

UNIVERSITÀ DEGLI STUDI DELLA TUSCIA DI VITERBO

DIPARTIMENTO DI SCIENZE AGRARIE E FORESTALI

Corso di Dottorato di Ricerca in

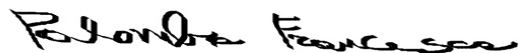
ECOSISTEMI E SISTEMI PRODUTTIVI - XXIX Ciclo.

PARBLEU: un modello di qualità ecologica empirico con valutazione dell'incertezza

SSD: AGR/08

Tesi di dottorato di:

Dott. Francesca palomba



Coordinatore del corso

Prof. Alessandro Sorrentino

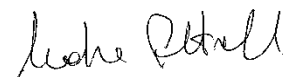
Firma



Tutore

Prof. Ing. Andrea Petroselli

Firma



13/07/2017

A mio padre...

*La ricerca della verità è l'occupazione
più nobile dell'uomo; la sua
pubblicazione è un
dovere*

RINGRAZIAMENTI

Al termine di questo lavoro desidero ringraziare coloro che mi hanno sostenuta ed incoraggiata nel corso di questa ricerca e che hanno contribuito in vario modo alla mia crescita umana e intellettuale, in particolare i carissimi colleghi e amici dell'Autorità di Bacino del Fiume Tevere: Gianni Colatosti, Filippo Marasco, Leonardo Gatta, Emanuele Sillato e Paola Schiavella.

Ringrazio infinitamente il Dirigente Tecnico dell'Autorità di Bacino del Fiume Tevere Remo Pelillo per la linea guida che mi ha fornito nell'elaborazione del modello, per le nozioni di base relative all'inferenza statistica, per la disponibilità e il tempo dedicatomi nell'approfondimento dei temi fondamentali relativi alla modellistica.

Ringrazio inoltre i Professori Andrea Petroselli e Salvatore Grimaldi per avermi dato l'opportunità di realizzare la presente ricerca e di scrivere un articolo scientifico.

Un ringraziamento è inoltre rivolto al professore Leonardo Varvaro, al Segretario Generale dell'Autorità di Bacino del fiume Tevere Dott. Ing. Giorgio Cesari e al Coordinatore del corso di Dottorato di ricerca in EcoSistemi e Sistemi Produttivi, Professor Alessandro Sorrentino.

