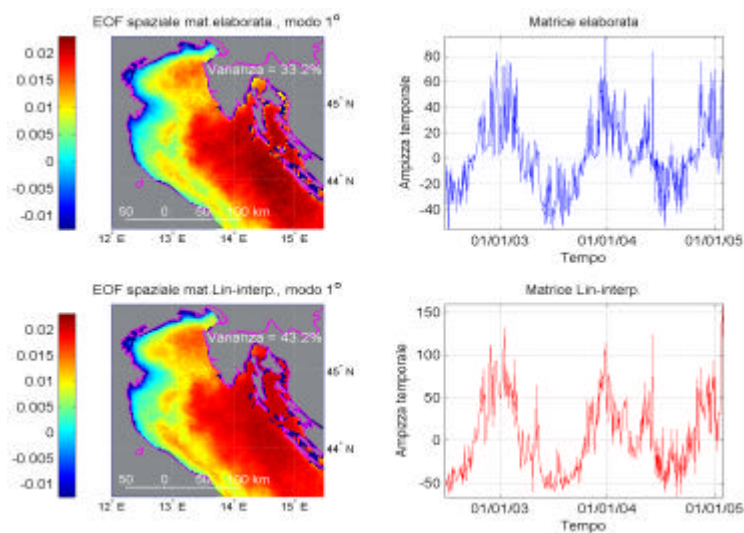


INTERPOLAZIONE DI DATI ATTRAVERSO METODO EOF: APPLICAZIONE A IMMAGINI MODIS NEL MEDIO E ALTO ADRIATICO



ŽIVKO JUŽNIC-ZONTA, ELENA MAURI E PIERRE-MARIE POULAIN

Approvato da:

.....

Dr. Alessandro Crise
Direttore del Dipartimento di Oceanografia



Indice

1. Introduzione	3
2. Dati.....	3
2.1 Elaborazione del set di dati	3
2.2 Statistica semplice applicata ai dati iniziali e a quelli interpolati.....	5
3. Metodo d'interpolazione attraverso metodo EOF.....	10
3.1 Diagramma di flusso della funzione <i>InterpEOF.m</i>.....	11
3.2 Errore del ciclo e numero di modi massimo (e la variante logaritmica).....	14
3.2 Risultati ottenuti per mezzo del metodo EOF, variante lineare.....	18
3.4 Risultati ottenuti per mezzo del metodo EOF, variante logaritmica.....	24
4. Conclusioni	29
5. Referenze.....	29

1. Introduzione

L'obiettivo di questo rapporto tecnico è di testare il metodo di interpolazione con EOF (Empirical Orthogonal Function) (Beckers e Rixen, 2003) su mappe di concentrazione di clorofilla nel medio e alto Adriatico. I dati sono quelli derivati dal sensore MODIS (AQUA). Il metodo applicato è stato descritto e utilizzato in un precedente rapporto tecnico (Južnic -Zonta et al., 2004) per lo studio dell'interpolazione del K490 nel Golfo di Trieste. Con questo metodo si calcola il numero ottimale di modi EOF, tramite il test di cross-validazione, con il quale interpolare il set di dati. Questo metodo, oltre a fornire un set di immagini complete, descrive attraverso i modi principali e le loro ampiezze la varianza del segnale nel data set. In questo rapporto viene descritto solo il metodo ma non si tratterà dell'interpretazione dei risultati.

2. Dati

Il set di dati utilizzato nell'interpolazione attraverso il metodo EOF, consiste in dati giornalieri di clorofilla dal 1 luglio 2002 al 1 febbraio 2005. La clorofilla è espressa in mg/m^3 , con una risoluzione spaziale di 1 km. Tali dati sono forniti dal sensore MODIS e sono messi a disposizione dalla NASA (National Aeronautics and Space Administration) sulla pagina web: <http://daac.gsfc.nasa.gov/data/datapool/>. I dati scaricati sono già corretti per l'effetto atmosferico (Livello 2). I passi giornalieri con copertura totale o pressoché totale non sono stati scaricati, per cui il set di dati ha un totale di 594 immagini, su un totale di 951 giorni. I dati sono acquisiti ogni giorno attorno alle ore 12:00 GMT.

2.1. Elaborazione del set di dati

Il programma WIM/WAM (Windows Image Manager / Windows Automatic Module) è stato utilizzato per estrarre le aree di interesse e applicare la proiezione. La matrice di dati ottenuta da ciascun passo ha una risoluzione uniforme dei pixel, in formato *hdf*.

Il data set è stato successivamente ridotto di risoluzione tramite un programma Matlab, che prevede una finestra mobile 3×3 , sui cui elementi è calcolata la mediana. Se più del 60% dei pixel

della finestra sono noti, al valore centrale della finestra è attribuito il valore della mediana, mentre nel caso la percentuale sia minore, il valore attribuito è zero.

Per ciascun passo sono stati eliminati e i dati presenti sulla terra. Per eliminare i dati terrestri si è sviluppata un'*immagine maschera* (figura 1b) che presenta sulla terra ed in qualche pixel vicino alla costa, valori NaN. Si è stabilito che se il pixel è rappresentato in meno del 30% dei passi viene mascherato e considerato al pari della terra. In figura 1a sono rappresentate le percentuali di dati mancanti per pixel nell'area e per il periodo di interesse.

Il set di dati così ottenuto è stato strutturato in una matrice spazio-temporale, le cui righe sono i pixel relativi al mare e le cui colonne rappresentano i dati dei passi giornalieri. In questa matrice, i pixel nuvolosi corrispondono al valore zero.

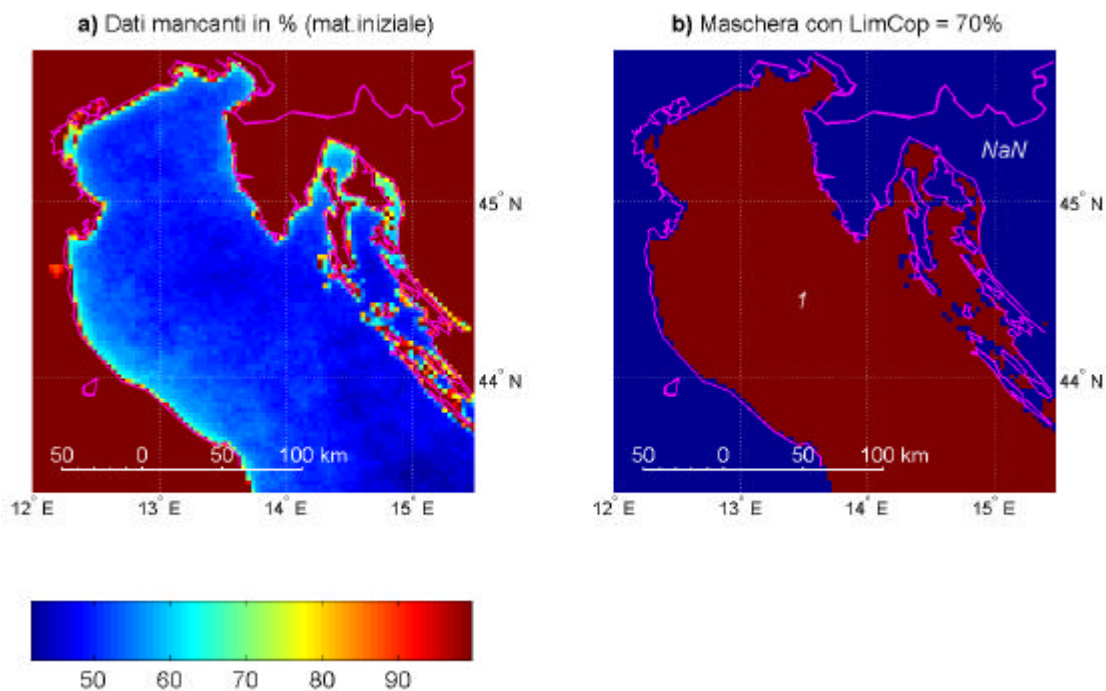


Figura 1: a) Percentuale di dati mancanti per pixel e b) l'immagine maschera ottenuta ponendo a NaN tutti i pixels che presentano piu' del 70% di dati mancanti nell'intervallo di tempo studiato.

Ulteriori riduzioni dei dati si sono operate:

- togliendo due immagini giornaliere: il 20 gennaio 2004 e il 11 febbraio 2004, con problemi di navigazione;
- ponendo i valori superiori a $64 \text{ mg} / \text{m}^3$ a zero;

- utilizzando un metodo statistico iterativo. La varianza viene calcolata per ogni singolo pixel nel tempo e si costruisce la funzione di distribuzione cumulata, *CDF*. Viene poi calcolato il valore x_{99} e x_{68} , definito come gli estremi superiori dei valori che si verificano per il 99% e il 68% delle probabilità. Il x_{99} viene usato come soglia sopra la quale i dati vengono uguagliati a zero. A questo punto non sono più utilizzati i dati superiori a x_{99} ed è ricalcolata una nuova *CDF*. Il procedimento è reiterato fino a convergenza del valore x_{68} per ogni pixel (tolleranza del centesimo).
- eliminando le immagini con una percentuale di dati mancanti (o messi a zero a causa della nuvolosità o dei test citati qui sopra) maggiore al 95%. Tale scelta è stata compiuta, perché questi dati non contribuiscono sostanzialmente alla varianza del segnale.

Il set di dati così ottenuto è composto da 499 immagini giornaliere.

2.2. Analisi statistica applicata ai dati iniziali e a quelli ridotti

Le figure 2a e 2b rappresentano rispettivamente i minimi e i massimi temporali della matrice iniziale. I valori massimi particolarmente elevati e con una pronunciata discontinuità spaziale, rispetto al resto del bacino, si osservano lungo la costa occidentale. I minimi sono relativamente elevati in certi punti lungo la costa italiana. Le figure 3a e 3b rappresentano i minimi e i massimi spaziali riferiti alla matrice ridotta. Dal confronto delle due matrici iniziale e ridotta, si osserva minor variabilità spaziale dovuta alla riduzione dei valori massimi mentre i minimi non cambiano, poiché i valori piccoli non sono coinvolti nella riduzione statistica iterativa. Si noti la diversa scala di colore con valore massimo di 64 nella matrice di dati ridotti mentre in quella iniziale il massimo è molto più alto.

Le medie temporali della matrice iniziale e di quella ridotta sono rappresentate rispettivamente nelle figure 4a e 4b. La media della matrice ridotta è leggermente più bassa su tutto il bacino, ma la struttura spaziale evidenziata è la stessa.

In figure 5a e 5b sono riportate le deviazioni standard temporali relative alla matrice iniziale e a quella ridotta. Si osserva che la deviazione standard più grande si concentra nella regione influenzata dal fiume Po e lungo la costa occidentale. La deviazione standard della matrice ridotta è più piccola e ha una distribuzione spaziale più uniforme. Spariscono valori isolati di deviazione standard elevati e in generale tutte le deviazioni standard sono ridotte rispetto ai dati iniziali.

I coefficienti d'asimmetria (skewness) temporali sono riportati rispettivamente per la matrice iniziale e quella ridotta per ogni pixel in figure 6a e 6b. La matrice iniziale presenta una notevole asimmetria, con dispersione dei dati verso la destra rispetto alla media, mentre la matrice ridotta indica che generalmente il segnale ha una distribuzione che si avvicina a quella normale.

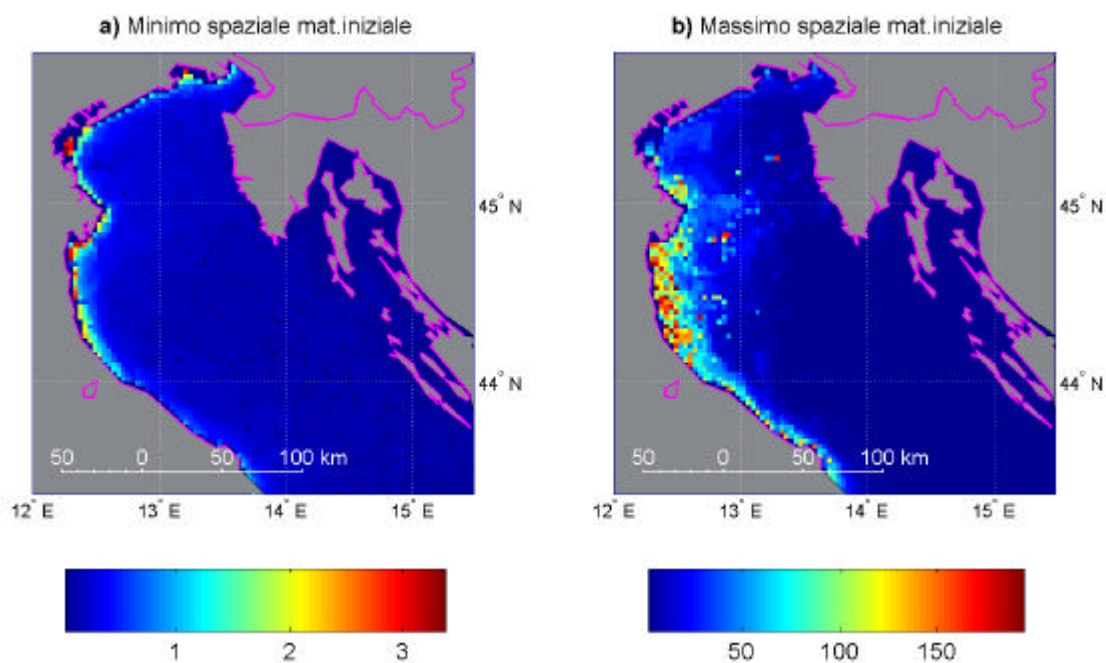


Figura 2: a) Minimo e b) massimo temporali riferiti alla matrice iniziale non ridotta.

