matsuda and F. Komaki. Multivariate time series decomposition into oscillation components. *Neural Comput.*, 29(8):2055–2075, 2017.
- [3] B. Whitcher, P. Guttorp, and D. B. Percival. Wavelet analysis of covariance with application to atmospheric time series. *Journal of Geophysical Research: Atmospheres*, 105(D11):14941–14962, 2000.
- [4] R. Zhang, P. Robinson, and Q. Yao. Identifying cointegration by eigenanalysis. *J. Amer. Statist. Assoc.*, 114(526):916–927, 2019.

Appendix.

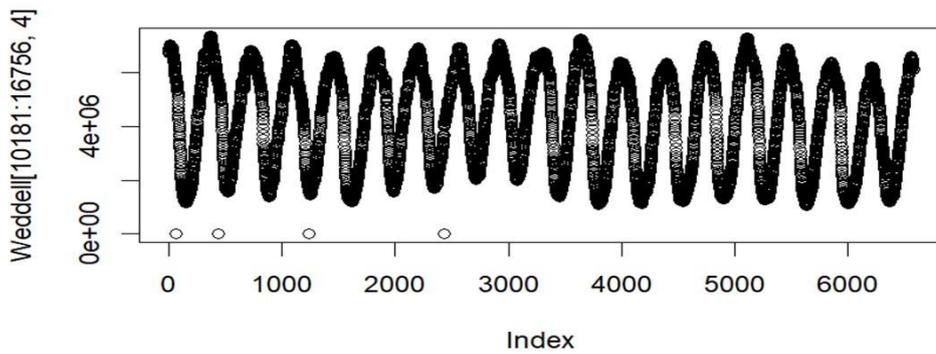
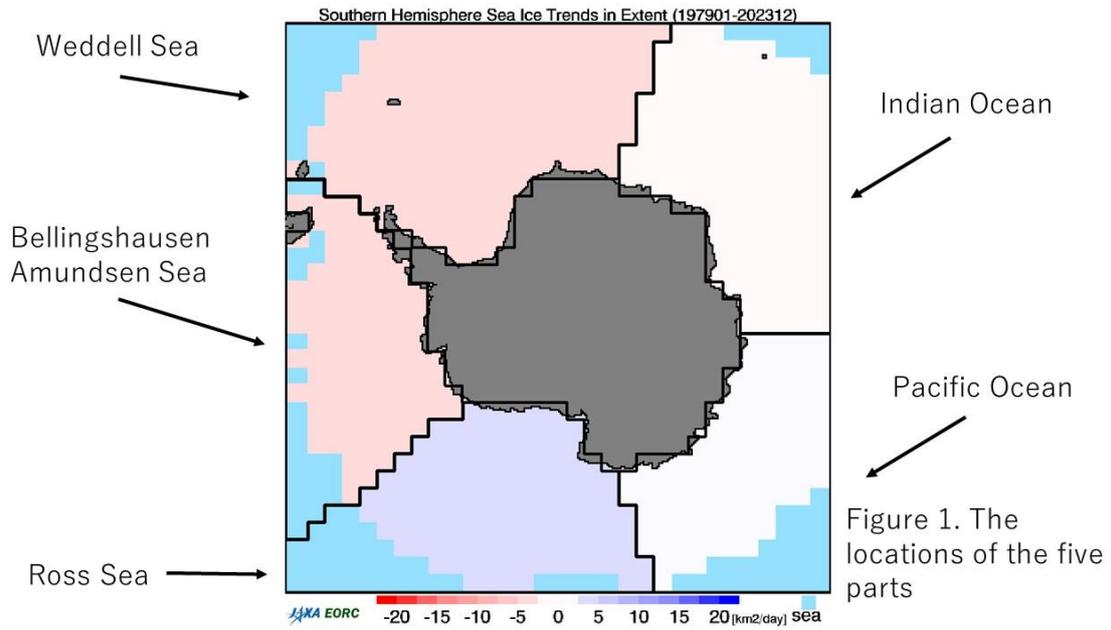