1. Introduzione

1.1. Obiettivi

1.2. La scelta del modello

1.2.1. Criteri generali di scelta

1.2.2. Il modello fisicamente basato

1.2.3. Considerazioni per la scelta: il modello empirico

1.2.4. La quantificazione dell'incertezza

2. Area di studio

3. Materiali e metodi

3.1. Organizzazione del modello

3.1.1. I dati in ingresso

3.1.1.1. Macroinvertebrati bentonici

3.1.1.2. Macrofite

3.1.1.3. Diatomee

3.1.1.4. Fauna ittica

3.1.2. La costruzione del campione statistico

3.1.3. La scelta della variabile dipendente e dei regressori: la specificazione del modello

3.1.3.1. La matrice di correlazione

3.1.3.2. Scatterplots

3.1.3.3. Regressione lineare semplice

3.1.3.4. Analisi di regressione lineare multipla

3.1.3.5. Il secondo approccio per la specificazione: “partial residual plots”

3.1.3.6. Metodo backward

3.1.4. Stima dei parametri con verifica delle assunzioni di base della regressione

3.1.4.1. Omoschedasticità

3.1.4.2. Normalità della distribuzione degli errori

3.1.4.3. Collinearità

3.1.5. Calcolo incertezza attraverso la stima degli intervalli di confidenza

3.1.6. Risoluzione e rappresentazione delle regressioni

4. Risultati

4.1. Analisi preliminare

4.1.1. Matrice di correlazione

4.1.2. Scatter-plots

4.2. Specificazione del modello

4.2.1. Regressione lineare semplice

4.2.1.1. RQE_IBMR

4.2.1.2. RQE_ICMi

4.2.1.3. RQE_STAR_ICMi

4.2.1.4. ISECI

4.2.2. Regressione lineare multipla

4.2.2.1. RQE_IBMR

4.2.2.1.1. Test di significatività

4.2.2.1.2. Grafici dei residui

4.2.2.1.3. Verifica della normalità della distribuzione degli errori

4.2.2.1.4. Confronto tra modelli

4.2.2.2. RQE_ICMi

4.2.2.2.1. Test di significatività

4.2.2.2.2. Grafici dei residui

4.2.2.2.3. Verifica della normalità della distribuzione degli errori

4.2.2.2.4. Confronto tra modelli

4.2.2.3. RQE_STAR_ICMi

4.2.2.3.1. Test di significatività

4.2.2.3.2. Grafici dei residui

4.2.2.3.3. Verifica della normalità della distribuzione degli errori

4.2.2.3.4. Confronto tra modelli

4.2.2.4. ISECI

4.2.2.4.1. Test di significatività

4.2.2.4.2. Grafici dei residui

4.2.2.4.3. Verifica della normalità della distribuzione degli errori

4.2.2.4.4. Confronto tra modelli

4.1.3. Stima dei parametri con verifica assunzioni del sotto-modello finale

4.1.3.1. I test di significatività

4.1.3.2. Verifica dell'assunzione di normalità ed omoschedasticità

5. Discussioni

5.1. Le violazioni delle assunzioni nella regressione del RQE-ICMi

5.2. L'interpretazione delle doppie frecce

5.3. L'effetto di N-NH₄ ed N-NO₃ sugli indici biotici

5.4. L'effetto del PT sugli indici biotici

5.5. L'effetto dell'OD sugli indici biotici

5.6. L'effetto della variabile IDRO-ECOREGIONE sugli indici biotici

5.7. L'utilità del sotto-modello

6. Conclusioni

BIBLIOGRAFIA

1. Introduzione

Il ‘Dottorato in ecosistemi e sistemi Produttivi’ mira all’individuazione di strumenti teorici e metodologici che consentano di effettuare analisi avanzate relative ai sistemi di gestione sostenibile delle risorse naturali e ai sistemi di trasformazione e impiego delle medesime, valutandone le implicazioni nel contesto ambientale e socioeconomico. L’attività di ricerca svolta presso l’Autorità di bacino del Fiume Tevere (AbT) risponde all’obiettivo descritto attraverso l’elaborazione ed il processamento di un modello che esplicita le relazioni esistenti tra le attività socio-economiche e le risorse naturali, facendo riferimento, in particolare, agli elementi di qualità biologica che costituiscono l’elemento base della qualità ambientale dei corpi idrici superficiali.

Il modello risponde all’esigenza dell’AbT di definire misure per ottenere un prefissato obiettivo di qualità ambientale attraverso il miglioramento dello stato di qualità dei corpi idrici superficiali. Tale modello è uno strumento di gestione e di supporto alle decisioni nel processo di pianificazione inerente alla gestione della risorsa idrica per l’implementazione della Direttiva n. 2000/60/CE (WFD) così come recepita con il D.Lgs. 3 aprile 2006, n. 152.

Il corpo idrico (superficiale) identifica il luogo delle interazioni delle diverse componenti dell’ecosistema direttamente legato alle varie categorie di acque presenti all’interno di un distretto idrografico: in questo lavoro ci occuperemo del sub-distretto dei bacini abruzzesi all’interno del più vasto distretto dell’Appennino Centrale.

In termini più generali, i corpi idrici (e nello specifico quelli superficiali) contribuiscono alla formazione dei caratteri del “paesaggio” così come definito nel D.Lgs. 22 gennaio 2004, n. 42. In questo contesto il modello, anche se non direttamente, guarda alla capacità di autodepurazione del corso d’acqua avendo come obiettivo strumentale quello di conservarne la massima potenzialità nella produzione dei cosiddetti servizi ecosistemici.

L’interazione tra le attività socio-economiche ed il sistema naturale può essere studiata attraverso semplici relazioni causali; questa assunzione è all’origine dello schema noto come DPSIR (*Driving forces – Pressures – State – Impact - Response*) adottato dall’Agenzia Ambientale Europea (1999) per proporre con esso una struttura di riferimento generale e un approccio integrato nei processi di *reporting* sullo stato dell’ambiente. Tale approccio integrato permette di rappresentare l’insieme degli elementi e delle relazioni che caratterizzano un qualsiasi fenomeno ambientale, attraverso l’intreccio delle dinamiche

naturali con quelle delle politiche esercitate su di esse. Lo schema DPSIR è fondato, quindi su una struttura di relazioni causali che lega nella WFD tra loro i seguenti elementi:

- 1) determinante (*driver*): attività antropica in grado di produrre un effetto sull'ambiente;
- 2) pressione: effetto diretto dei determinanti sull'ecosistema del corpo idrico (e.g.: un effetto che causa un cambiamento nel flusso e/o nella chimica delle acque);
- 3) stato: la condizione nella quale vengono a trovarsi le componenti biotiche e abiotiche del corpo idrico derivante dalle pressioni esercitate su di esso;
- 4) impatti: l'effetto ambientale della pressione (e.g.: un aumento della concentrazione dei nutrienti nelle acque);
- 5) risposte: le misure prese per migliorare lo stato del corpo idrico (e.g.: ridurre il prelievo di acqua dai corpi idrici, limitare le fonti di scarico puntuale, ecc).

Il modello DPSIR prevede l'utilizzo di indicatori, valutati attraverso grandezze osservabili e misurabili (programma di monitoraggio), dal contenuto informativo elevato ed articolato: la loro funzione principale è la rappresentazione sintetica dei problemi indagati in modo da conservare in fase di analisi il necessario contenuto informativo.

La Direttiva n. 2000/60 CE, introducendo un approccio innovativo nell'amministrazione e gestione delle risorse naturali, si propone di raggiungere gli obiettivi generali delineati nell'articolo 4 all'interno del quale convivono:

- la necessità di perseguire in un determinato orizzonte temporale per tutti i corpi idrici almeno lo stato di qualità “buono” (§ 1 dell'art. 4);
- la possibilità che una parte di questi corpi idrici possa essere esentata dall'obiettivo “buono” per ragioni tecniche e socio-economiche (§§ 4, 5, 6 e 7 dell'art. 4).

La Direttiva stabilisce che i singoli Stati Membri affrontino la tutela delle acque a livello di “bacino idrografico” che rappresenta l'unità territoriale di riferimento organizzata all'interno del “distretto idrografico”, costituito da uno o più bacini idrografici limitrofi e dalle rispettive acque superficiali (fiumi, laghi, acque di transizione e marino-costiere) e sotterranee.

In ciascun distretto idrografico gli Stati membri devono adoperarsi affinché vengano effettuati:

- un'analisi delle caratteristiche del distretto;
- un esame dell'impatto provocato dalle attività umane sullo stato delle acque superficiali e sotterranee;
- un'analisi economica dell'utilizzo idrico.

Nel presente lavoro ci occuperemo del secondo aspetto con particolare riferimento alle acque superficiali tenendo in considerazione i vincoli ed i condizionamenti imposti dal terzo.