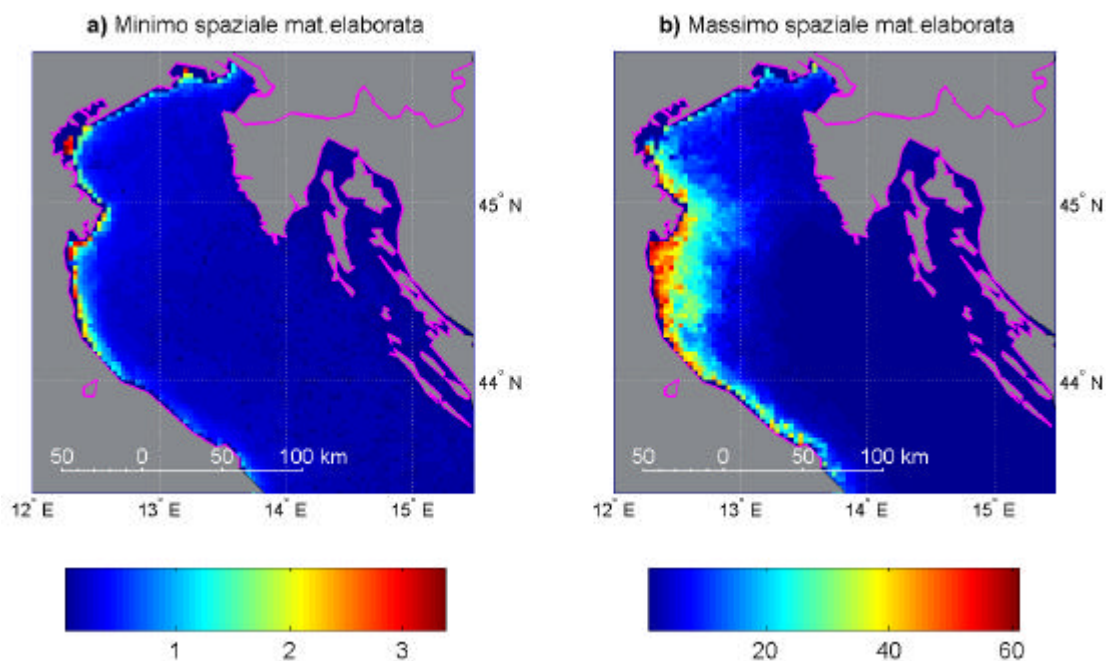


Figura 3: a) Minimo e b) massimo temporali riferiti alla matrice ridotta (o elaborata).

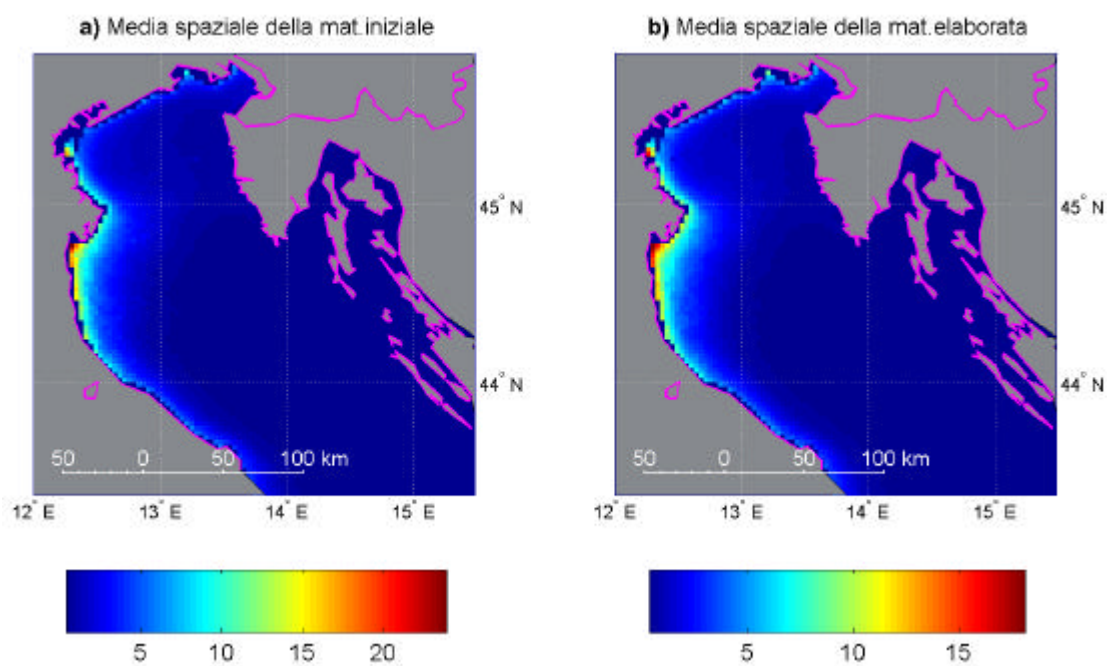


Figura 4: a) Media temporale della matrice iniziale e b) quella della matrice ridotta.

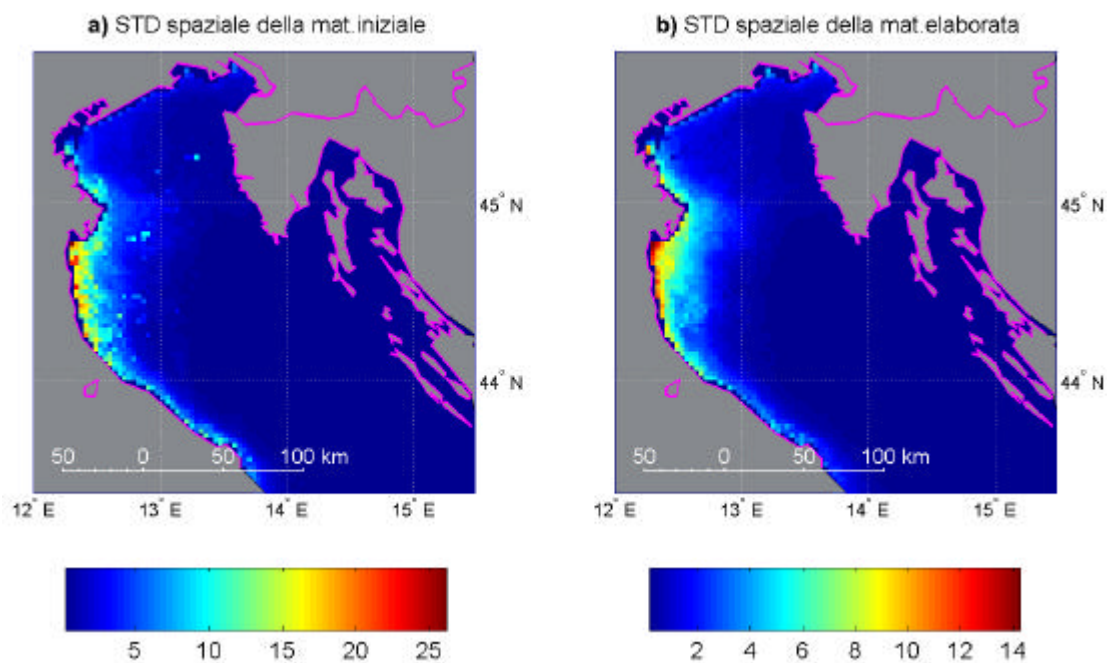


Figura 5: a) Deviazione standard temporale della matrice iniziale e b) quella della matrice ridotta.

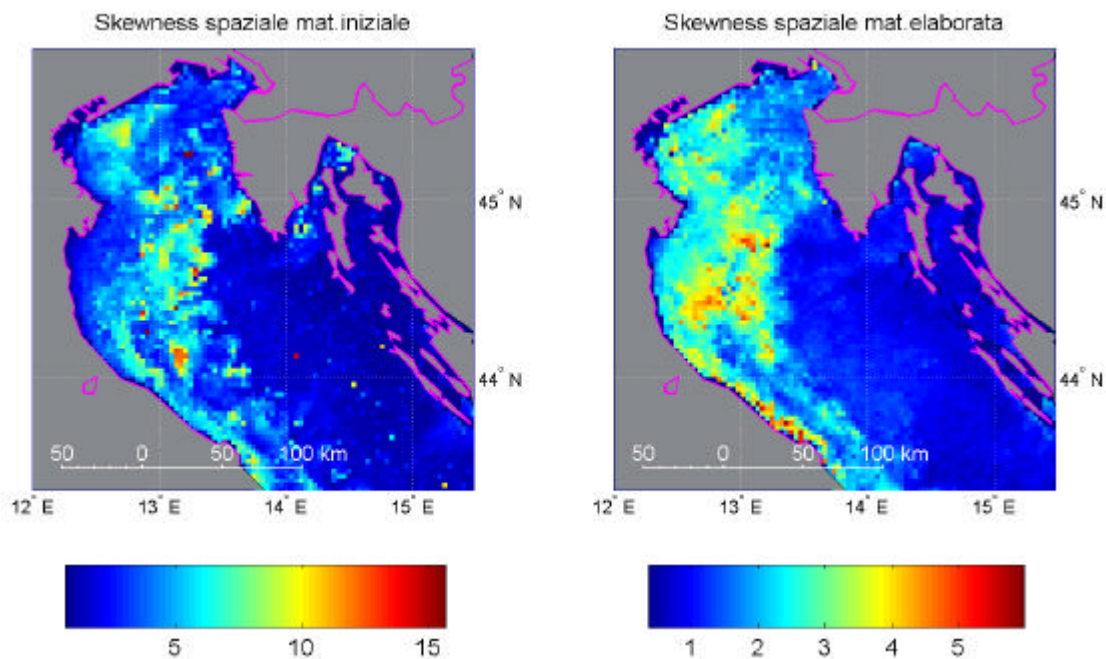


Figura 6: a) Coefficiente di asimmetria (skewness) della matrice iniziale e b) della matrice ridotta.

In figura 7 è riportata la percentuale di dati mancanti (o messi a zero) per ogni immagine giornaliera verso il tempo. I dati utilizzati sono quelli ottenuti dopo l'eliminazione dei valori superiori a 64 mg/m^3 e l'applicazione del metodo statistico iterativo. Durante i periodi invernali, la percentuale di dati mancanti è maggiore che non durante i periodi estivi.

La percentuale temporale media dei dati mancanti della matrice iniziale è del 42,1%, dopo la riduzione della matrice (utilizzando x_{99}), la percentuale è del 42,8%. Se si utilizza x_{95} la percentuale media sale a 44,2%.

La distribuzione spaziale della percentuale (per ogni pixel) di dati mancanti, riferita alla matrice ridotta, è riportata nella figura 8. Si distinguono due regioni: la prima con un'elevata percentuale di dati mancanti lungo la costa, e una seconda con un percentuale relativamente bassa nel centro e al sud del bacino.

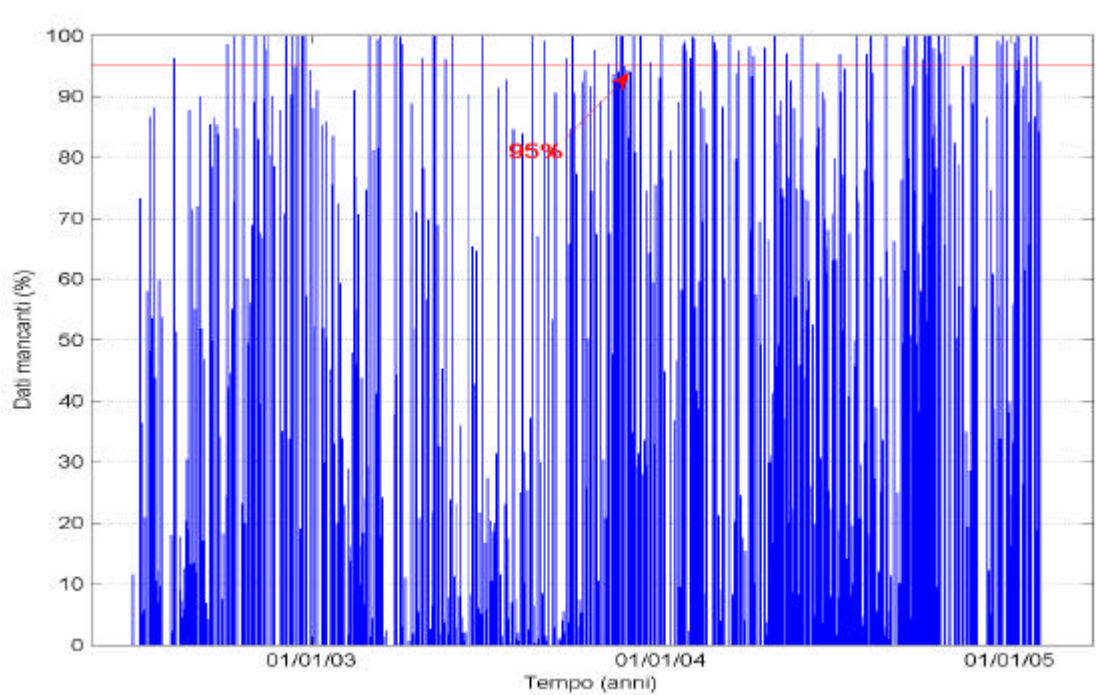


Figura 7: Percentuale di dati mancanti in funzione del tempo, riferita alla matrice iniziale ridotta dal metodo statistico iterativo. Ogni barra rappresenta un'immagine giornaliera.