Figure 2. The sea ice extent [km²] for the Weddell Sea from September 15, 2006 to September 15, 2024

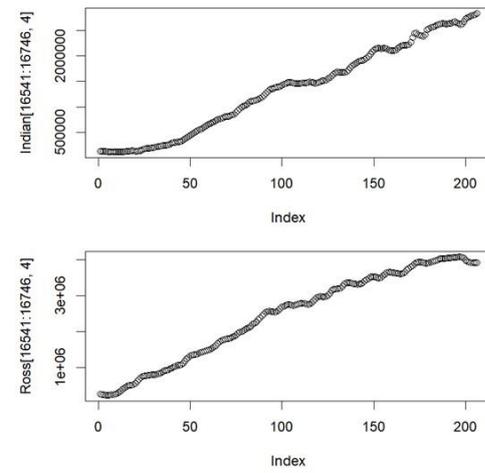
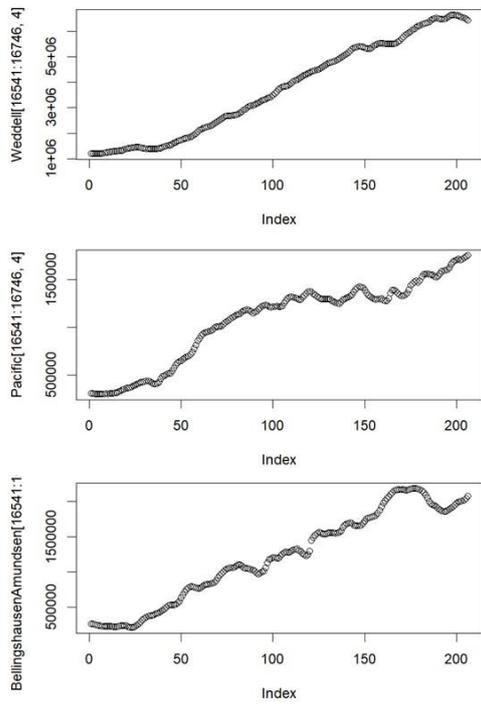


Figure 3. The sea ice extent [km²] for five seas from February 13, 2024 until September 5, 2024

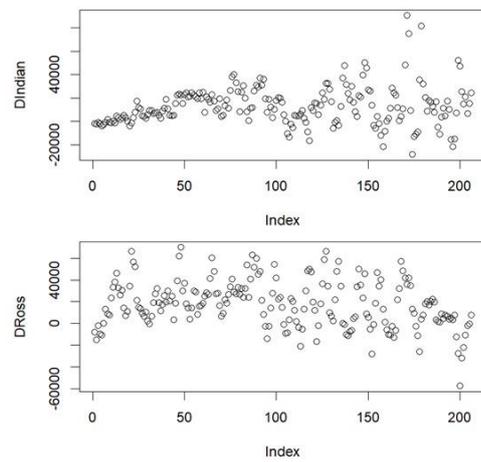
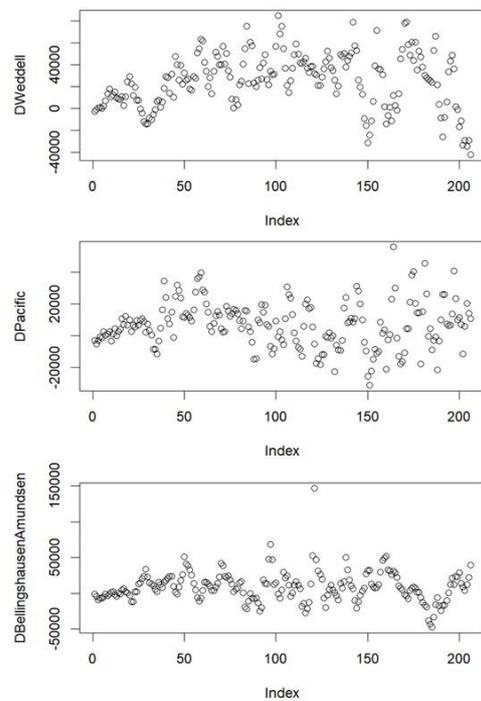


