

Akurasi Peramalan Long Horizon dengan Singular Spectrum Analysis

Awit Marwati Sakinah^{1, a)}

¹Jurusan Komputerasi Akuntansi, STMIK Tasikmalaya

^{a)}email: awitsakinah@gmail.com

Abstrak

Perubahan iklim akhir-akhir ini tidak dapat dihindari. Salah satu penyebab perubahan iklim adalah perubahan suhu udara. Untuk itu, perlu dilakukan peramalan suhu agar penyimpangannya dapat diantisipasi. Dalam penelitian ini akan dibandingkan akurasi hasil peramalan dengan menggunakan model Singular Spectrum Analysis (SSA) dengan *R-forecasting* dan *V-Forecasting*. Peramalan dengan metode SSA *R-forecasting* dan *V-Forecasting* pada suhu Jakarta menggunakan *window length* $L = 204$ dan $r = 3$ menghasilkan ramalan yang tidak jauh berbeda (aproksimasi kekontinuannya hampir sama). Berdasarkan hasil analisis, didapat MAPE untuk hasil peramalan dengan SSA *R-forecasting* sebesar 5,0029 yang lebih besar dari MAPE SSA *V-Forecasting* sebesar 4,0067. Ini menunjukkan bahwa peramalan suhu untuk long horizon lebih akurat dengan menggunakan *V-Forecasting* dibandingkan dengan *R-Forecasting*.

Kata kunci: akurasi, Singular Spectrum Analysis, *R-Forecasting*, *V-Forecasting*

Abstract

Climate change lately cannot be avoided. One of the causes of climate change is changes in air temperature. It is necessary to forecast the temperature so that the deviation can be anticipated. In this study, the accuracy of forecasting results will be compared using the Singular Spectrum Analysis (SSA) model with *R-forecasting* and *V-Forecasting*. Forecasting with the SSA *R-forecasting* method and *V-Forecasting* at Jakarta temperature using the window length $L = 204$ and $r = 3$ produces predictions that are not much different (approximation is almost the same continuity). Based on the results of the analysis, obtained MAPE for SSA *R-forecasting* results of 5.0029 which is higher than 4.0067 for SSA *V-Forecasting*. This shows that temperature forecasting for long horizon is more accurate by using *V-Forecasting* compared to *R-Forecasting*.

Keywords: accuracy, Singular Spectrum Analysis, *R-Forecasting*, *V-Forecasting*

Pendahuluan

Perubahan iklim akhir-akhir ini tidak dapat dihindari. Salah satu penyebab perubahan iklim adalah perubahan suhu udara. Suhu udara memegang peranan penting dalam berbagai bidang, diantaranya dalam bidang pertanian, yaitu dalam semua proses pertumbuhan. Dalam bidang kesehatan, suhu udara yang fluktuatif akan menyebabkan timbulnya penyakit. Dalam bidang penerbangan, perubahan suhu berpengaruh pada jadwal keperluan *start engine* saat *take off*. Selain itu, suhu udara yang terlalu tinggi akan rentan menyebabkan kebakaran hutan dan peningkatan polusi udara.

Adanya perubahan pada satu faktor seperti pada suhu mampu merubah pola yang dibentuk oleh faktor-faktor secara keseluruhan. Untuk itu, perlu dilakukan peramalan suhu agar penyimpangannya dapat diantisipasi. Menurut horizon waktunya, peramalan dibagi dalam tiga bagian, yaitu peramalan jangka panjang (*long-horizon*) untuk 2-10 tahun, medium untuk 1-24 bulan, dan jangka pendek atau *short-horizon* untuk 1-5 minggu.

Peramalan dapat digunakan jika data masa lalu mampu merefleksikan masa depan [1]. Salah satu teknik peramalan yang dapat digunakan adalah SSA (*Singular Spectrum Analysis*). SSA adalah teknik analisis *time series* baru dan powerful yang menggabungkan unsur *time series* klasik, statistik multivariat, geometri multivariat, sistem dinamis dan pemrosesan sinyal [2]. Broomhead dan Raja [3] memberikan kontribusi pertama pada kemunculan SSA. Mereka menunjukkan bahwa *Singular Value Decomposition* (SVD) efektif mengurangi *noise*.