Relativamente ad ogni distretto, deve essere predisposto un “programma di misure” che tenendo conto delle analisi effettuate consenta di raggiungere gli obiettivi ambientali fissati dalla Direttiva riducendo le pressioni in relazione ai loro effetti diretti (carichi immessi nel reticolo idrografico e negli acquiferi). I programmi di misure sono integrati nei Piani di Gestione che ogni Autorità di Distretto deve predisporre/aggiornare ogni sei anni; i primi rappresentano pertanto lo strumento di attuazione per il conseguimento degli obiettivi fissati nei secondi.

Il processo di formazione dei piani e dei relativi programmi è iterativo e interattivo comprendendo un determinato numero di *step* eseguiti in un prefissato periodo temporale. Poiché l'approccio include di solito tre stadi principali (valutazione della condizione attuale e degli scenari futuri caratterizzati dalle pressioni ridotte; definizione degli obiettivi di sostenibilità ambientale ed economica; sviluppo di programmi di misure idonee a conseguire gli obiettivi) il modello rappresenta il necessario strumento di supporto alle decisioni del soggetto istituzionale.

Occorre tuttavia considerare che l'unità territoriale di riferimento (bacino e distretto) rappresenta anch'essa una necessaria semplificazione in quanto sia i processi socio-economici sia le dinamiche naturali assumono “ambiti” di riferimento non definibili “una volta per tutte”.

In effetti, alla intrinseca “complessità” dei sistemi “naturali” si somma un'ulteriore complessità che inducendo una forte “stocasticità” a livello di bacino (e di distretto) ha consigliato, come si vedrà oltre, di ricorrere non ad un modello fisicamente basato ma ad un modello empirico basato sulla cosiddetta “analisi a scatola chiusa” (*black-box analysis*).

Un ulteriore aspetto, assolutamente non secondario, è costituito dalla forma dei processi decisionali che accompagnano la formazione e l'attuazione del Piano di Gestione e del relativo programma delle misure.

Il progressivo passaggio delle competenze sul demanio idrico dal Governo centrale ai Governi regionali (compiutamente determinata dal D. Lgs. n. 112/1998) e la compresenza di più Regioni su diversi bacini idrografici facenti parte del Distretto hanno determinato la necessità di definire in accordo con la Commissione Europea il rapporto tra il Piano di Gestione del Distretto dell'Appennino Centrale (PGDAC) e i diversi Piani Regionali di Tutela delle Acque (PTA). Il PGDAC (così come i piani di gestione degli altri distretti) si configura come un piano strategico che definisce misure e obiettivi a scala di distretto che i PTA provvedono a specializzare territorialmente a livello di corpo idrico (o di gruppi di corpi idrici) in ragione delle criticità locali.

Questo doppio livello di definizione e di risoluzione richiede strumenti differenziati e proporzionalmente adeguati e motiva, unitamente alle considerazioni precedenti, la scelta del tipo di modello, illustrata nei successivi paragrafi, ma soprattutto accende i riflettori su una lettura e su un approccio semplificati che vedono nello schema DPSIR un riferimento immediatamente operativo.

1.1. Obiettivi

La WFD impone agli Stati membri di redigere per ciascun distretto idrografico, il piano di gestione e il programma di misure, che forniscono il contesto per coordinare l'attuazione degli obiettivi chiave della Direttiva. Il programma di misure deve soddisfare gli obiettivi ambientali e gli interessi di tutti gli stakeholder coinvolti; diventa essenziale trovare un collegamento tra le misure, l'impatto delle attività umane e gli obiettivi ambientali. In questa parte del processo di pianificazione, la descrizione di ciò che potrebbe accadere è spesso fornita attraverso giudizi di esperti, ipotesi ragionevoli (ad es.: estrapolazioni) o, il più delle volte, prendendo in considerazione diversi scenari basati su simulazioni modellistiche. In risposta alle disposizioni della WFD, l'AbT ha sviluppato un sistema modellistico che contiene tre moduli principali: il primo relativo alla risorsa idrica naturale, il secondo relativo alla definizione di scenari per la gestione delle risorse idriche, il terzo relativo all'analisi economica e ambientale. Il cuore del sistema modellistico è il secondo modulo e il suo nome è PARBLEU (acronimo di Pressures/impacts/responses Assessment into River Basins Leading to Environmental Uses).

E' noto che la verifica del conseguimento dello stato di qualità "buono", per i corpi idrici superficiali, previsto dalla WFD, è effettuata attraverso la determinazione dello "stato

ecologico” e dello “stato chimico”. Per “stato ecologico” si intende l'espressione della qualità della struttura e del funzionamento degli ecosistemi acquatici associati alle acque superficiali, supportati da specifici indicatori chimico-fisici e idromorfologici; lo stato chimico è invece valutato sulla base del confronto tra le concentrazioni degli inquinanti monitorati e gli standard di qualità previsti dalla normativa. Poiché lo stato chimico non rappresentava per l'area oggetto di studio un fattore di rischio, solo alcuni degli elementi usati per la definizione dello stato ecologico sono stati usati nello sviluppo del modello.

In Italia, per la valutazione dello stato ecologico, i seguenti indici sono stati determinati per i quattro elementi di qualità biologica (EQB, macrofite, diatomee, macroinvertebrati e fauna ittica): STAR_ICMi (macroinvertebrati), ICMi (diatomee bentoniche), IBMR (macrofite), and ISECI (fauna ittica). La normativa italiana ha individuato il LIMeco come indice della qualità chimico-fisica, con il fine di sostenere gli elementi biologici. Il LIMeco si concentra sullo stato trofico delle acque dolci, considerando nutrienti (azoto ammoniacale (N-NH_4), azoto nitrico (N-NO_3), fosforo totale (PT) e ossigeno disciolto (OD)) (vedi Decreto Ministeriale Ambiente, 8/11/2010, n. 260).