Figure 4. Differences ∇Y_{it} , $i = \text{Weddell, Indian, Pacific, Ross, Bellingshausen Amundsen}$ $t = \text{Feb 13, 2024, ..., Sep 5, 2024}$

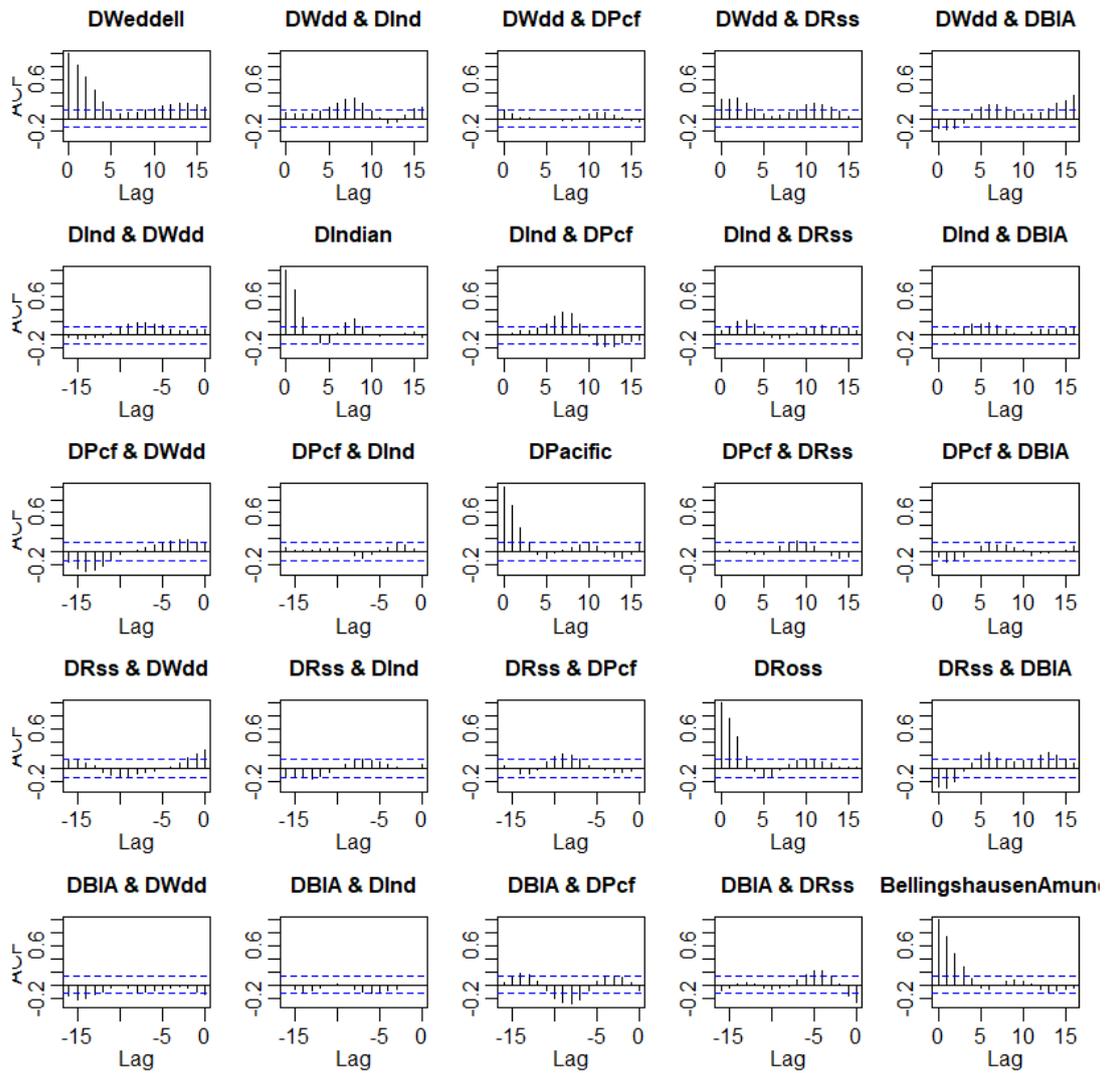


Figure 5. The sample cross-correlogram of the differences ∇Y_{it} of the sea ice extent

R Outputs.

```
> var = VAR(data, lag.max=5, ic="SC"); summary(var)
```

VAR Estimation Results:

=====

Endogenous variables: DWeddell, DIndian, DPacific, DRoss, DBellingshaus
enAmundsen

Deterministic variables: const

Sample size: 205

Log Likelihood: -11102.162

Roots of the characteristic polynomial:

0.8201 0.7424 0.7424 0.7299 0.7299

Call:

VAR(y = data, lag.max = 5, ic = "SC")

Estimation results for equation DWeddell:

=====

DWeddell = DWeddell.l1 + DIndian.l1 + DPacific.l1 + DRoss.l1 + DBellings
hausenAmundsen.l1 + const

	Estimate	Std. Error	t value	Pr(> t)
DWeddell.l1	0.82928	0.04239	19.562	<2e-16 ***
DIndian.l1	-0.02103	0.05943	-0.354	0.7239
DPacific.l1	-0.09239	0.06854	-1.348	0.1792
DRoss.l1	0.05843	0.04881	1.197	0.2327
DBellingshausenAmundsen.l1	-0.04954	0.04870	-1.017	0.3103
const	4510.91756	1829.98918	2.465	0.0145 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13930 on 199 degrees of freedom