Banyak penelitian dengan menggunakan SSA, Hassani [2] membandingkan SSA dengan metode SARIMA, ARAR, dan *seasonal Holt-Winter* pada peramalan data kematian akibat kecelakaan di USA dari tahun 1973 sampai dengan 1978. Hasilnya menunjukkan bahwa SSA merupakan metode terbaik dibanding ketiga metode lainnya, hal ini terlihat dari nilai *Mean Absolute Error* (MAE) dan *Mean Relative Absolute Error* (MRAE) yang terkecil.

Myung [4] menunjukkan bahwa SSA adalah teknik analisis *time series* yang baik untuk mengekstrak tren dalam data dan komponen lainnya dengan struktur data sederhana. SSA telah terbukti sangat baik dan telah menjadi metode standar dalam analisis *time series* klimatologi, meteorologi dan geofisika sehingga tidak asing jika metode ini banyak ditemukan di jurnal-jurnal klimatologi, jurnal *Atmospheric Sciences* dan jurnal penelitian geofisika [5].

Terdapat dua metode peramalan dalam SSA, di antaranya metode *recurent* dan vektor. Metode *recurent* adalah metode dasar yang sering digunakan karena relatif lebih mudah [6]. Sedangkan metode vektor merupakan hasil modifikasi dari metode *recurent* yang dikembangkan oleh Nekrutkin tahun 1999. Perbedaan antara kedua metode peramalan tersebut terletak pada kontinuasinya. Metode *recurent* melakukan kontinuitas secara langsung (dengan bantuan LRF), sedangkan peramalan dengan metode vektor berhubungan dengan *L-continuation*. Ini menyebabkan dalam *approximate continuation*-nya biasanya memberikan hasil yang berbeda.

Metode

Singular Spectrum Analysis (SSA) adalah teknik analisis *time series* baru dan powerful yang menggabungkan unsur *time series* klasik, statistik multivariat, geometri multivariat, sistem dinamis dan pemrosesan sinyal [2]. SSA terbuat dari dua istilah yaitu *singular* dan *spectrum*. Istilah *singular* didefinisikan dekomposisi spektral atau dekomposisi eigen dari matriks lintasan. Istilah *spektrum* didefinisikan untuk penjumlahan satu set nilai eigen setelah dekomposisi spektral [7].

Algoritma Dasar SSA (*Singular Spectrum Analysis*)

Terdapat dua tahap algoritma dalam SSA. Tahap algoritma pertama disebut dekomposisi dan yang kedua adalah rekonstruksi. Pada tahap dekomposisi terdiri dari *embedding* dan *Singular Value Decomposition* (SVD). Sedangkan tahap rekonstruksi dilakukan pengelompokan dan *diagonal averaging*.

Pada tahap pertama, *embedding* dengan menguraikan seri data asli unidimensi ke dalam seri multidimensi. Seri multidimensi inilah yang nantinya akan membentuk matriks lintasan X dengan dimensi $L \times K$. *Window length* L dengan $2 < L < N/2$ merupakan parameter yang berperan penting dalam langkah ini. Sedangkan K adalah banyaknya vektor lag.

Tahap SVD mendekomposisi nilai singular dari matriks lintasan dan kemudian menyajikan matriks lintasan sebagai jumlah dari beberapa matriks sesuai dengan nilai singularnya. Dalam tahap ini dihasilkan *eigen triple* yang terdiri dari *singular value* $\sqrt{\lambda_i}$, *eigenvector* U_i dan *principal component* V_i .

$$\begin{aligned}
 \mathbf{X} &= \mathbf{X}_1 + \mathbf{X}_2 + \dots + \mathbf{X}_d \\
 &= U_1 \sqrt{\lambda_1} V_1^T + \dots + U_d \sqrt{\lambda_d} V_d^T \\
 &= \sum_{i=1}^d U_i \sqrt{\lambda_i} V_i^T
 \end{aligned} \tag{1}$$