I valori finali degli indici relativi agli elementi di qualità biologica sono espressi utilizzando una scala numerica tra zero e uno in 'Rapporti di qualità ecologica' (RQE). I RQE rappresentano il rapporto tra il valore del parametro biologico osservato per un dato corpo idrico superficiale e il valore previsto in condizioni di riferimento. La valutazione dello stato ecologico è basata sul principio “one-out all-out” (fuori uno, fuori tutti), nel senso che se un singolo elemento di qualità non raggiunge lo stato buono per un particolare corso d'acqua, allora l'intero corpo idrico è classificato come “fallire il raggiungimento degli obiettivi”. Quindi, strumenti in grado di valutare gli scenari relativi allo stato di ogni elemento di qualità biologica potrebbero diventare utili al fine di simulare la qualità dell'acqua. La valutazione dello stato ecologico e la quantificazione delle sue relazioni con le pressioni antropiche dipende dalla conoscenza degli elementi abiotici che influenzano la comunità biologica, la sua composizione e la dinamica; l'elemento abiotico rappresenta, allora, il legame tra le pressioni e la componente biotica. Inoltre, la biocomplexità dell'ecosistema fluviale suggerisce, per definire lo 'stato ecologico', la necessità di considerare sia lo stato di un EQB e sia le sue relazioni con gli altri.

Così, per quanto riguarda i corpi idrici superficiali, PARBLEU è diviso in due sotto-modelli principali: il primo collega le pressioni agli impatti (descritti dai parametri fisico-chimici a

sostegno dello stato ecologico (LIMeco), il secondo collega il LIMeco ai RQE (obiettivo ambientale).

I sotto-modelli lavorano in sequenza (i risultati del primo sono utilizzati come input nel secondo) e raggiungono la convergenza verso la soluzione ottimale attraverso un processo iterativo. Ma, difficoltà dovute all'acquisizione di un'informazioni affidabile relativa alle pressioni hanno portato a sviluppare solo il secondo sotto-modello.

Nel presente lavoro, quindi, le interazioni reciproche tra ogni EQB e gli elementi chimici sono stati investigati utilizzando la regressione multipla, che permette una rapida quantificazione dell'incertezza. Infatti, il modello fornisce una stima del valore medio dei RQE corrispondenti ad una particolare combinazione di valori per le variabili esplicative, per ciascuna stazione di campionamento. Per ogni valore medio previsto, l'intervallo di confidenza è stato calcolato per esprimere il livello di incertezza associato con i valori previsti di ciascun RQE, con un livello di confidenza specificato. La quantificazione dell'incertezza diventa il fulcro di questo lavoro perché permette all'utente di valutare il rischio di un futuro "fallimento" ambientale a fronte di un impegno economico attuale. Il vantaggio di utilizzare i RQE è l'immediatezza della rappresentazione dei risultati, senza altre conversioni. Ma, essi sono espressi come valori medi annuali. Questi valori riferiti ad un singolo anno, diventano rappresentativi dello stato di un corpo idrico per un ciclo di pianificazione di sei anni, nel caso di monitoraggio di sorveglianza (vedi WFD) e per tre anni nel caso di monitoraggio operativo (vedi WFD). Il modo migliore per identificare le relazioni tra variabili ambientali dovrebbe essere l'applicazione di un'analisi spazio-temporale; ma i dati disponibili in questo momento sui RQE non possono dare ancora un'informazione sufficiente per sviluppare modelli basati su serie temporali.

Lo scopo di questo studio è:

- 1) l'individuazione di un sistema di relazioni che permettono all'utente di determinare se lo stato ecologico di un corpo idrico considerato sta migliorando o peggiorando, per un periodo che coincide con il ciclo di sei anni di pianificazione del Piano di Gestione e del Programma di Misure;
- 2) la quantificazione dell'incertezza associata con la singola valutazione di stato dei RQE.

1.2. La scelta del modello

Com'è noto, il problema dell'individuazione del tipo di modello deve tener conto dei seguenti elementi: scala spaziale e temporale, complessità delle relazioni, disponibilità dei dati in ingresso, caratteristiche dei parametri da processare.

1.2.1. Criteri generali di scelta

Un'esauriente ma sintetica rappresentazione dei diversi tipi di modelli è riportata nella Fig. 1.1., tratta da B. A. Cox (2003).