Multiple R-squared: 0.6932, Adjusted R-squared: 0.6854

F-statistic: 89.91 on 5 and 199 DF, p-value: < 2.2e-16

Estimation results for equation DIndian:

=====

$$DIndian = DWedde11.11 + DIndian.11 + DPacific.11 + DRoss.11 + DBellingshausenAmundsen.11 + const$$

	Estimate	Std. Error	t value	Pr(> t)
DWedde11.11	-5.157e-03	3.624e-02	-0.142	0.8870
DIndian.11	6.919e-01	5.081e-02	13.617	<2e-16 ***
DPacific.11	4.177e-02	5.859e-02	0.713	0.4767
DRoss.11	6.043e-02	4.173e-02	1.448	0.1492
DBellingshausenAmundsen.11	2.599e-02	4.163e-02	0.624	0.5331
const	2.687e+03	1.564e+03	1.718	0.0874 .

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11910 on 199 degrees of freedom

Multiple R-squared: 0.492, Adjusted R-squared: 0.4792

F-statistic: 38.54 on 5 and 199 DF, p-value: < 2.2e-16

Estimation results for equation DPacific:

=====

$$DPacific = DWedde11.11 + DIndian.11 + DPacific.11 + DRoss.11 + DBellingshausenAmundsen.11 + const$$

	Estimate	Std. Error	t value	Pr(> t)
DWedde11.11	0.01488	0.03095	0.481	0.6312
DIndian.11	0.04210	0.04339	0.970	0.3331
DPacific.11	0.69316	0.05004	13.852	<2e-16 ***
DRoss.11	-0.01486	0.03564	-0.417	0.6772
DBellingshausenAmundsen.11	-0.07792	0.03555	-2.192	0.0296 *
const	2217.48393	1336.01373	1.660	0.0985 .