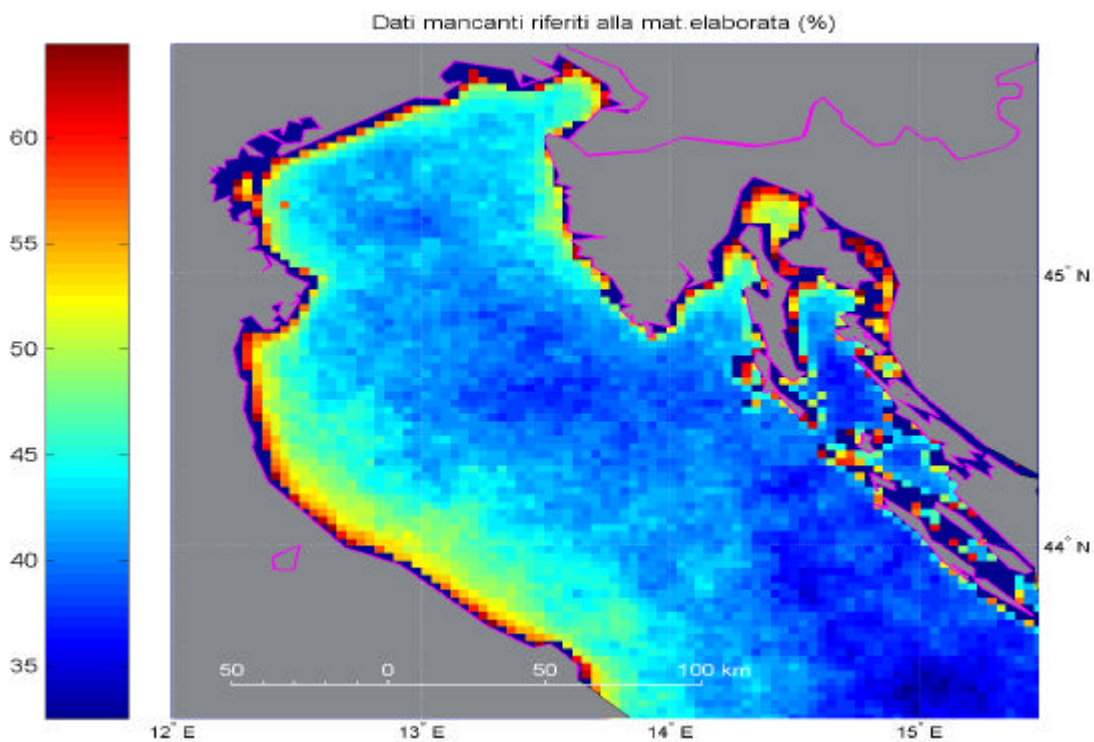


Figura 8: Percentuale dei dati mancanti per ogni pixel, riferito alla matrice ridotta.

3. Metodo d'interpolazione attraverso metodo EOF

Il metodo di interpolazione attraverso EOF, proposto da Beckers e Rixen (2003), è riassunto brevemente qui di seguito, per maggiori dettagli riguardanti la parte teorica sulle funzioni ortogonali empiriche, EOF, si rimanda il lettore a Emery e Thomson (1997). Il metodo prevede:

- a) Si calcola la media e la deviazione standard temporali della matrice dei dati ridotti, $\mathbf{y}(\vec{x}, t)$; si costruisce la matrice normalizzata e si inizializza i dati mancanti con il valore zero. La

matrice normalizzata è definita come:
$$N(\vec{x}_i, t) = \frac{\mathbf{y}(\vec{x}_i, t) - \overline{\mathbf{y}(\vec{x}_i, t)}_{tempo}}{std(\mathbf{y}(\vec{x}_i, t))}$$
 per ogni

$i = 1, 2, \dots, \text{num.pixel}$.

- b) Si compiono i seguenti passi fino a convergenza dei cicli d'interpolazione:

⇒ per ottenere una prima stima delle *EOF* spaziali e ampiezze temporali, è eseguita la decomposizione ai valori singolari, *SVD* (singular value decomposition) della matrice dei dati ridotti, $\mathbf{y}(\vec{x}, t)$;

⇒ i valori mancanti sono rimpiazzati dai valori ottenuti dalla combinazione lineare dell'ampiezza temporale con l'*EOF* spaziale del primo modo:

$$D_{i,j} = \mathbf{y}(\vec{x}_i, t_j) = \sum_{p=1}^k a_p(t_j) \mathbf{f}_{pi}, \text{ dove } k = 1 \text{ e } j = 1, 2, \dots, \text{num.immagini};$$

⇒ ottenuta una stima dei dati mancanti, si ripetono i due punti precedenti con i nuovi valori. Tale procedimento è iterato fino a raggiungere, per mezzo di una determinata tolleranza, la convergenza.

- c) Il passo successivo consiste nel rifare il ciclo di convergenza usando sia il primo che il secondo modo. L'avvalersi di una quantità crescente di modi, consente di ottenere dati interpolati con $1, 2, \dots, k$ *EOF*. Da questo momento si userà l'espressione *numero di modi k*, per indicare che nell'interpolazione dei dati sono stati usati $1, 2, \dots, k$ modi.

- d) Il numero di modi ottimo è ottenuto dal test di cross-validazione: s'inserisce (casualmente) alcuni valori zero al posto dei dati buoni. Il numero dei dati presi per il test è un compromesso tra la robustezza statistica e la stabilità dell'informazione principale contenuta nei primi modi. Il numero di modi ottimo corrisponde a il massimo della funzione di cross-correlazione, che è funzione del numero di modi usati per l'interpolazione.

e) Una volta che il numero di modi ottimo è noto, l'interpolazione è ripetuta usando tutti i dati del set normalizzato.

Si nota che dalla matrice ridotta utilizzata per il ciclo d'interpolazione viene tolta immagine media (figura 4b) e la matrice viene di seguito normalizzata con la deviazione standard temporale della figura 5b. Questa procedura è diversa di quella di Južnic-Zonta et al. (2004) per cui dalla matrice dei dati ridotti è sottratta la media spazio-temporale.

3.1. Diagramma di flusso della funzione *InterpEOF.m*

Nella figura 9 è rappresentato il diagramma di flusso della funzione *InterpEOF.m*. Essa ha come variabili in entrata la matrice di dati ridotti da interpolare, *Cubo_iniz*, il numero di modi massimo da usare nel test di cross-validazione *nmodi*, l'errore di troncamento del ciclo di convergenza, *err*, dove l'errore del ciclo è ottenuto usando un numero di modi *k* ed è definito come:

$$err_k = \left| \sum_{i=1}^k S_{i,i}^j - \sum_{i=1}^k S_{i,i}^{j-1} \right|, \quad (3.1)$$

dove *S* è la matrice degli autovalori singolari ottenuta dalla decomposizione *SVD* per il ciclo di convergenza *j*-esimo, con $j=1,2,\dots,numdicicli$. Altre variabili in entrata sono il percentuale massimo dei dati mancanti per ogni immagine giornaliera, *LimCoper* (fissato a 95%), il valore limite massimo che può assumere il dato, *LimVal* (fissato a 64 mg/m^3), il limite di probabilità *LimStat* (fissato a 99%), usato dal metodo statistico iterativo, e per ultimo, una stringa di caratteri, '*str*', che indica la variante del metodo d'interpolazione da usare. Sono due le varianti del metodo possibili: quella *lineare* e quella *logaritmica*. Quella logaritmica è descritta nel paragrafo successivo.

In uscita è restituita la matrice interpolata, *Cubo_ott_interp*, il coefficiente di cross-correlazione, *cross_corr*, e l'errore quadratico medio del residuo, *RMS*, entrambi in funzione dei modi usati durante il test di cross-validazione. Infine, per mezzo della cella *Stat*, è riportato l'errore del ciclo in funzione del numero di cicli, riferito al test di cross-validazione e all'interpolazione ottima.

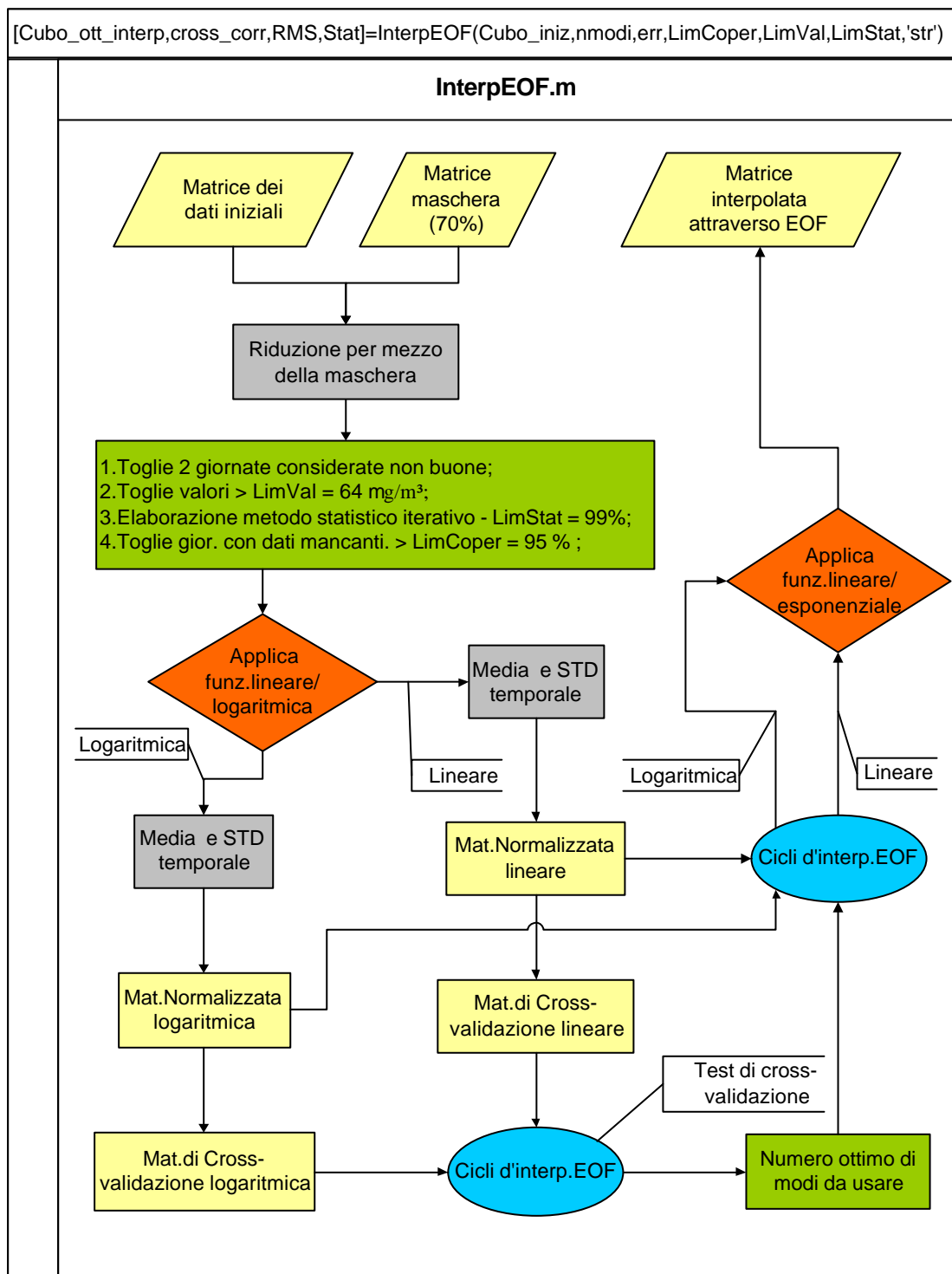


Figura 9: Diagramma di flusso della funzione *InterpEOF.m*.

Descriviamo brevemente il flusso logico e le varianti della funzione *InterpEOF.m*. La matrice dei dati in entrata è la matrice ridotta con i metodi già descritti. Si possono ora scegliere 2 diversi modi di procedere: *variante lineare* o *variante logaritmica*. Nel primo caso la matrice è

normalizzata, mentre nel modo logaritmico si calcola il logaritmo naturale di ogni elemento diverso dal valore zero e poi si normalizza la matrice. Da ora in poi se non viene specificata la variante si intenderà quella lineare.

Si procede (indipendentemente dalla variante scelta) al calcolo del numero ottimale di modi EOF da utilizzare per l'interpolazione attraverso il metodo di cross-validazione. Il metodo parte dalla sostituzione casuale con zeri del 3% dei dati dalla matrice normalizzata e consiste nel successivo confronto di questi valori con quelli interpolati utilizzando k modi. Il coefficiente di cross-correlazione indica la bontà dell'interpolazione, ed è definito come:

$$\mathbf{r}_k = \frac{\frac{1}{M-1} \sum_{i=1}^M (x_i - m_x)(y_{i,k} - m_y^j)}{\mathbf{s}_x \mathbf{s}_y^j}, \quad (3.2)$$

dove x_i è l' i -esimo dato prelevato casualmente dalla matrice dei dati normalizzati con $i = 1, 2, \dots, M$, dove M è la quantità di tali dati, $y_{i,j}$ è l' i -esimo dato interpolato tramite il k -esimo numero di modi; m_x è la media dei dati x_i , m_y^k è la media dei dati $y_{i,k}$ ricavato con il rispettivo numero di modi e \mathbf{s}_x , \mathbf{s}_y^k sono le deviazioni standard rispettive. Poiché il coefficiente indica come covariano i due set di dati, potrebbe accadere che questi differiscano per un termine costante e/o per un fattore moltiplicativo. E' quindi necessario esaminare la distribuzione delle differenze tra i valori interpolati e i valori reali. Pertanto, oltre al coefficiente di cross-correlazione si è calcolato anche l'errore quadratico medio del residuo, RMS , definito come:

$$RMS_k = \sqrt{\sum_{i=1}^M \frac{(x_i - y_{i,k})^2}{M}}, \quad (3.3)$$

Una volta definito il numero di cicli da usare nell'interpolazione della matrice normalizzata, si eseguono i cicli per l'interpolazione vera e propria. Nel caso della variante logaritmica, la matrice restituita dal ciclo d'interpolazione è ritrasformata per mezzo della funzione esponenziale.