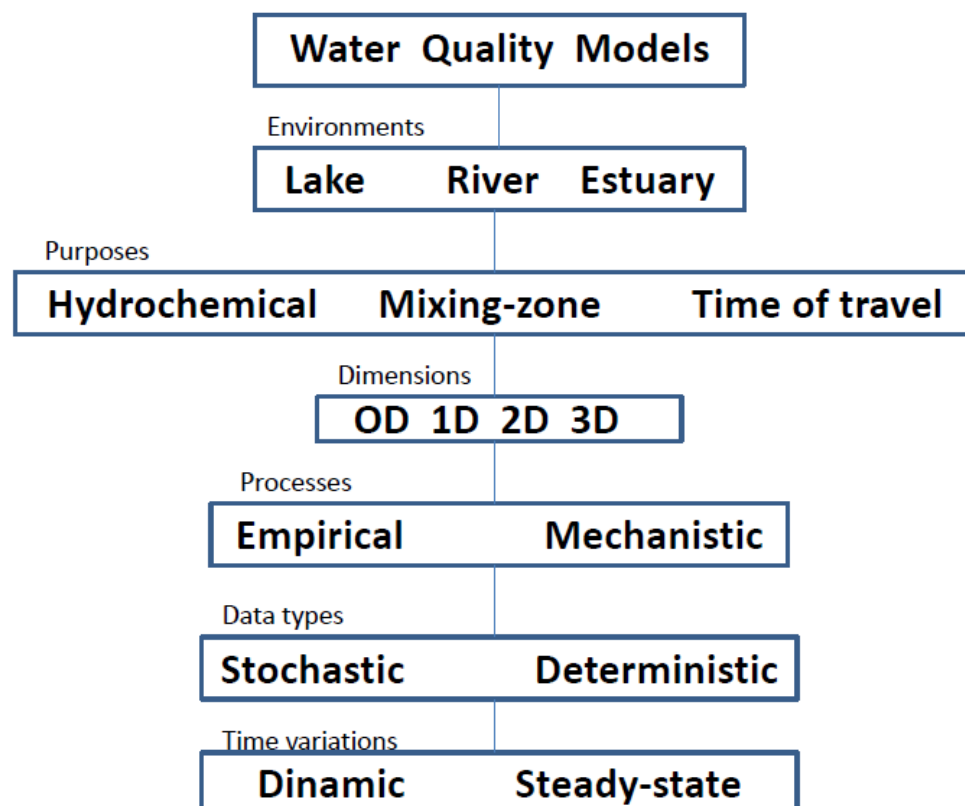


Figura 1.1. Classificazione dei modelli di qualità delle acque

Con riferimento all'alternativa tra modelli empirici e modelli meccanicistici, scrive Cox:

"In larger systems (i. e. greater than 10,000 km²), water quality data are generally collected to provide qualitative analyses while intensive monitoring and data collection will only occur at the small scale (i. e. less than about 10 km²) where the objective is to quantify model parameters for specific processes. Thus, it will generally be the case that data will need to be collated from a number of sources when modelling large systems. The limiting factor of data availability explains the popularity of hybrid/ stochastic models within the EA. These models

will provide the necessary statistical water quality standards, but they also require relatively little data because they do not attempt to represent hydrochemical processes other than by a simple first-order decay rate. Because of this, they can be applied to almost any river catchment with no detailed knowledge of that catchment, but the assumptions made and the limitations of using a model with such simple processes must be acknowledged ...”.

Ancora B. A. Cox e P. G. Whitehead in “Parameter Sensitivity and Predictive Uncertainty in a New Water Quality Model, Q²” (Journal of Environmental Engineering, 2005): “... *it is more important to obtain high quality **forcing functions** than to obtain improved parameter estimates ... there is a need to ensure **reasonable simulation of a range of water quality determinands** ...”.*

Scala spaziale e temporale di lavoro e attenzione alle “forzanti” più che ai parametri costituiscono dunque i primi criteri di orientamento nella scelta del tipo di modello.

Un altro criterio-guida generale è quello che deve condurre alla scelta tra modelli di simulazione e modelli di ottimizzazione. Poiché l’obiettivo finale è quello di scegliere il miglior sistema di misure per conseguire il miglior obiettivo possibile (occorre tener conto che si opera in condizioni di risorse economiche scarse) sarebbe naturale rivolgersi verso modelli di ottimizzazione.

Peraltro la difficoltà di rappresentare in forma adeguata la funzione da ottimizzare (connessa soprattutto ma non solo alla “monetizzazione” dei valori ambientali) e il trovarsi in condizioni molto distanti dalla programmazione lineare portano inevitabilmente verso i modelli di simulazione (composizione di scenari) integrati di volta in volta con la valutazione delle risorse necessarie all’attuazione dello scenario simulato.

In ogni caso, anche se fosse possibile scrivere la funzione di minimizzazione degli impatti o di massimizzazione della qualità ambientale (il Distretto conta più di 600 corpi idrici), la definizione dei vincoli economici (tra i quali non ultima la disponibilità da parte dei portatori d’interesse a pagare i costi e/o ad accettare i condizionamenti alle loro attività per l’attuazione delle misure) necessiterebbe di modelli comportamentali degli agenti socio-economici rappresentabili solo in forma di processi dinamici stocastici.

Il modello in definitiva deve produrre un “insieme” di scenari di impatto sulla base di ipotesi di sistemi di misure tecnicamente fattibili ed economicamente sostenibili. Dagli scenari possibili si determina una “rosa” di impatti caratterizzata dalla sostenibilità economica (sociale

e finanziaria) delle corrispondenti misure. Dalla "rosa" si sceglie lo scenario che determina il miglior risultato in termini di qualità ambientale (secondo un criterio predeterminato).

Infine, a differenza di un modello di ottimizzazione, un modello di simulazione ha il vantaggio di una maggiore comprensione (e in definitiva di una maggiore utilizzabilità) nei processi di partecipazione pubblica che la WFD pone al centro del processo di definizione del piano di gestione.

1.2.2. Il modello fisicamente basato

Un modello fisicamente-basato (o meccanistico) è fondato su una descrizione dettagliata della fisica dei fenomeni (generalmente a parametri distribuiti qualora si sia interessati ad una definizione dei risultati in più punti del dominio di calcolo) e perciò usualmente espresso in termini di equazioni differenziali che esprimono i principi fisici fondamentali quali i bilanci della quantità di moto, dell'energia e della massa. Il modello necessita di una grande quantità di dati ed anche di un notevole impiego di tempo e di capacità computazionale. La difficoltà di tradurre in termini computazionali le equazioni differenziali e l'interazione continua (nel tempo e nello spazio) dei numerosi componenti chimico-fisico-biologici con notevoli difficoltà di fissare gli algoritmi risolutivi sono il maggiore ostacolo da superare e richiedono di ricorrere pertanto a una serie di semplificazioni che “allontanano” sempre di più (via via che si introducono semplificazioni) il modello da un'affidabile rappresentazione dei fenomeni che vuole rappresentare.