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10170 on 199 degrees of freedom
Multiple R-Squared: 0.5157, Adjusted R-squared: 0.5036
F-statistic: 42.39 on 5 and 199 DF, p-value: < 2.2e-16

Estimation results for equation DRoss:

=====

DRoss = DWedde11.11 + DIndian.11 + DPacific.11 + DRoss.11 + DBellingshausenAmundsen.11 + const

	Estimate	Std. Error	t value	Pr(> t)
DWedde11.11	4.213e-03	4.098e-02	0.103	0.918219
DIndian.11	-6.026e-02	5.745e-02	-1.049	0.295549
DPacific.11	-6.129e-02	6.626e-02	-0.925	0.356099
DRoss.11	7.481e-01	4.719e-02	15.853	< 2e-16 ***
DBellingshausenAmundsen.11	-1.088e-01	4.707e-02	-2.311	0.021848 *
const	6.606e+03	1.769e+03	3.734	0.000246 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13460 on 199 degrees of freedom
Multiple R-Squared: 0.6189, Adjusted R-squared: 0.6094
F-statistic: 64.64 on 5 and 199 DF, p-value: < 2.2e-16

Estimation results for equation DBellingshausenAmundsen:

=====

DBellingshausenAmundsen = DWedde11.11 + DIndian.11 + DPacific.11 + DRoss.11 + DBellingshausenAmundsen.11 + const

	Estimate	Std. Error	t value	Pr(> t)
DWedde11.11	-3.674e-02	4.170e-02	-0.881	0.3793

DIndian.l1	4.588e-04	5.846e-02	0.008	0.9937
DPacific.l1	1.530e-01	6.742e-02	2.269	0.0243 *
DRoss.l1	6.356e-02	4.802e-02	1.324	0.1871
DBellingshausenAmundsen.l1	7.816e-01	4.790e-02	16.319	<2e-16 ***
const	8.251e+02	1.800e+03	0.458	0.6472

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13700 on 199 degrees of freedom
Multiple R-Squared: 0.5866, Adjusted R-squared: 0.5762
F-statistic: 56.47 on 5 and 199 DF, p-value: < 2.2e-16

Covariance matrix of residuals:

	DWeddell	DIndian	DPacific	DRoss
DWeddell	194017288	9518661	19093856	25684930
DIndian	9518661	141787847	-13625707	2385538
DPacific	19093856	-13625707	103410640	-2487336
DRoss	25684930	2385538	-2487336	181301633
DBellingshausenAmundsen	-11418520	-4405383	-6986462	-36206885
	DBellingshausenAmundsen			
DWeddell		-11418520		
DIndian		-4405383		
DPacific		-6986462		
DRoss		-36206885		
DBellingshausenAmundsen			187713272	

Correlation matrix of residuals:

	DWeddell	DIndian	DPacific	DRoss
DWeddell	1.00000	0.05739	0.13480	0.13695
DIndian	0.05739	1.00000	-0.11253	0.01488
DPacific	0.13480	-0.11253	1.00000	-0.01817
DRoss	0.13695	0.01488	-0.01817	1.00000
DBellingshausenAmundsen	-0.05983	-0.02700	-0.05014	-0.19626

	DBellingshausenAmundsen
DWeddell	-0.05983
DIndian	-0.02700
DPacific	-0.05014
DRoss	-0.19626
DBellingshausenAmundsen	1.00000

```
> varR = restrict(var); summary(varR)
```

VAR Estimation Results:

=====

Endogenous variables: DWeddell, DIndian, DPacific, DRoss, DBellingshausenAmundsen

Deterministic variables: const

Sample size: 205

Log Likelihood: -11109.56

Roots of the characteristic polynomial:

0.8407 0.7496 0.7496 0.7475 0.696

Call:

VAR(y = data, lag.max = 5, ic = "SC")

Estimation results for equation DWeddell:

=====

DWeddell = DWeddell.l1 + const

	Estimate	Std. Error	t value	Pr(> t)
DWeddell.l1	0.8407	0.0400	21.018	< 2e-16 ***
const	3910.1517	1418.9529	2.756	0.00639 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13970 on 203 degrees of freedom

Multiple R-Squared: 0.8475, Adjusted R-squared: 0.846
F-statistic: 564.1 on 2 and 203 DF, p-value: < 2.2e-16