3.2. Errore del ciclo e numero di modi massimo (e la variante logaritmica)

La definizione dell'errore del ciclo data dall'equazione (3.1) è ritenuta più appropriata di quella proposta da Južnic-Zonta et al. (2004), definita come:

$$err_k = \left| \max(D_k^j - D_k^{j-1}) \right|, \quad (3.4)$$

dove l'indice j indica il j -esimo ciclo e l'indice k indica il numero di modi usati. Lo svantaggio della (3.4) è di fornire una misura della convergenza troppo restrittiva, poiché essa si basa sul valore di un solo punto della matrice interpolata. Non si può escludere infatti, che tale punto possa rappresentare un'anomalia. L'equazione (3.1) fornisce una misura della convergenza generale della matrice, poiché esprime in modo indiretto il residuo della varianza contenuta. Dalla teoria della perturbazione per la decomposizione *SVD*, presentata da Beckers e Rixen (2003), infatti, è noto che passando da un ciclo all'altro la varianza nei modi principali aumenta a discapito di quelli minori. Quindi ad un certo numero di cicli, usando un numero di modi k , il trasferimento di varianza sarà molto piccolo per ogni ulteriore passo. L'equazione (3.1) è quindi capace di misurare tale trasferimento e, fissato un errore di ciclo minimo (errore di troncamento), di terminare il ciclo che corrisponde alla convergenza.

In figura 10 (variante lineare) è rappresentato l'errore del ciclo in funzione del numero di cicli, per i 10 modi. La convergenza, in tutti i modi, è raggiunta per valori dell'errore del ciclo tra 10 e 20 unità di varianza. Si è quindi fissato l'errore di ciclo minimo (errore di troncamento) a 15. Quando si parla di n° modo 1, n° modo 2, n° modo 3, si intende modo 1, modo 1 + modo 2, modo 1 + modo 2 + modo 3, ecc.).

Per l'interpolazione della matrice di cross-validazione è usato un numero di modi massimo uguale a 50 (cinquanta). In figura 11a sono rappresentati il coefficiente di cross-correlazione e l'*RMS* riferiti all'interpolazione EOF di variante lineare. Tale interpolazione è portata a termine dopo 186.7 minuti di lavoro macchina (Athlon 2400+, 712 RAM e Windows XP). Si osserva che la bontà dell'interpolazione è massima per il coefficiente di cross-correlazione uguale a 0.884, utilizzando 23 modi.

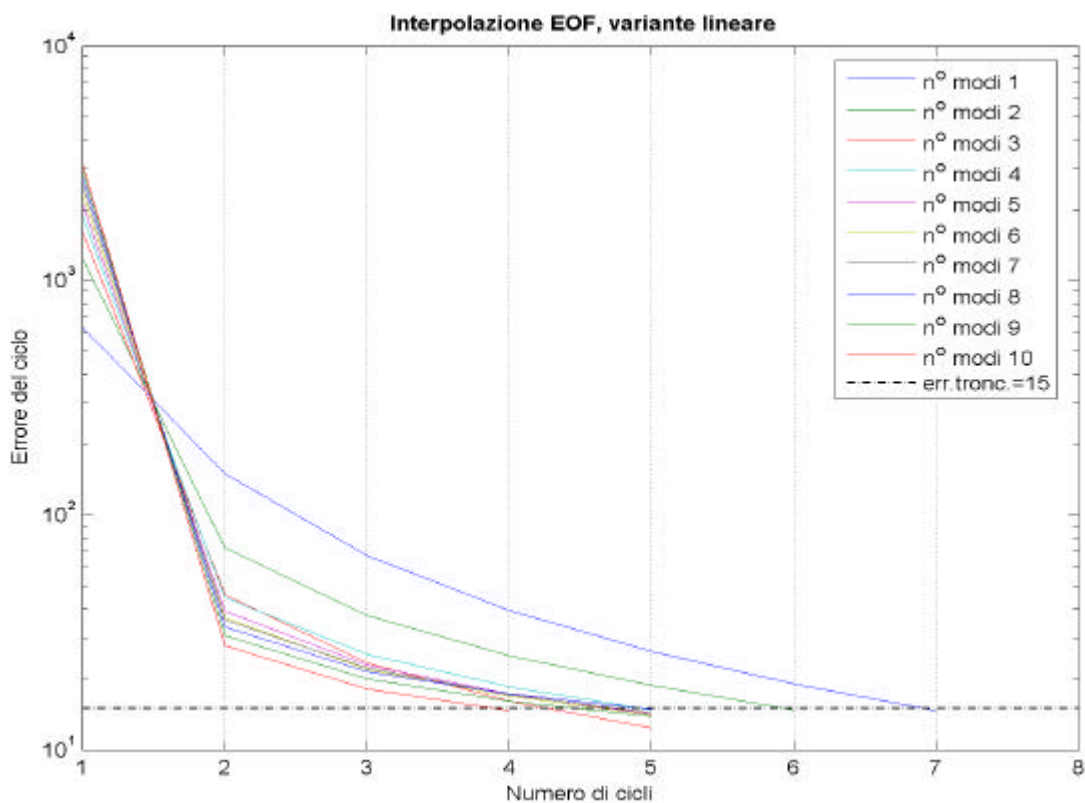


Figura 10: Errore del ciclo (scala logaritmica) in funzione del numero di cicli per 10 modi usati nell'interpolazione EOF, variante lineare.

Nel caso della variante logaritmica, il cui coefficiente di cross-correlazione e l'*RMS* sono rappresentati in figura 11b. Il metodo trova un massimo di 0.966 utilizzando 34 modi, e impiega solo 159.0 minuti. I coefficienti di cross-correlazione sono diversi per le due varianti e quello corrispondente alla variante logaritmica è più alto. Va tenuto comunque presente che i valori estratti casualmete sono solitamente circondati da valori noti e non raggruppati come nel caso delle nuvole. L'indice che si utilizza quindi è calcolato in condizioni relativamente più facili da interpolare rispetto alle immagini reali. Sarebbe quindi auspicabile un test di cross-validazione che riproduca condizioni più realistiche. Il prelievo, simulando la nuvolosità reale, dal set di dati per il test di cross-validazione, è già stato impiegato da Beckers e Rixen (2003).

L'idea di applicare agli elementi della matrice ridotta la funzione logaritmica di base naturale, nasce dalle proprietà del set di dati. Nella figura 12 è rappresentata la funzione di densità di probabilità, *PDF*, riferita alla matrice ridotta normalizzata. La forma è fortemente asimmetrica con molti valori bassi. Per un intervallo di campionamento molto piccolo, si osserva che la risoluzione digitale è di $1 \times 10^{-6} \text{ mg} / \text{m}^3$.

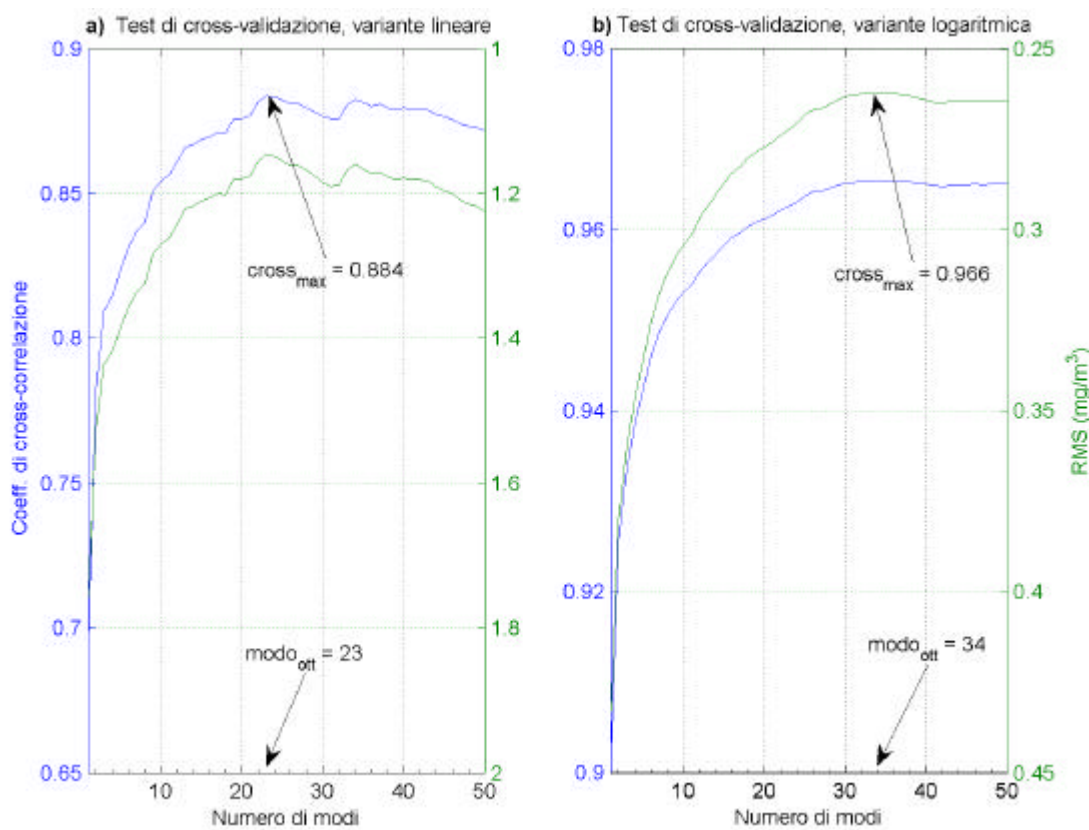


Figura 11: a) Coefficiente di cross-correlazione e RMS in funzione del numero di modi per la variante lineare e b) per la variante logaritmica.

Visto la *PDF* della matrice normalizzata si è pensato di utilizzare la funzione logaritmica per aumentare le differenze tra i valori piccoli. L'effetto di tale amplificazione è rappresentato dalla figura 13 che rappresenta la *PDF* della matrice logaritmica normalizzata (usata dalla variante logaritmica), confrontata con la *PDF* della matrice lineare normalizzata. Si osserva come la prima (in rosso) sia più simmetrica e con dati più dispersi, rispetto alla seconda (in blu). Si osserva che la *PDF* logaritmica, se confrontata con la lineare, dovrebbe interpolare meglio per valori compresi tra 0 mg/m^3 e 1 mg/m^3 , poiché la funzione logaritmica è più crescente di quella lineare per tale intervallo, mentre per i valori tra 1 mg/m^3 e 64 mg/m^3 , dovrebbe essere peggiore, poiché è meno crescente della lineare.

La varianza nei primi dieci modi, rispettivamente della matrice normalizzata e di quella logaritmica normalizzata è rappresentata nelle figure 14a e 14b. Si osserva che una maggiore quantità di varianza per i modi principali è espressa nel caso logaritmico che non in quello lineare.

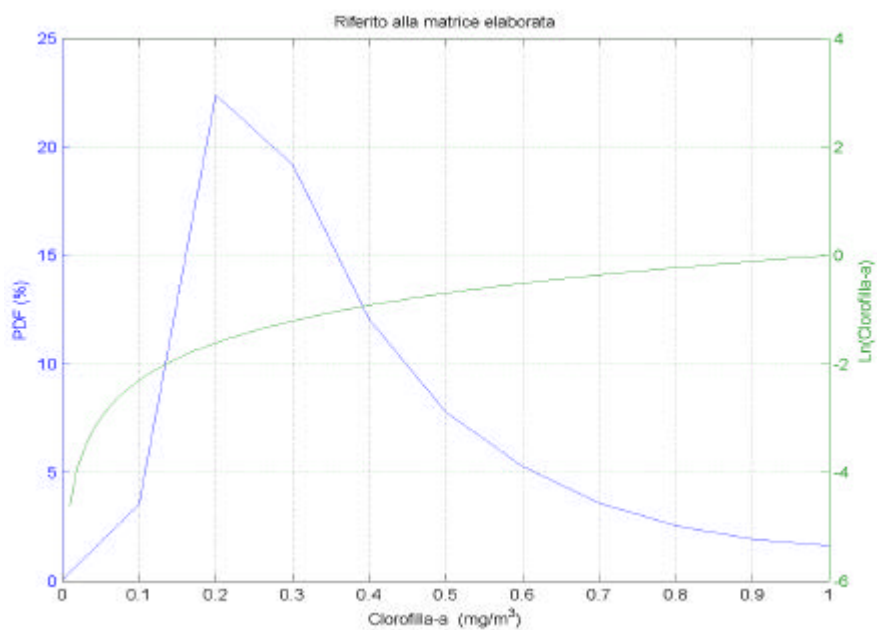


Figura 12: PDF della matrice ridotta e il logaritmo di base naturale in funzione dei valori della clorofilla sull'intervallo [0,1].