Esiste, in altri termini, un *trade-off* tra corretta rappresentazione dei fenomeni e capacità computazionale che riduce – come evidenziato da B. A. Cox – il campo d'azione dei modelli fisicamente basati a condizioni operative estremamente di dettaglio nelle quali il dominio di rappresentazione deve essere facilmente controllabile (sia in termini di grandezze d'interesse sia in termini di fenomeni occorsi).

Tornando alla Fig. 1.1., privilegiati campi d'azione sono il calcolo della “mixing zone” e quello del “time-of-travel” così come quello idro-chimico quando limitato a domini di modesta estensione.

Esempi di modelli fisicamente-basati inerenti alla qualità delle acque sono forniti dall'EPA (United States Environmental Protection Agency) come WASP e Qual2k. Il primo permette di interpretare e prevedere le risposte della qualità delle acque in funzione dei fenomeni naturali, di alcuni descrittori biologici e di inquinamento prodotto dall'attività umana. Esso si avvale di un software di modellazione comportamentale dinamica per sistemi acquatici che consente di

rappresentare sistemi fino a tre dimensioni ed anche dipendenti dal tempo e di simulare il trasporto e la trasformazione di sostanze inquinanti nella colonna d'acqua sovrastante e nel fondo sottostante, cioè nello strato bentonico. Le variabili di stato sono le seguenti: azoto, fosforo, ossigeno disciolto, BOD, domanda di ossigeno nei sedimenti, alghe, periphyton, prodotti chimici organici, metalli, mercurio, agenti patogeni, temperatura. Il modello implementa l'equazione di trasporto della concentrazione dei costituenti della qualità delle acque tenendo conto dei processi di dispersione e advezione, nonché dei carichi di massa puntuali e diffusi e degli scambi con l'ambiente esterno e discretizza in un dominio di calcolo rappresentato schematicamente da una serie di volumi di controllo definiti segmenti.

Il secondo è un modello di qualità delle acque dei fiumi e dei torrenti che rappresenta una versione aggiornata del QUAL2E. Quest'ultimo, usato per simulare la portata in stato stazionario ed il carico inquinante lungo il corso d'acqua, produce come output la concentrazione dell'inquinante e la portata in funzione della distanza. Come strumento di pianificazione può essere impiegato sia in stato-stazionario che dinamico; esso può essere utilizzato per studiare l'impatto degli scarichi inquinanti sulla qualità di un corso d'acqua o per identificare la quantità e la qualità di scarichi inquinanti diffusi, per analizzare gli effetti delle variazioni a ciclo giornaliero dei dati meteorologici sulla qualità dell'acqua (principalmente su ossigeno disciolto e temperatura) ed il comportamento dell'ossigeno disciolto in funzione della crescita algale e della respirazione. Il modello rappresenta il fiume come una serie di 'reaches' (elementi aventi caratteristiche idrauliche uniformi) e ciascun 'reach' viene diviso nuovamente in un numero intero di elementi computazionali. Per ciascun elemento computazionale, il modello esegue un bilancio di portate, un bilancio di massa espresso in termini di concentrazione per ciascun costituente ed un bilancio di calore in termini di temperatura. Le principali cinetiche considerate in QUAL2E sono: la trasformazione di nutrienti, la produzione algale, la domanda bentonica e dei sedimenti a composizione carbonacea, la re-areazione atmosferica (secondo diverse formulazioni) e gli effetti di questi processi sul bilancio dell'ossigeno disciolto, le cinetiche relative all'inquinamento microbiologico causato da Coliformi fecali e ad altre sostanze conservative e non.

L'approccio meccanicistico si basa, quindi, sull' "upscaling" da scale di campo minori a scale di campo maggiori semplicemente descrivendo i processi in modo spazialmente distribuito (Rode, 2010). Questo approccio ha, quindi, le sue debolezze: i modelli meccanicistici sono di solito over-parametrizzati e ci sono difficoltà nel calibrarli, specialmente in condizioni di

scarsità di dati (Keupers and Willems, 2015). In particolare Young (1983) scrive sull'over-parametrizzazione: *“as a result, the model has a degree of ‘surplus content’ not estimated from the available data, but based on a somewhat ad hoc evaluation of all available prior knowledge of the system and coloured by the analyst’s preconceived notions of its behavioral mechanisms”*.

1.2.3. Considerazioni per la scelta: il modello empirico.

Alcune applicazioni di modelli fisicamente-basati per la qualità delle acque sono state effettuate anche dall'Autorità di bacino Tevere pervenendo alla conclusione che la mancanza di informazioni e la scarsità di set differenti di dati di input necessari alla taratura e all'analisi di sensibilità, a causa della bassa frequenza con cui erano avvenuti i campionamenti, non aveva consentito di valutare correttamente la precisione del modello se non “forzando” il processo di taratura (operazione che Cox e Whitehead consigliano ovviamente di evitare). In altri termini l'approssimazione dei modelli fisicamente-basati alla realtà è così bassa da non poter definire scelte a carattere puntuale come richiede la WFD.