Langkah berikutnya yaitu *grouping*. Tujuan dari langkah ini adalah untuk memisahkan komponen aditif dari deret waktu dengan pemisahan matriks yang telah dihitung dalam SVD ke dalam beberapa subkelompok matriks dan menjumlahkan matriks dalam masing-masing kelompok. Misal $l = \{i_1, \dots, i_p\}$ adalah matriks resultan \mathbf{X}_l dengan indeks i_1, \dots, i_p sesuai dengan kelompok l dapat didefinisikan sebagai $\mathbf{X}_l = \mathbf{X}_{i_1} + \dots + \mathbf{X}_{i_p}$. Kemudian \mathbf{X}_l disesuaikan dengan kelompok $l = \{l_1, \dots, l_m\}$. Dengan demikian, $\mathbf{X} = \mathbf{X}_1 + \dots + \mathbf{X}_d$ diekspansi menjadi:

$$\mathbf{X} = \mathbf{X}_{l_1} + \dots + \mathbf{X}_{l_m} \tag{2}$$

Tahap *diagonal averaging* ini memetakan matriks lintasan menjadi deret itu sendiri. Dengan kata lain, dekomposisi seri awal menjadi komponen-komponen aditif untuk merekonstruksi *time series* baru dari subkelompok.

Peramalan dengan SSA *R-Forecasting* dan *V-Forecasting*

Peramalan untuk periode mendatang dapat dilakukan jika model sudah terbentuk. Dalam metode SSA, model dapat digambarkan dengan bantuan *Linear Recurrent Formula* (LRF).

$$y_{i+d} = \sum_{k=1}^d a_k y_{i+d-k} \text{ untuk } 1 \leq i \leq N-d \tag{3}$$

Perbedaan antara kedua metode peramalan menyebabkan dalam *approximate continuation*-nya biasanya memberikan hasil yang berbeda [6]. Kedua metode peramalan memiliki dua tahap umum: *diagonal averaging* dan *continuation*. Untuk *R-forecasting*, *diagonal averaging* digunakan untuk memperoleh seri rekonstruksi, dan *continuation* dilakukan dengan menerapkan LRF. Sedangkan dalam *V-forecasting*, kedua tahap digunakan dalam urutan terbalik; pertama, vektor peramalan dilakukan dan kemudian *diagonal averaging* memberikan nilai ramalan. Untuk mendapatkan M periode ke depan, metode vektor menggunakan prosedur $M+L-1$ langkah. Perbedaan ramalan antara metode *recurent* dan vektor secara teoritis akan sulit dijelaskan. Hal yang bisa dilakukan untuk melihat perbedaan antar kedua metode peramalan salah satunya dengan melihat akurasi hasil ramalannya.

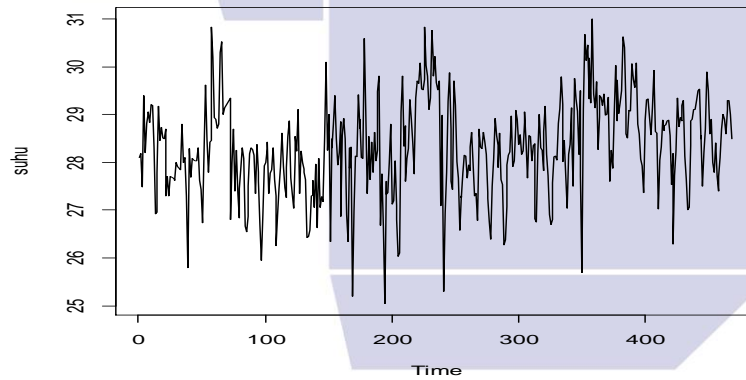
Akurasi Hasil Peramalan

Setelah dilakukan tahapan demi tahapan, selanjutnya menghitung seberapa besar ketepatan peramalan tersebut. Dalam Makridakis [8] dijelaskan bahwa ukuran ketepatan peramalan dipandang sebagai kriteria penolakan untuk memilih suatu metode peramalan sehingga dapat digunakan untuk menentukan kemungkinan yang lebih baik. Metode yang digunakan adalah *Mean Absolute Percentage Error* (MAPE). Berikut adalah bentuk perhitungannya:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{T_t - \hat{Y}_t}{Y_t} \right| \times 100\% \tag{4}$$

Hasil dan Diskusi

Data yang digunakan adalah data bulanan suhu Jakarta (*observatory*) dari bulan Januari 1979 hingga Desember 2017. Peramalan akan dilakukan untuk 10 tahun mendatang.