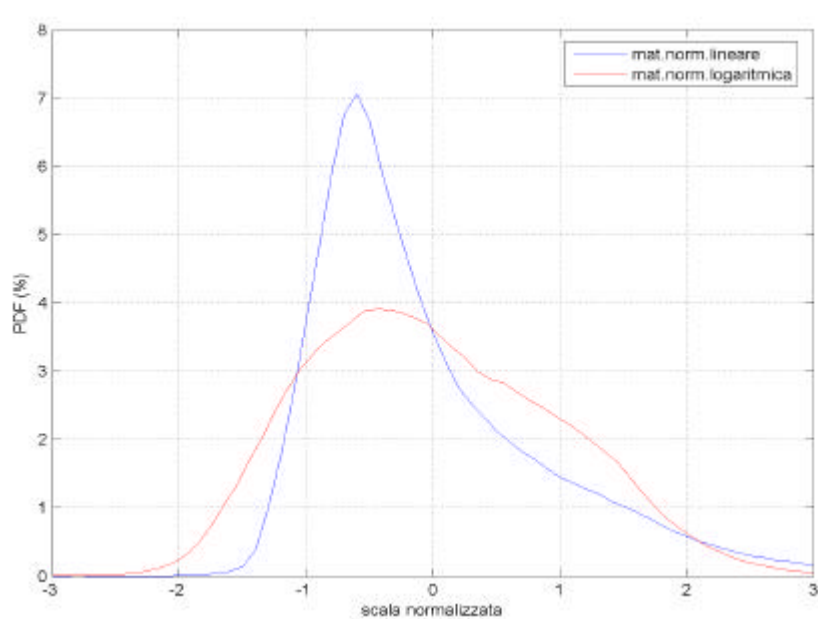


Figura 13: PDF della matrice normalizzata lineare (in blu) e quella riferita alla matrice normalizzata logaritmica (in rosso). I valori mancanti (uguali a zero) non sono considerati.

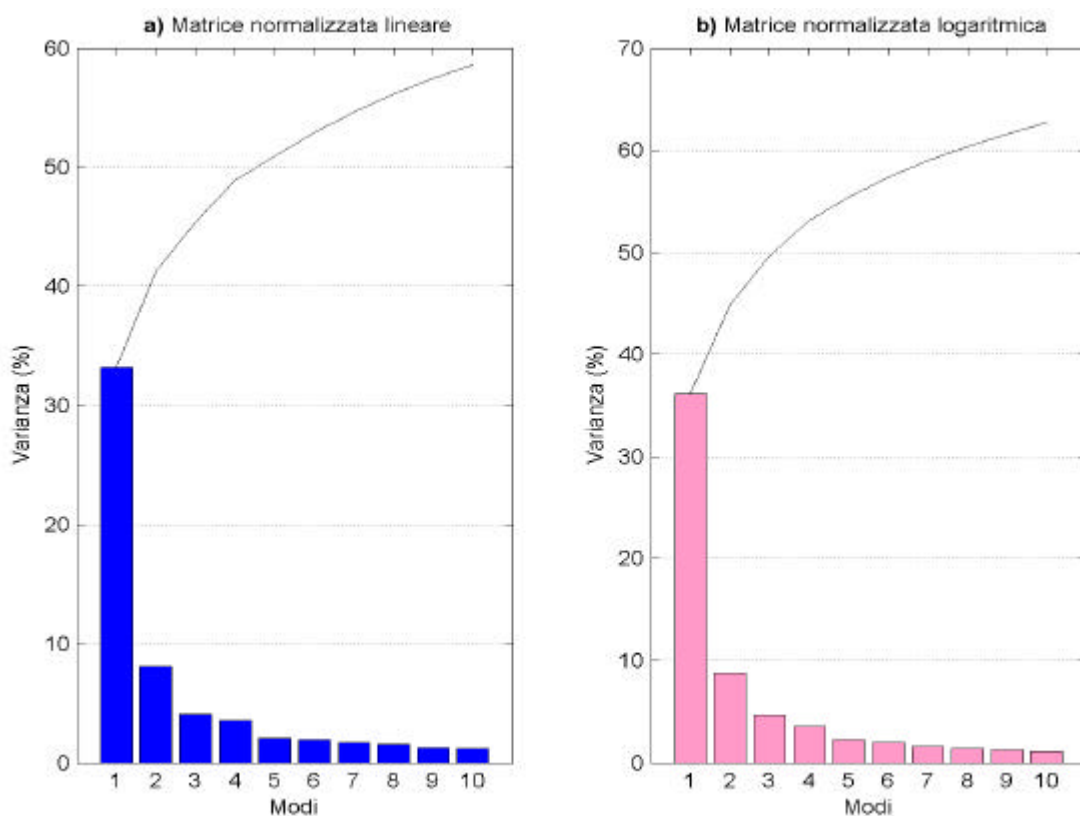


Figura 14: a) Varianza in funzione dei modi, riferita alla matrice normalizzata della variante lineare e b) di quella logaritmica.

3.3. Risultati ottenuti per mezzo del metodo EOF, variante lineare

Nonostante il fatto che la variante logaritmica sembra essere migliore di quella lineare, essa presenta delle difficoltà citate nel paragrafo precedente, non presenti nella seconda. Si presenteranno per cui i risultati di entrambe le varianti. Affinché la variante logaritmica possa essere considerata valida quanto quella lineare, il confronto deve far emergere un grado sensato di coerenza dei risultati.

In primo luogo sono presentate le caratteristiche della matrice interpolata con variante lineare, per mezzo delle misure statistiche semplici. Le figure 15a e 15b rappresentano rispettivamente i minimi e i massimi spaziali. Si osserva che la variante lineare crea dei valori negativi per il set di dati interpolati e dei valori maggiori del valore limite *LimVal*. In entrambi i casi, la regione problematica è quella lungo la costa occidentale, dove la varianza del segnale è più intensa. La percentuale dei valori negativi è del 1.35%, mentre per i valori maggiori a *LimVal* è del 0.01%. La media e la deviazione standard della matrice interpolata (figure 16a e 16b), se confrontate rispettivamente con

quelle della matrice ridotta (figura 4b e figura 5b), sono cambiate leggermente di valore. Inoltre il coefficiente d'asimmetria (skewness), rappresentato nella figura 17, se confrontato con quello della matrice ridotta (figura 6b), è più uniforme, con una concentrazione di valori alti nei pressi del fiume Po.

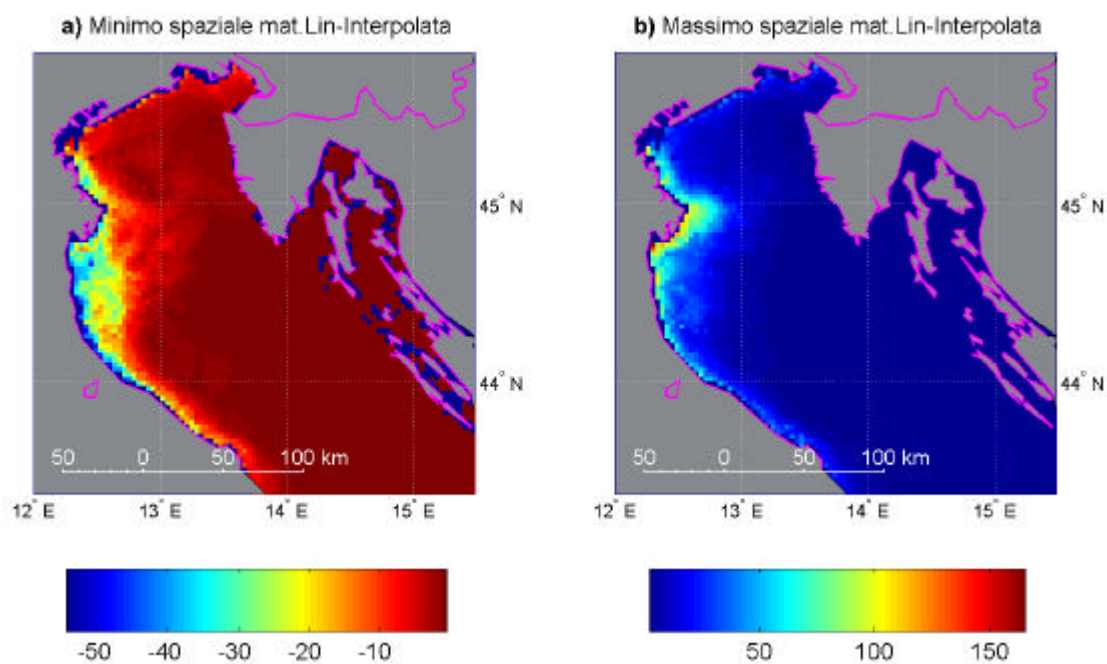


Figura 15: a) Minimi e b) massimi spaziali della matrice interpolata con variante lineare.

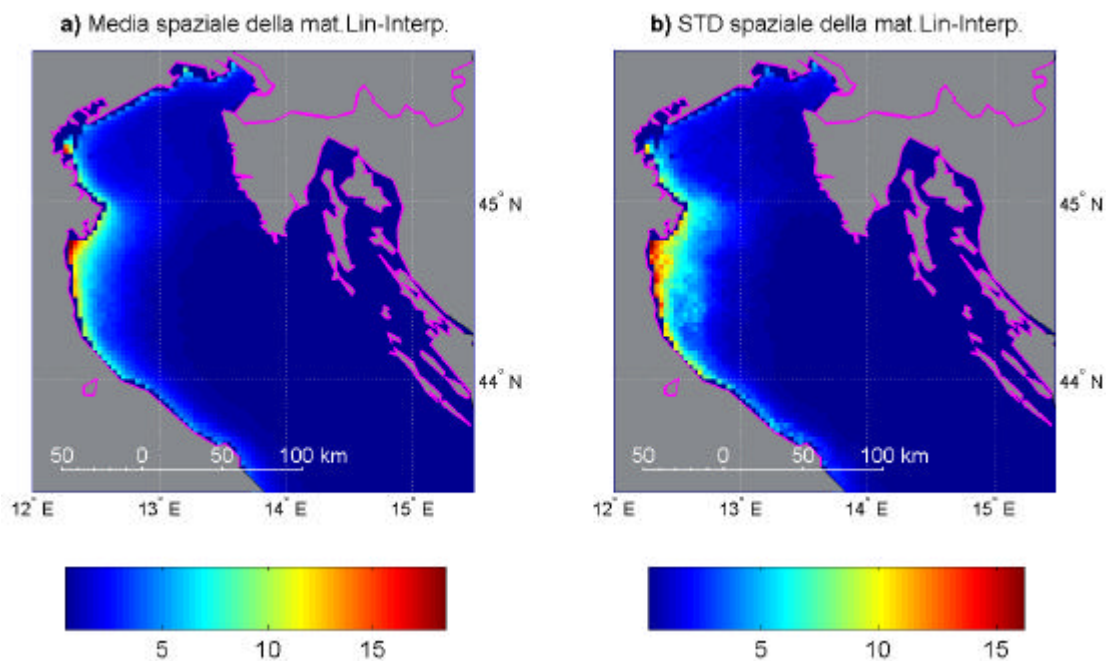


Figura 16: a) Media e b) deviazione standard spaziale riferite alla matrice interpolata con variante lineare.

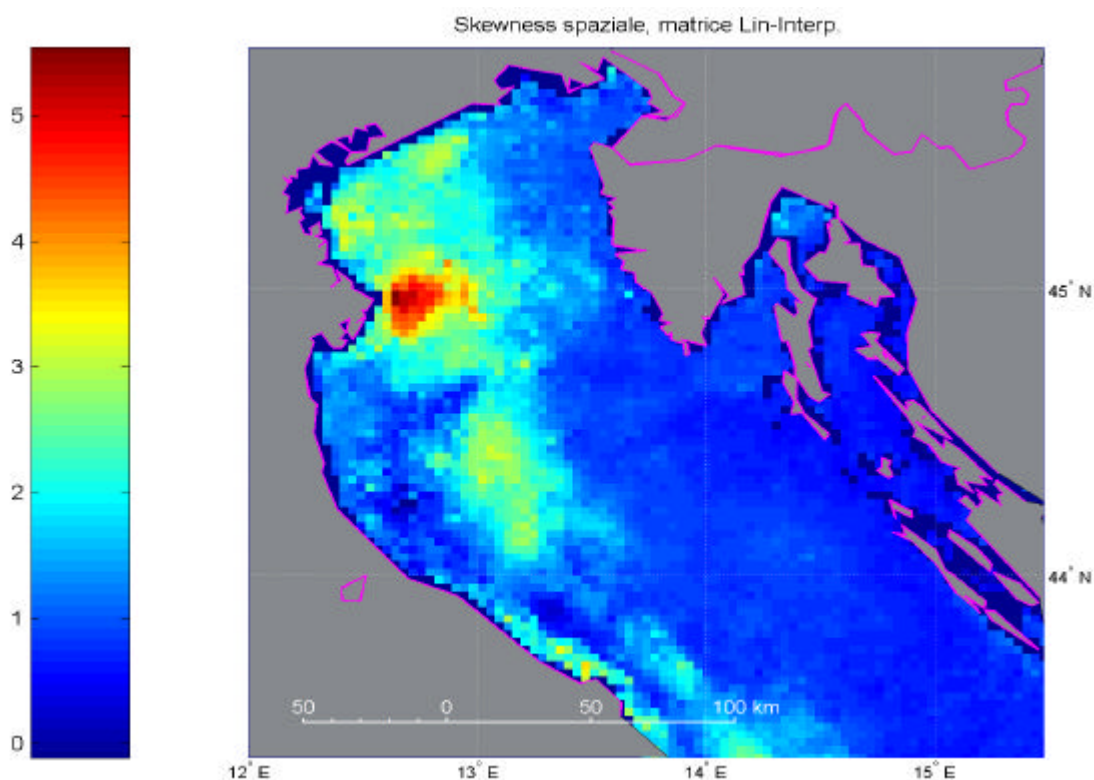


Figura 17: Coefficiente di asimmetria (skewness) spaziale della matrice interpolata con variante lineare.