Per poter, infatti, effettuare tale valutazione, è necessario considerare:

- l'aderenza dei risultati del modello rispetto ai dati osservati disponibili operando sul minimo numero di parametri possibile al fine di garantire l'aderenza del modello alla realtà fisica evitando il rischio di una over-parametrizzazione;
- il grado di rappresentatività dei dati simulati rispetto alla fenomenologia fisica attesa nella sua interezza e nella sua complessità, fornendo in forma oggettiva il grado di incertezza che necessariamente affligge ogni risultato modellistico.

Quest'ultimo aspetto, in particolare, dipende strettamente dal livello con cui il monitoraggio (qui inteso in senso lato, dai dati sulle pressioni sul corpo idrico fino ai parametri chimico-fisico-biologici del corpo idrico) riesce ad esprimere il fenomeno osservato e deve essere ritenuto in assoluto il più critico non solo per la numerosità dei dati richiesti (che aumenta nel passaggio da un modello che opera in “steady-state condition” in “dynamic condition”) ma soprattutto per il dettaglio richiesto nella caratterizzazione delle pressioni come “carichi” in ingresso al sistema da simulare (si pensi, ad esempio, alla difficoltà di trasformare la

pressione dovuta alle attività agricole in carichi inquinanti diffusi entranti nel reticolo idrografico).

Per questo motivo, è parsa più valida (anche in ragione dei compiti dell'Autorità) l'alternativa costituita da un modello empirico, con l'obiettivo di creare uno strumento per governare un sistema a grande scala all'interno del quale fosse possibile successivamente individuare delle soluzioni puntuali, ottenute attraverso un'applicazione di maggiore risoluzione spaziale e temporale, anche con l'uso di modelli fisicamente-basati. In questa situazione infatti il modello fisicamente basato può essere convenientemente semplificato, senza per questo perdere nella capacità di rappresentare in modo affidabile la realtà fisica in quanto predominanti sono gli aspetti gestiti attraverso un set di dati di minore impegno (conseguito attraverso gli usuali programmi monitoraggio in sorveglianza ed in operativo eventualmente assistiti da specifici monitoraggi d'indagine – si rimanda al Capitolo 1.3 dell'Allegato V della WFD).

1.2.4. La quantificazione dell'incertezza

E' noto che i modelli debbano includere informazioni sull'incertezza dei loro output quando essi sono utilizzati per il supporto alle decisioni (Beven and Binley, 1992; Beven, 2002; Pahl-Wostl, 2002; Jakeman and Letcher, 2003; Refsgaard and Henriksen, 2004; Pahl-Wostl, 2007; Vandenberghe et al., 2007; Uusitalo et al., 2015). L'incertezza si sviluppa dall'informazione inadeguata e dalle assunzioni incorrette, nonché dalla variabilità dei processi naturali. Esistono diverse fonti di incertezza associate ai modelli di simulazione della qualità delle acque; esse sono:

- Quella relativa alla conoscenza imperfetta dei valori dei parametri associati con i processi;
- Quella dovuta alla naturale variabilità sia temporale che spaziale dei dati di misura;
- Quella relativa alla incorretta rappresentazione dei processi.

In riferimento all'incertezza strutturale associate ai modelli meccanicistici, Loucks et. al. (2005) scrivono nel loro libro "Water Resources Systems Planning and Management": *"Placing a priority on process description usually leads to the development and use of*

complex mechanistic models rather than simpler mechanistic or empirical models. In some cases this may result in unnecessarily costly analyses. In addition, physical, chemical and biological processes in terrestrial and aquatic environments are far too complex to be fully represented in even the most complicated models. For water quality management, the primary purpose of modelling should be to support decision-making. The inability to describe all relevant processes completely contributes to the uncertainty in the model predictions”.

Essi aggiungono anche: *“Increasing model complexity in order to more closely represent the complexity of the real system may not only add to the cost of data collection, but may also introduce even more parameters, and thus even more potential sources of error in model output”.*

Dunque, la maggior parte dell’analisi dell’incertezza è tradizionalmente focalizzata sui parametri modellistici e sui dati di input e non su quella strutturale che rappresenta spesso la principale fonte di incertezza nelle predizioni modellistiche (Refsgaard et al., 2006). Ci sono, comunque, ancora difficoltà a condurre un’analisi completa di incertezza inerente ai parametri per i modelli meccanicistici. La difficoltà relativa a tale quantificazione nasce dalla necessità di dover stimare molti parametri durante la fase di calibrazione del modello. Mentre la calibrazione che mira ad individuare una stima puntuale del parametro riducendo la discrepanza tra i dati ed i valori calcolati dal modello è un problema di ottimizzazione, quello inerente alla stima legata all’incertezza dei parametri è un problema di analisi multivariata. Come Tao (2008) sottolinea, il carico computazionale aumenta con l’incremento delle dimensioni parametriche e per i modelli meccanicistici, sarebbe necessario sviluppare nuovi metodi più efficienti ed efficaci da usare nelle applicazioni modellistiche future.

Un recente studio ha valutato lo stato attuale dei modelli di qualità delle acque fisicamente basati a parametri distribuiti relativi ai nutrienti da inquinamento diffuso: i risultati hanno evidenziato che solo il 10% tra 257 di tali studi modellistici tra gennaio 1992 e luglio 2010 ha cercato di quantificare l’incertezza legata alla predizione modellistica in modo quantitativo (Wellen, 2015).

Nei modelli di regressione, invece, l’incertezza strutturale è associata alla possibilità di una specificazione errata.