Gambar 1 Plot Suhu Jakarta Januari 1973 hingga Desember 2011

Berdasarkan plot di atas, suhu Jakarta memiliki fluktuasi yang tajam. Ini terlihat dari turun naiknya suhu pada beberapa periode dengan ekstrim. Plot memiliki pola yang cenderung tren lambat. Pola periodik tidak nampak secara jelas, tetapi menurut informasi tambahan dari BMKG, suhu cenderung memiliki periode 12 bulanan. Untuk itu, diasumsikan periodisitas data adalah 12 sehingga dalam pemilihan parameter *window length* perlu diperhatikan besarnya L yang bersesuaian dengan periode suhu tersebut.

1. Dekomposisi

Pada penelitian ini, dengan melihat fluktuasi yang tajam maka diperlukan *window length* L yang cukup besar agar dekomposisi yang dihasilkan lebih detail dan memiliki *separability* yang baik. Dalam penelitian ini dipilih *window length* $L = 204$ proporsional dengan periodisitas data suhu (12 bulan). Tahap SVD yang akan menghasilkan *eigen triple* sebanyak 204 ($L=204$).

2. Singular Value

Singular value diperoleh dari akar kuadrat *eigenvalue* ($\sqrt{\lambda_1}, \dots, \sqrt{\lambda_L}$) dari matriks $S = \mathbf{X}\mathbf{X}^T$.

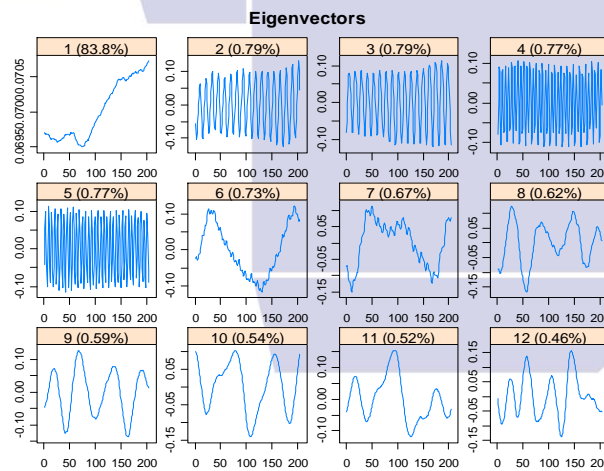


Gambar 2 Logaritma dari 50 *Eigenvalue* Pertama

Dari plot di atas, terlihat bahwa *eigenvalue* pertama memiliki nilai logaritma yang paling besar. *Eigenvalue* selanjutnya cukup jauh berbeda dengan yang pertama dan cenderung menurun. Dari penurunan ini dapat diketahui dua hal, yaitu komponen harmonik dan *noise*. Komponen harmonik dengan periodisasi yang berbeda akan memiliki nilai singular yang berdekatan [2]. Ini terjadi pada nilai singular 2 hingga 19. Sedangkan penurunan perlahan menunjukkan adanya *noise* (mulai terjadi pada *eigenvalue* ke-20).

3. Eigenvector

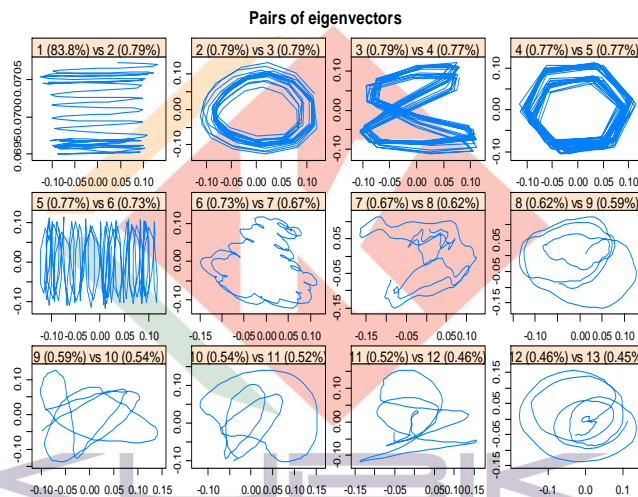
Selanjutnya, *eigenvector* akan disajikan dalam satu dimensi untuk mempermudah menemukan komponen dalam data.



Gambar 3 Ke-12 Eigenvector Pertama

Berdasarkan gambar di atas, eigenvector pertama memiliki pola tren, eigenvector kedua hingga kelima memiliki pola harmonik yang menunjukkan adanya unsur periodik (misalnya pola musiman). Eigenvector keenam dan selanjutnya memiliki pola yang aperiodik yang menunjukkan noise.