In secondo luogo si analizzano i risultati per mezzo della decomposizione *SVD*, applicata alla matrice ridotta e a quell'interpolata di variante lineare. Per quanto riguarda la seconda i valori

negativi sono posti uguali a zero e quelli maggiori di $LimVal$ sono posti uguali a tale limite superiore, in modo da garantire un minimo grado d'affidabilità dei dati. Entrambe le matrici sono normalizzate, poiché tale operazione assicura che la varianza di un pixel non sia privilegiata nell'analisi (a tutti i pixel è data la stessa possibilità di contribuire alla varianza). Con riferimento alla figura 18, che rappresenta per ogni giornata interpolata la percentuale di valori maggiori di $LimVal$, si osserva una giornata con percentuale maggiore al 10%; essa crea disturbo nella decomposizione SVD e per tale ragione è tolta dal set di dati interpolati. A questo punto è possibile rappresentare i primi quattro modi delle EOF spaziali e delle ampiezze temporali. Si può notare inoltre, che a causa dell'operazione citata sopra, le varianze dei quattro modi, riportate nelle figure sottostanti, non coincidono con quelle della figura 14a.

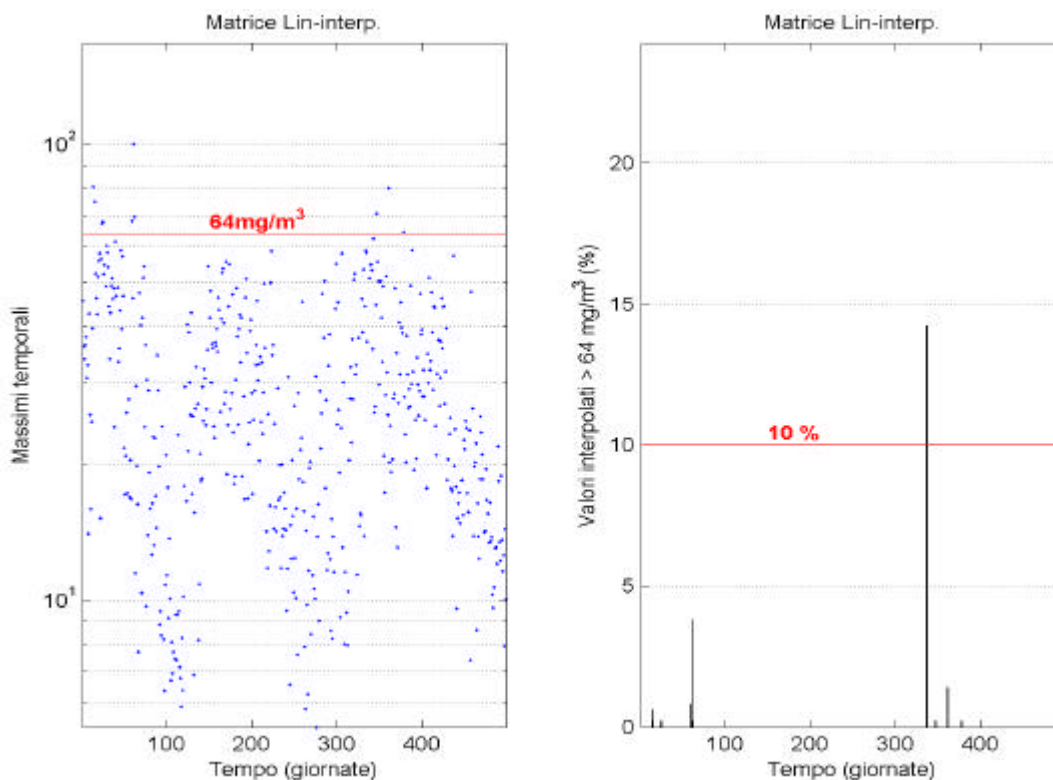


Figura 18: a) Massimi giornalieri della matrice interpolata di variante lineare. b) Percentuale temporale di valori eccedenti il limite superiore di $64 \text{ mg} / \text{m}^3$.

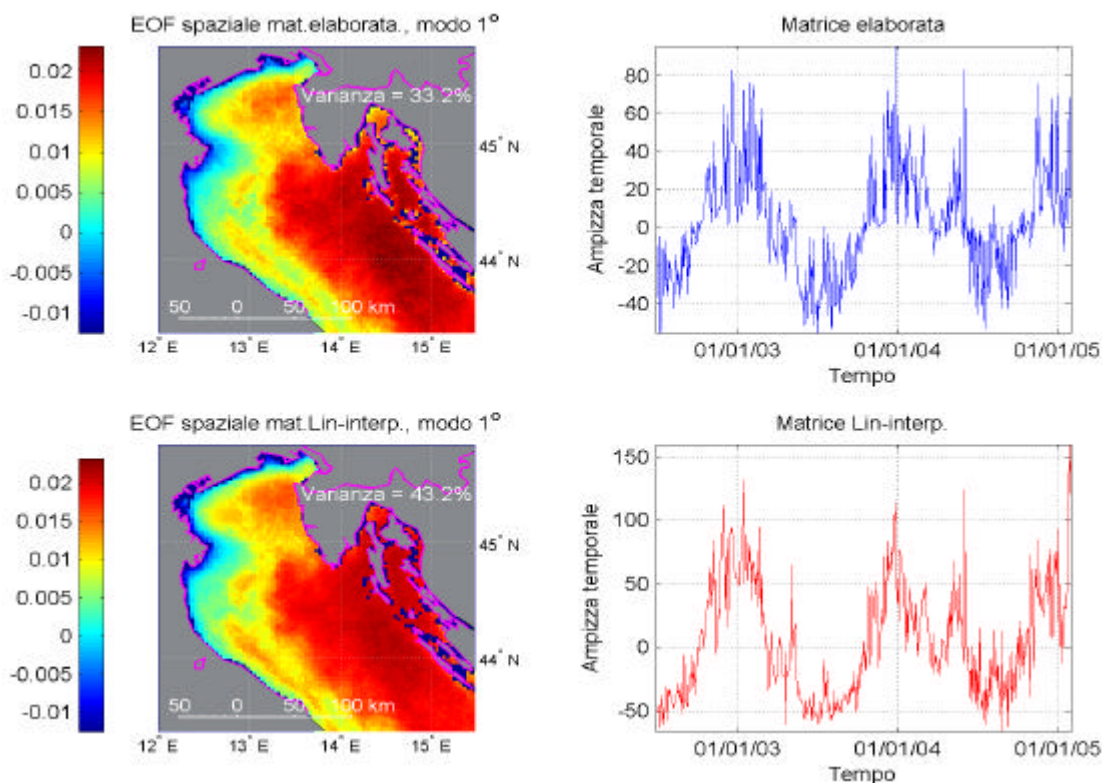


Figura 19: EOF spaziali e ampiezze temporali nel 1° modo per la matrice ridotta e quella interpolata utilizzando la variante lineare. Entrambe le matrici sono normalizzate.

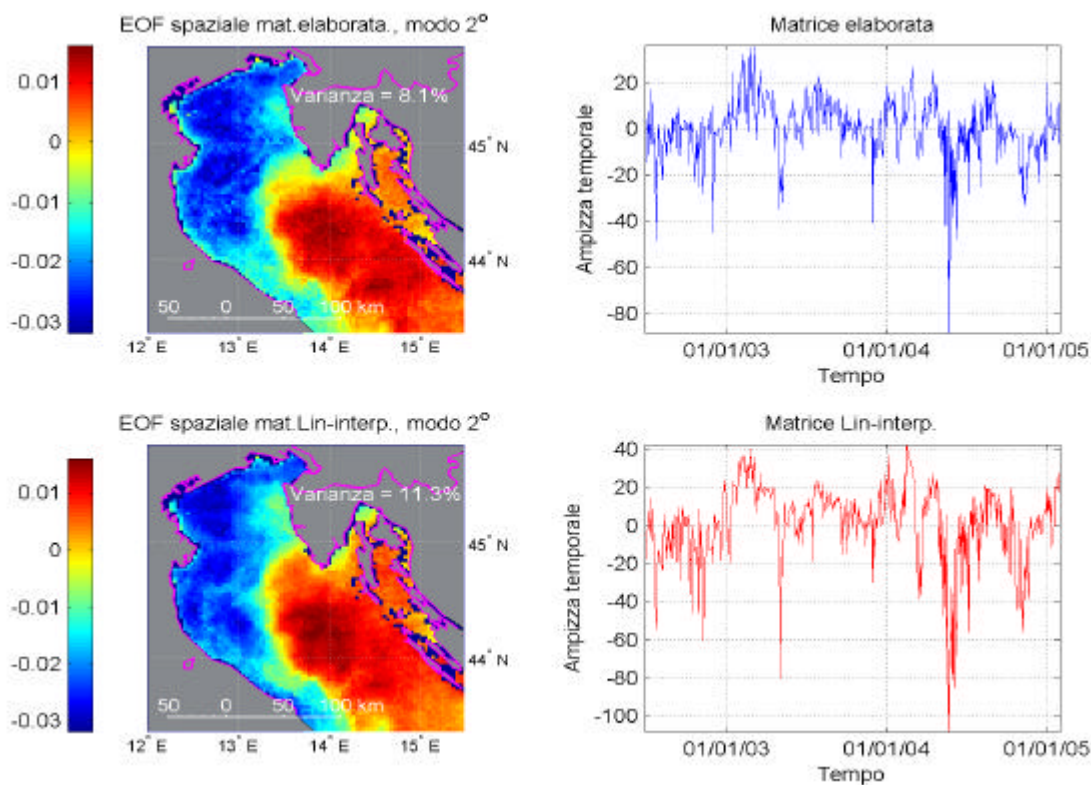


Figura 20: EOF spaziali e ampiezze temporali nel 2° modo per la matrice ridotta e quella interpolata utilizzando la variante lineare. Entrambe le matrici sono normalizzate.

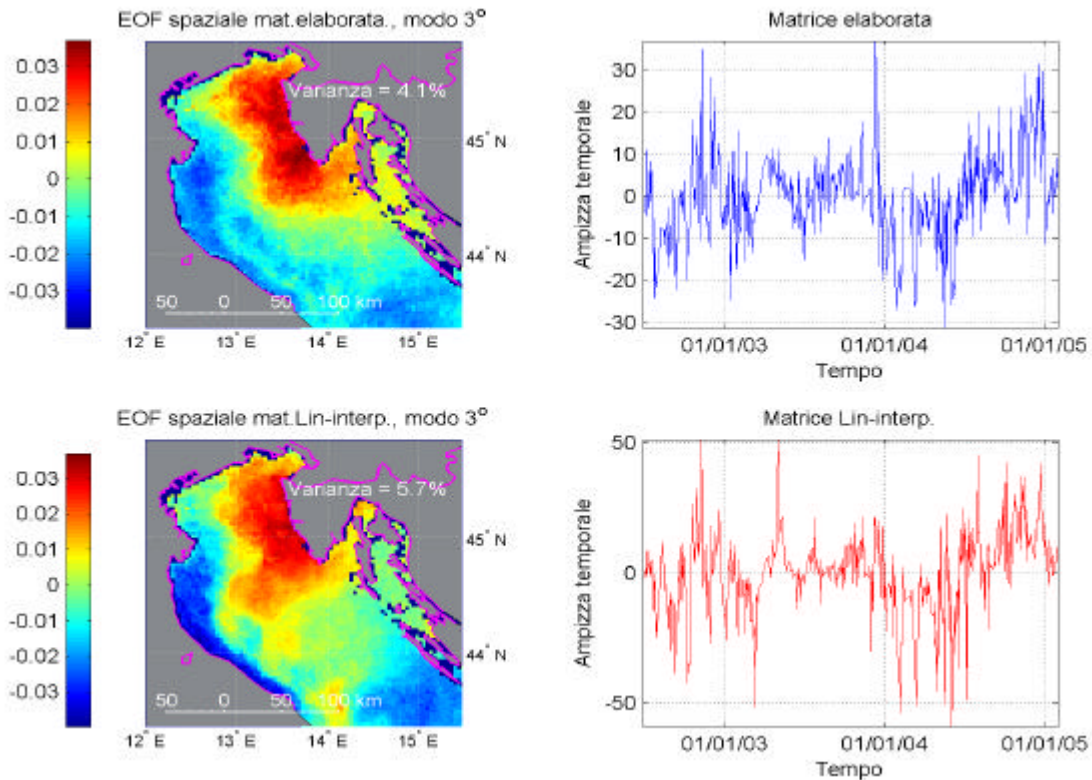


Figura 21: EOF spaziali e ampiezze temporali nel 3° modo per la matrice ridotta e quella interpolata utilizzando la variante lineare. Entrambi le matrici sono normalizzate.