Estimation results for equation DIndian:

=====

DIndian = DIndian.l1 + const

	Estimate	Std. Error	t value	Pr(> t)	
DIndian.l1	6.960e-01	5.031e-02	13.834	< 2e-16	***
const	4.092e+03	1.057e+03	3.871	0.000146	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11870 on 203 degrees of freedom
Multiple R-Squared: 0.6861, Adjusted R-squared: 0.683
F-statistic: 221.8 on 2 and 203 DF, p-value: < 2.2e-16

Estimation results for equation DPacific:

=====

DPacific = DPacific.l1 + DBellingshausenAmundsen.l1 + const

	Estimate	Std. Error	t value	Pr(> t)	
DPacific.l1	0.69725	0.04926	14.154	< 2e-16	***
DBellingshausenAmundsen.l1	-0.07592	0.03397	-2.235	0.026531	*
const	2839.27675	849.92985	3.341	0.000996	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 10130 on 202 degrees of freedom
Multiple R-Squared: 0.6069, Adjusted R-squared: 0.6011
F-statistic: 104 on 3 and 202 DF, p-value: < 2.2e-16

Estimation results for equation DRoss:

=====

DRoss = DRoss.l1 + DBellingshausenAmundsen.l1 + const

	Estimate	Std. Error	t value	Pr(> t)
DRoss.l1	0.74746	0.04526	16.514	< 2e-16 ***
DBellingshausenAmundsen.l1	-0.10591	0.04669	-2.269	0.0244 *
const	5487.32317	1367.70027	4.012	8.47e-05 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13430 on 202 degrees of freedom
Multiple R-Squared: 0.7726, Adjusted R-squared: 0.7692
F-statistic: 228.7 on 3 and 202 DF, p-value: < 2.2e-16

Estimation results for equation DBellingshausenAmundsen:

=====

DBellingshausenAmundsen = DPacific.l1 + DBellingshausenAmundsen.l1

	Estimate	Std. Error	t value	Pr(> t)
DPacific.l1	0.17238	0.05989	2.878	0.00443 **
DBellingshausenAmundsen.l1	0.78714	0.04242	18.557	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13670 on 203 degrees of freedom
Multiple R-Squared: 0.6433, Adjusted R-squared: 0.6398
F-statistic: 183 on 2 and 203 DF, p-value: < 2.2e-16

Covariance matrix of residuals:

	DWeddell	DIndian	DPacific	DRoss
DWeddell	199075316	9995051	18535185	27256332
DIndian	9995051	143659130	-13934987	1933868
DPacific	18535185	-13934987	104098982	-3181026
DRoss	27256332	1933868	-3181026	183081713
DBellingshausenAmundsen	-8863038	-3245909	-7520858	-35856919

	DBellingshausenAmundsen
DWeddell	-8863038
DIndian	-3245909
DPacific	-7520858
DRoss	-35856919
DBellingshausenAmundsen	189968438

Correlation matrix of residuals:

	DWeddell	DIndian	DPacific	DRoss
DWeddell	1.00000	0.05910	0.12876	0.14277
DIndian	0.05910	1.00000	-0.11395	0.01192
DPacific	0.12876	-0.11395	1.00000	-0.02304
DRoss	0.14277	0.01192	-0.02304	1.00000
DBellingshausenAmundsen	-0.04558	-0.01965	-0.05348	-0.19227

	DBellingshausenAmundsen
DWeddell	-0.04558
DIndian	-0.01965
DPacific	-0.05348
DRoss	-0.19227
DBellingshausenAmundsen	1.00000

```
> fac <- Factors(data,lag.k=5)
```

```
> r_hat <- fac$factor_num; r_hat
```

The estimated number of factors

1

```
> loading_Mat <- fac$loading.mat; loading_Mat
```

```
[1] 0.8487129 0.1312050 0.1402631 0.4382938 -0.2251585
```