4. Pairwise Eigenvector

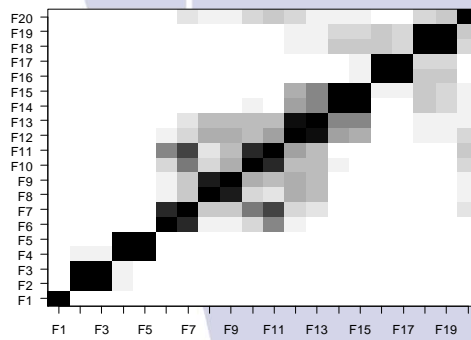


Gambar 4 Ke-12 Eigenvector Berpasangan (Pairwise Scatterplot)

Pola lingkaran mengidentifikasi adanya komponen periodik. Semakin menyerupai lingkaran, periode komponen akan semakin besar. Dengan melakukan pemeriksaan terhadap periodenya, diketahui bahwa untuk pasangan eigenvector 2 dan 3 memiliki periode 12. Komponen periodik juga teridentifikasi pada pasangan eigenvector 4 dan 5 dengan periode 6. Adanya kesesuaian periode ini (6 dan 12) akan menjadikan eigenvector ke-2, 3, 4, dan 5 kedalam satu komponen. Sementara pasangan eigenvector ke-6 dan ke-7 dan eigenvector selanjutnya menunjukkan pola yang tidak beraturan yang diidentifikasi sebagai noise.

5. Rekonstruksi

Dari informasi eigentriple yang terbentuk, dapat disimpulkan bahwa terdapat 3 group ($r=3$) yang terdiri dari komponen trend, periodik, dan noise. Walaupun identifikasi awal terhadap komponen telah dilakukan, tetapi keterpisahan (separability) dari komponen yang mencirikan seberapa baik komponen yang berbeda dapat dipisahkan satu sama lain belum diketahui.



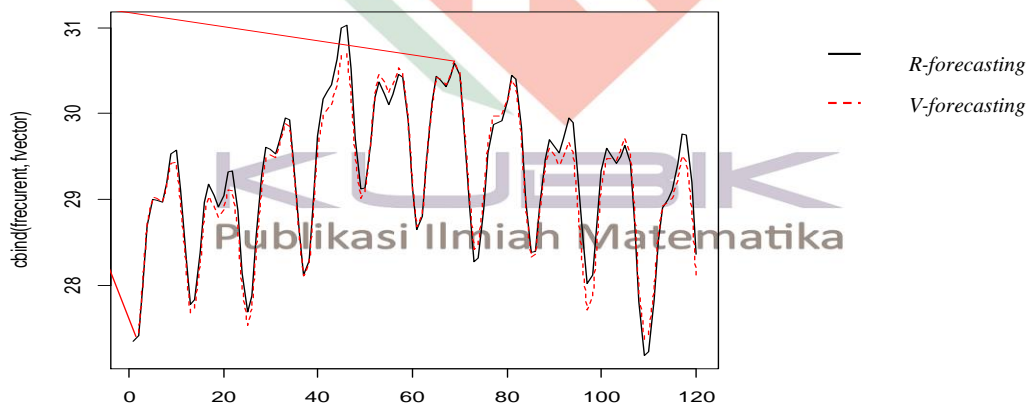
Gambar 5 Matriks korelasi ω untuk 20 komponen pertama

Dari matriks di atas, besarnya korelasi ditunjukkan oleh gradasi warna dari warna muda hingga tua. Semakin tua warnanya semakin tinggi korelasinya. Dari matriks di atas terlihat bahwa komponen satu tidak berkorelasi dengan komponen lainnya. Komponen 2 memiliki korelasi yang sangat kuat dengan komponen 3, begitu pun dengan komponen 4 dan 5. Ini menunjukkan bahwa komponen 2 – 3 satu kelompok. Untuk komponen 6-19, terdapat korelasi yang cukup kuat dengan komponen lainnya yang diidentifikasi *noise*.