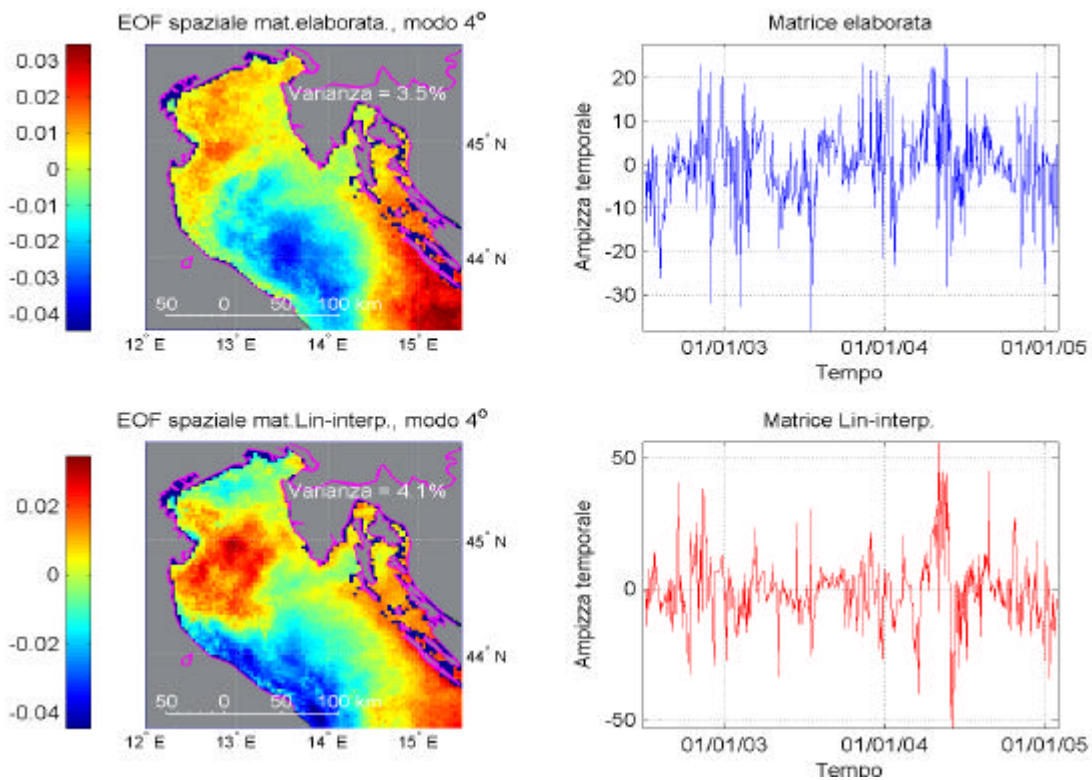


Figura 22: EOF spaziali e ampiezze temporali nel 4° modo per la matrice ridotta e quella interpolata utilizzando la variante lineare. Entrambe le matrici sono normalizzate.

3.4. Risultati ottenuti per mezzo del metodo EOF, variante logaritmica

La percentuale dei valori negativi, riferita alla matrice interpolata di variante logaritmica, è del 0%, mentre per i valori maggiori a $LimVal$ è del 0.07%. Il valore zero della prima percentuale è una prova che la variante logaritmica funziona particolarmente bene per valori minori a $1 \text{ mg}/\text{m}^3$: i valori piccoli sono stimati con grande precisione, infatti non si hanno valori negativi come nel caso della variante lineare. Il valore della seconda percentuale, nonostante essa sia apparentemente buona, risulta contenere valori molto grandi, dell'ordine del $1 \times 10^2 \div 1 \times 10^6$, come indicato dalla figura 23a che rappresenta i massimi giornalieri della matrice interpolata. Dalla figura 23b, che riporta la percentuale dei valori maggiori di $LimVal$ per ogni giornata, si osserva che ci sono diverse giornate interpolate con troppi valori sopra il limite superiore (sette giornate).

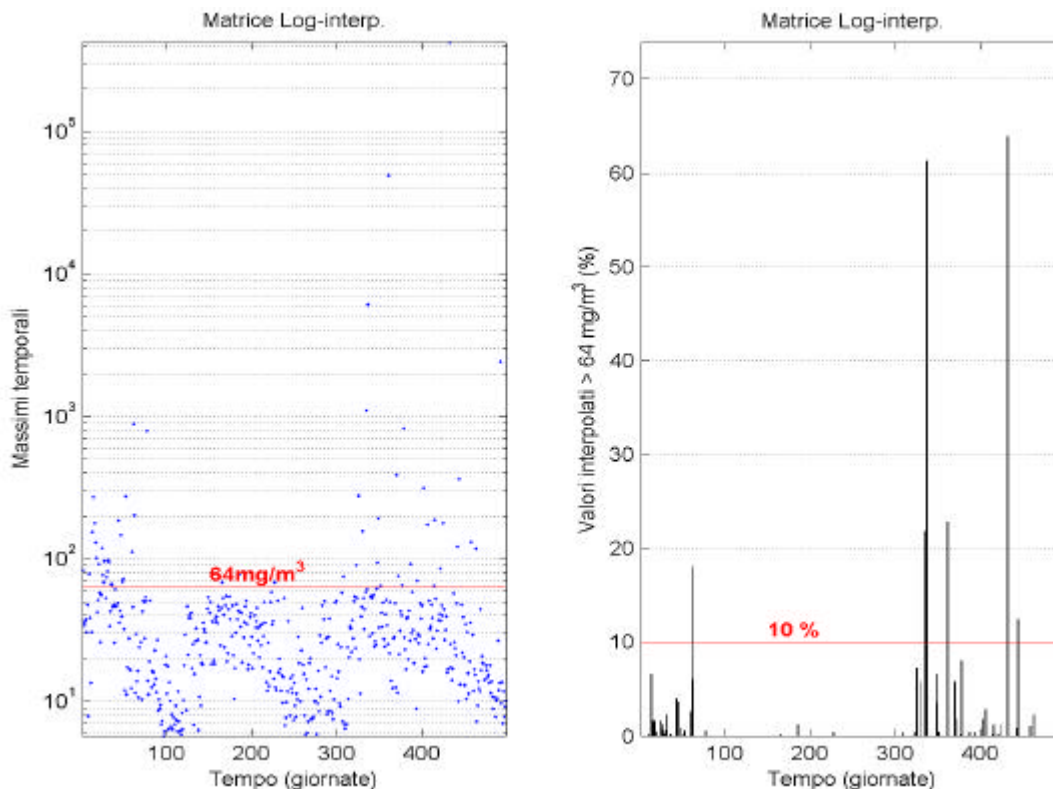


Figura 23: a) Massimi giornalieri della matrice interpolata di variante logaritmica. b) Percentuale temporale di valori eccedenti il limite superiore di $64 \text{ mg}/\text{m}^3$.

Si può ottenere una stima degli indici statistici se si riduce la matrice interpolata come nel caso visto nel paragrafo precedente, per la decomposizione SVD. Dalle figure 24a e 24b, che rappresentano i minimi e i massimi temporali, si osserva che la variante logaritmica crea valori maggiori di *LimVal* vicino alla varianza alta del segnale, come nel caso della variante lineare. La media e la deviazione standard della matrice interpolata (figure 25a e 25b) sono confrontabili con quelle della matrice interpolata di variante lineare (figure 16a e 16b): infatti, si ha un leggero cambiamento di valori presso l'estuario del Po. Il coefficiente d'asimmetria spaziale, rappresentato nella figure 26, sembra privo di significato per quanto riguarda i valori alti (ai quali è particolarmente sensibile).

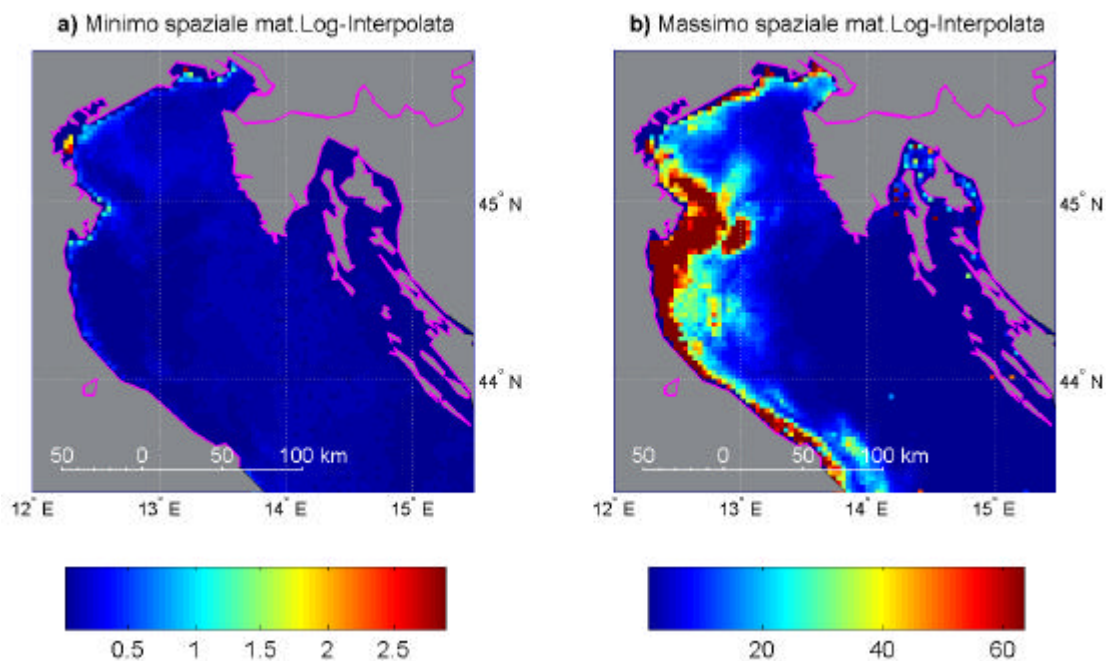


Figura 24: a) Minimi e b) massimi giornalieri della matrice interpolata con variante logaritmica.

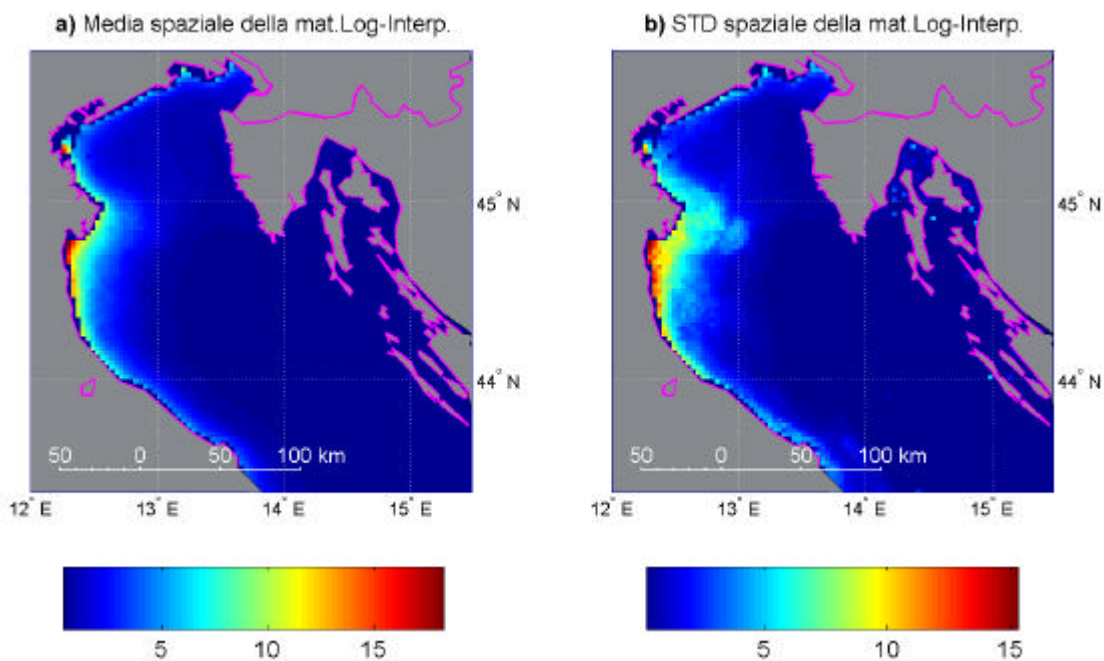


Figura 25: a) Media e b) deviazione standard, riferite alla matrice interpolata con variante logaritmica.

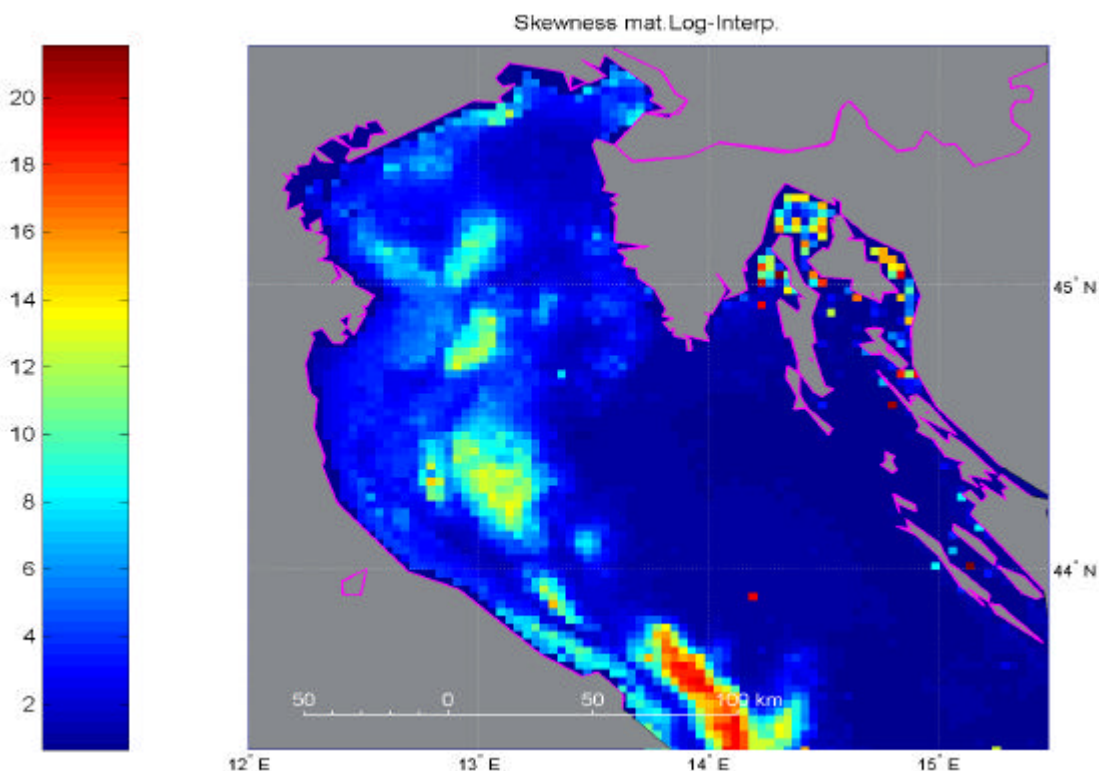


Figura 26: Coefficiente di asimmetria spaziale della matrice interpolata con variante logaritmica.