Dari informasi eigentriple dan *separability*-nya, komponen pertama adalah tren. Komponen 2 dan 3 memiliki korelasi yang kuat dan dari hasil pemeriksaan periode pada pasangan eigenvectornya memiliki periode 12. Tidak jauh berbeda dengan komponen ke- 4 dan 5 yang memiliki periode 6. Dengan demikian, komponen 2, 3, 4 dan 5 dikelompokkan sebagai periodik. Sementara komponen 6 dan seterusnya yang memiliki pola aperiodik dan memiliki korelasi yang cukup kuat antar satu dengan yang lain dikelompokkan sebagai *noise*.

6. Hasil Peramalan dengan R-Forecasting dan V-Forecasting

Setelah didapatkan koefisien LRF, maka peramalan dapat dilakukan.



Gambar 6 Hasil SSA *R-forecasting* dan *V-forecasting* Suhu untuk 10 Tahun Mendatang

Jakarta

Dari grafik di atas, terlihat bahwa peramalan suhu untuk 10 tahun mendatang baik dengan metode *R-forecasting* maupun *V-forecasting* menghasilkan ramalan yang tidak jauh berbeda. Artinya aproksimasi kekontinuannya hampir sama. Ini terjadi karena parameter *window length L* dan *grouping* yang sesuai dengan data. Hasil peramalan menunjukkan bahwa suhu akan berfluktuasi dari 27°C hingga 31°C. Suhu tertinggi diramalkan akan terjadi pada tahun ke-4 (tahun 2015).

7. Perbandingan Akurasi SSA *R-forecasting* dan *V-forecasting*

Berikut adalah hasil MAPE hasil peramalan SSA *R-forecasting* dan *V-forecasting*.

Tabel 1. Hasil MAPE untuk Akurasi SSA *R-forecasting* dan *V-forecasting*

Metode SSA	MAPE
<i>R-Forecasting</i>	5.0029
<i>V-Forecasting</i>	4.0067

Berdasarkan hasil perhitungan, didapat MAPE untuk hasil permalan dengan SSA *R-forecasting* sebesar 5,0029 bernilai lebih besar dibandingkan dengan SSA *V-Forecasting* sebesar 4,0067. Ini menunjukkan bahwa peramalan suhu untuk long horizon lebih akurat dengan menggunakan *V-Forecasting* dibandingkan dengan *R-Forecasting*. Hasil permalan kedua metode telah lebih kecil dari 0%, artinya akurasi hasil peramalan sangat baik [9].

Kesimpulan

SSA *R-forecasting* dan *V-Forecasting* pada suhu Jakarta menggunakan *window length* $L=204$ dan $r=3$ menghasilkan ramalan yang tidak jauh berbeda (aproksimasi kekontinuannya hampir sama). Ini terjadi karena parameter *window length* L dan *grouping* yang sesuai dengan data. Secara empiris didapat bahwa peramalan dengan metode *V-forecasting* lebih akurat dibanding dengan *R-forecasting*.

Referensi

- [1] A. Pepelyshev and A. Zhigljavsky, "Assesing The Stability of Long-Horizon SSA Forecasting," *Stat. Its Inference*, vol. X, no. XXXX, pp. 1–6, 2009.
- [2] H. Hassani, "Singular Spectrum Analysis: Methodology and Comparison," *J. Data Sci.*, vol. 5, pp. 239–257, 2007.
- [3] D. S. Broomhead and G. P. King, "Extracting qualitative dynamics from experimental data," *Phys. D*, vol. 20, pp. 217–236, 1986.
- [4] N. K. Myung, "Singular Spectrum Analysis," University of California, 2009.
- [5] P. Vautard, R. Yiou and M. Ghil, "Singular-spectrum analysis: A toolkit for short, noisy chaotic signal," *Phys. D*, vol. 58, pp. 95–126, 1992.
- [6] N. Golyandina, V. Nekrutkin, and A. Zhigljavsky, *Analysis of Time Series Structure: SSA and Related Techniques*. New York: Chapman & Hall/crc, 2001.
- [7] J. B. Elsner and A. A. Tsonis, *Singular Spectrum Analysis. A New Tool in Time Series Analysis*. New York: Plenum Press, 1996.
- [8] S. Makridakis and S. C. Wheelwright, *Metode dan Aplikasi Peramalan. Edisi Ke-2. Terjemahan Hari Suminto*. Jakarta: Binarupa Aksara, 1999.
- [9] C. D. Lewis, *International and Business Forecasting Methods*. London: Butterworths, 1982.