Si passa ora alla presentazione dei risultati ottenuti per mezzo della decomposizione *SVD*, svolta in maniera analoga a quella del paragrafo precedente (la matrice interpolata è ridotta).

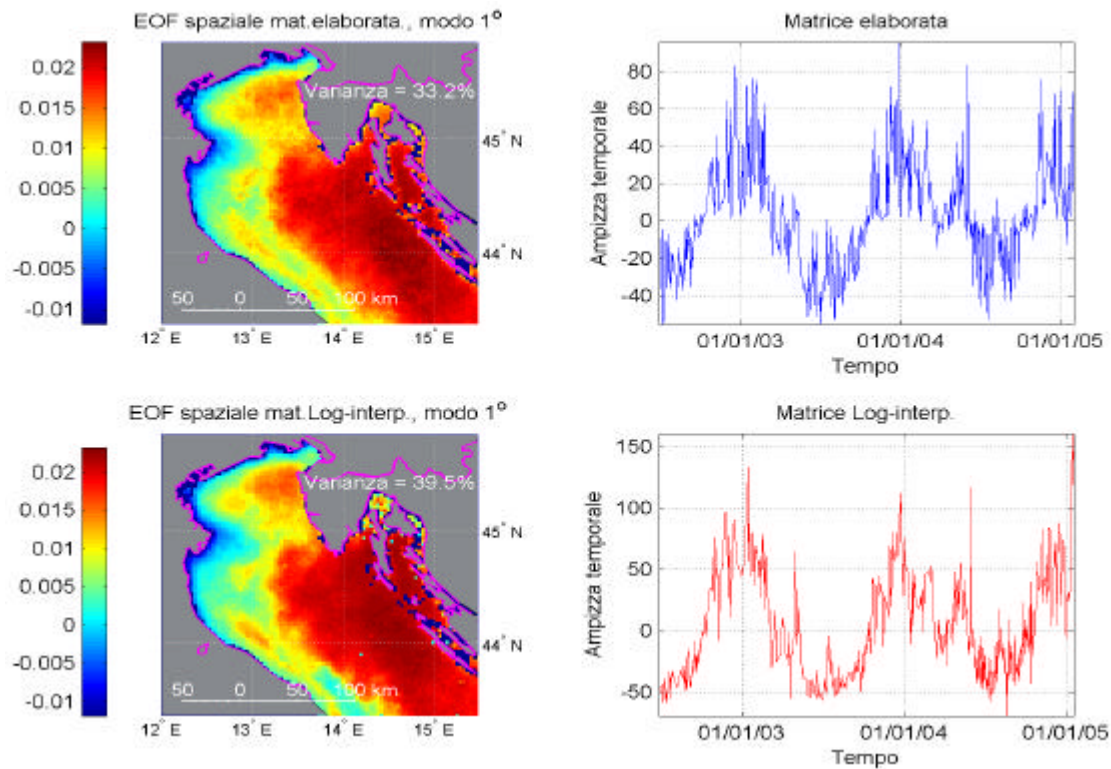


Figura 27: EOF spaziali e ampiezze temporali nel 1° modo per la matrice ridotta e quella interpolata utilizzando la variante logaritmica. Entrambe le matrici sono normalizzate.

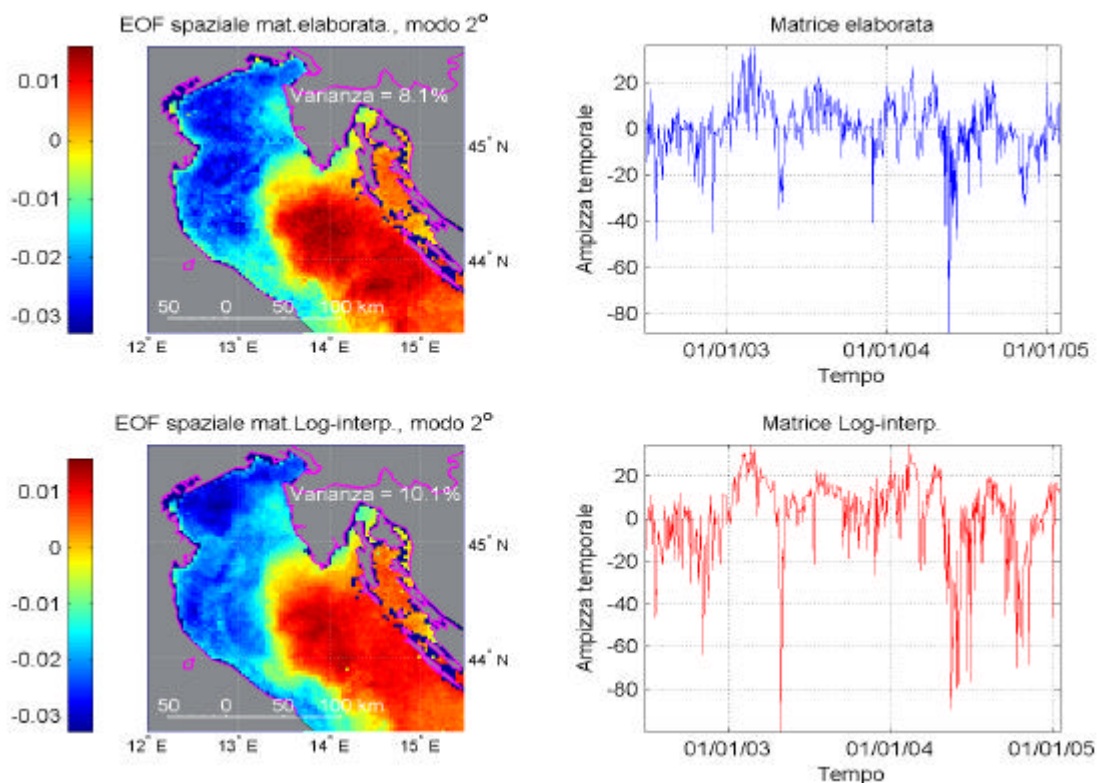


Figura 28: EOF spaziali e ampiezze temporali nel 2° modo per la matrice ridotta e quella interpolata utilizzando la variante logaritmica. Entrambe le matrici sono normalizzate.

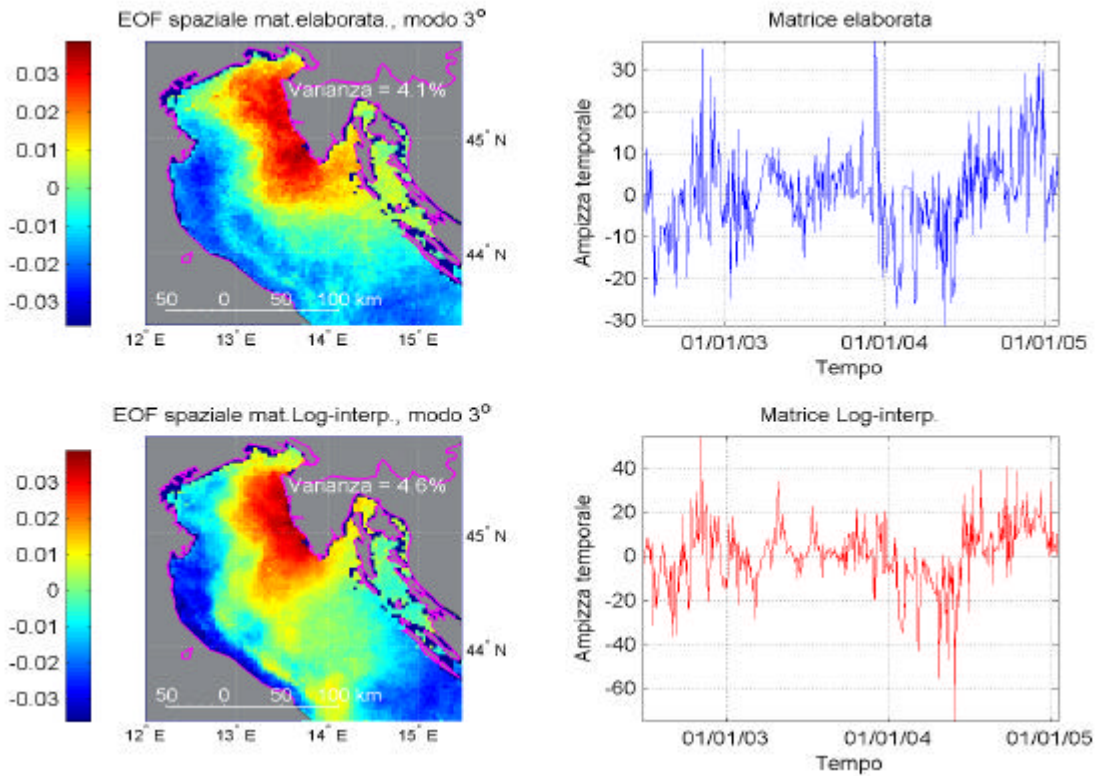


Figura 29: EOF spaziali e ampiezze temporali nel 3° modo per la matrice ridotta e quella interpolata utilizzando la variante logaritmica. Entrambe le matrici sono normalizzate.

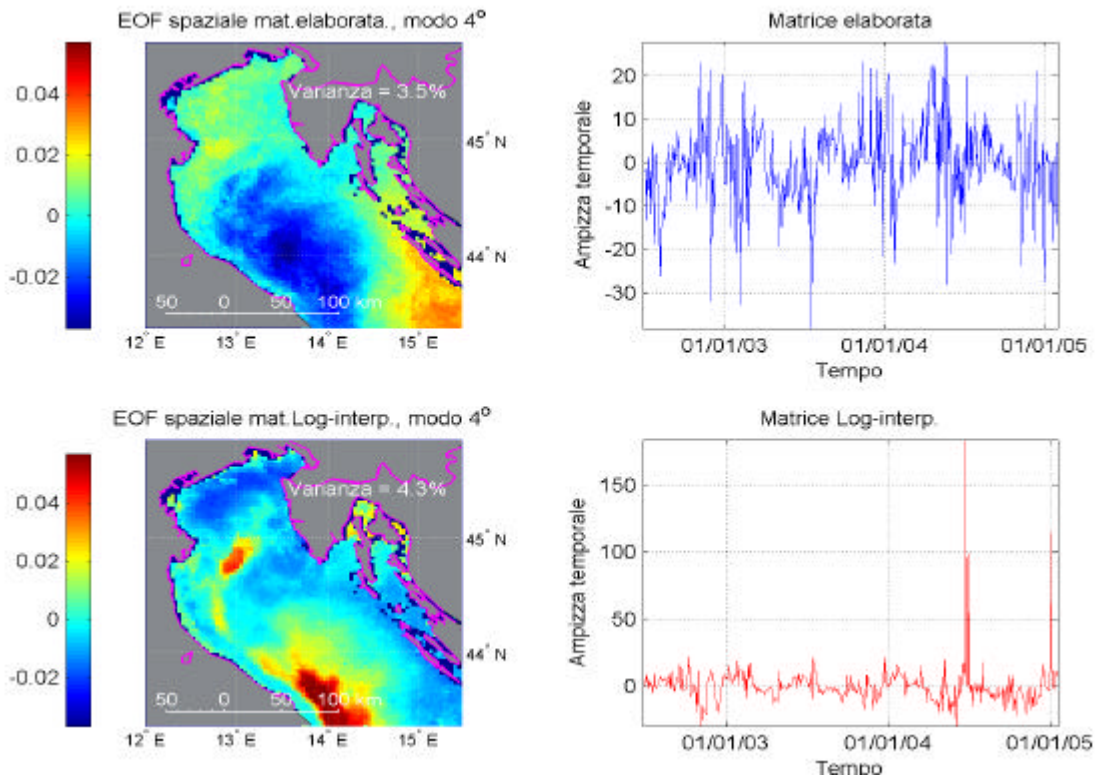


Figura 30: EOF spaziali e ampiezze temporali nel 4° modo per la matrice ridotta e quella interpolata utilizzando la variante logaritmica. Entrambe le matrici sono normalizzate.



4. Conclusioni

Nella presente relazione si sono illustrati i risultati del metodo di interpolazione con le EOF applicato ai dati MODIS di clorofilla superficiale nell'alto e medio Adriatico fra luglio 2002 e febbraio 2005, sia con la variante lineare che con quella logaritmica. I risultati ottenuti, fatta eccezione per alcuni valori negativi ottenuti con il metodo lineare, sono molto simili, ciò quindi non sembra influire pesantemente sul metodo. Si suggerisce quindi di utilizzare in ulteriori indagini la variante lineare per semplicità sia di calcolo che interpretativa.

5. Referenze

Beckers, J.-M. e M. Rixen (2003) EOF Calculations and Data Filling from Incomplete Oceanographic Datasets, *Journal of Atmospheric and Oceanic Technology*, 20, 1839-1856.

Južnic-Zonta, Ž., E. Mauri e P.-M. Poulain (2004) Interpolazione di dati attraverso metodo EOF: applicazione a immagini MODIS nel Golfo di Trieste, Rel.51/2004/OGA/14, 46 pp.