Quella che generalmente viene quantificata nei modelli di regressione lineare è dovuta solo a due fonti: incertezza associata al termine di errore casuale (dovuta all’inevitabile variabilità naturale) che può essere stimata dall’errore standard e l’incertezza associata alle stime parametriche. La quantificazione dell’incertezza si basa su principi statistici e per i modelli

empirici è inevitabilmente semplice. La struttura base del classico modello statistico è costituito dalla somma di una componente deterministica con una casuale. Il termine di errore è rappresentativo della variazione nei dati che non può essere spiegata dal modello. Per quanto riguarda i modelli lineari statistici, la stima dei parametri, sebbene semplice si basa sul principio della identificabilità (Smith, 2004) e la varianza ad essi associata diventa una misura dell'incertezza.

2. Area di studio

Questo studio si focalizza nelle stazioni di campionamento lungo i fiumi della regione Abruzzo (Fig. 2.1.), una gran parte della quale appartiene al Distretto dell'Appennino Centrale, costituito da bacini idrografici autonomi ma anche da bacini idrogeologici, più o meno interconnessi, i cui "confini" travalicano il distretto stesso.

I maggiori fiumi sono il Sangro, l'Aterno-Pescara ed il Tronto che costituisce insieme con il Tordino ed il Vomano un'unica entità idrologica a causa delle forti interconnessioni idrauliche afferenti al grande schema idroelettrico regolato dal grande invaso artificiale del Lago di Campotosto

La struttura della montagna che corre parallela alla costa determina il corso dei fiumi: la rete idrografica che drena verso il mare Adriatico è direttamente relazionata alle caratteristiche della catena montuosa dell'Appennino Centrale che include vette come il Massiccio del Gran Sasso (2912 metri sul livello del mare) ed la Maiella (2793 m), costituito principalmente da rocce carbonatiche. I fiumi più importanti sono alimentati da sorgenti della parte interna della catena e hanno portate costanti che forniscono acqua per molti usi, soprattutto civili ed idroelettrici, anche attraverso dighe e serbatoi. Nei tratti iniziali, essi scorrono lungo gli assi longitudinali delle valli per modificare la loro strada attraverso gole strette e profonde dove si trovano gli impianti idroelettrici. Nel tratto finale essi raggiungono il mare attraverso la pianura costiera, dopo aver fornito l'acqua per l'irrigazione e dopo che hanno ricevuto acque reflue urbane trattate. Altri fiumi hanno corsi più brevi e portate minori: provengono da sorgenti poste a livelli più bassi della catena montuosa e fluiscono attraverso le colline e la fascia pedemontana, prima di raggiungere il mare (per esempio: Vibrata, Salinello, Fino, Tavo, Foro), oppure appartengono completamente alle colline (es: Piomba, Alento, Orsento, Sinello). Due diversi scenari di percorsi fluviali possono essere distinti nella Regione Abruzzo: la zona interna, con bassa densità di popolazione, è essenzialmente dedicata all'agricoltura, mentre lungo la costa ed i bassi corsi dei fiumi, con una densità fino oltre molte centinaia di abitanti per chilometro quadrato, sono situati importanti insediamenti civili (industriali e residenziali) e centri di turismo.

E' importante sottolineare che il modello non è sviluppato sulla rete idrografica complessiva dell'area di studio, ma solo sugli elementi identificati come corpi idrici che in base alla Direttiva europea rappresentano le unità a cui fare riferimento per riportare ed accertare la

conformità con gli obiettivi ambientali e che devono essere sottoposti ad attività conoscitive ed osservative. Dalla normativa diverse tipologie di monitoraggio sono previste (Sorveglianza, Operativo, Indagine) ognuna delle quali ha specifiche finalità ed obiettivi; l'organizzazione dei programmi di monitoraggio dipende dal risultato dell'analisi di rischio associato al mancato raggiungimento degli obiettivi ambientali.

In generale, si può affermare che per i corpi idrici “non a rischio” viene attuato un monitoraggio definito di “sorveglianza”, mentre per i corpi idrici “a rischio” il monitoraggio è di tipo “operativo”. Il monitoraggio di indagine, invece, si attua quando sono sconosciute le cause del mancato raggiungimento degli obiettivi ambientali non essendo disponibile il monitoraggio operativo, o per valutare l'estensione e l'impatto dovuti ad eventi di inquinamento accidentale. All'interno della rete di monitoraggio di sorveglianza si individua anche una rete nucleo, realizzata per tenere sotto osservazione l'evoluzione dello Stato Ecologico dei siti di riferimento, valutare le variazioni a lungo termine derivanti da una diffusa attività antropica e quelle a lungo termine delle condizioni naturali.

La Regione Abruzzo ha individuato 26 bacini idrografici con un totale di 111 corpi idrici con un totale di 121 stazioni di campionamento (Fig. 2.1.). Nel 2010 ha organizzato il programma di monitoraggio disponendo 58 stazioni per la rete di sorveglianza, 13 per la rete nucleo e 50 per la rete operativa.

E' necessario inoltre menzionare che la WFD ha introdotto il concetto di Idro-ecoregione ossia di aree geografiche, definite sulla base di fattori quali l'orografia, la geologia e il clima, all'interno delle quali gli ecosistemi di acqua dolce dovrebbero presentare una limitata variabilità di caratteristiche chimiche, fisiche e biologiche. Il territorio della Regione Abruzzo è diviso in tre principali idroecoregioni corrispondenti alle diverse condizioni di struttura geomorfologica, regime idrologico, caratteristiche ecologiche e modello socio-economico (in base alla legislazione italiana). Esse sono precisamente: Costa Adriatica (cod. 12), Appennino Meridionale (cod. 18) Appennino Centrale (cod. 13).

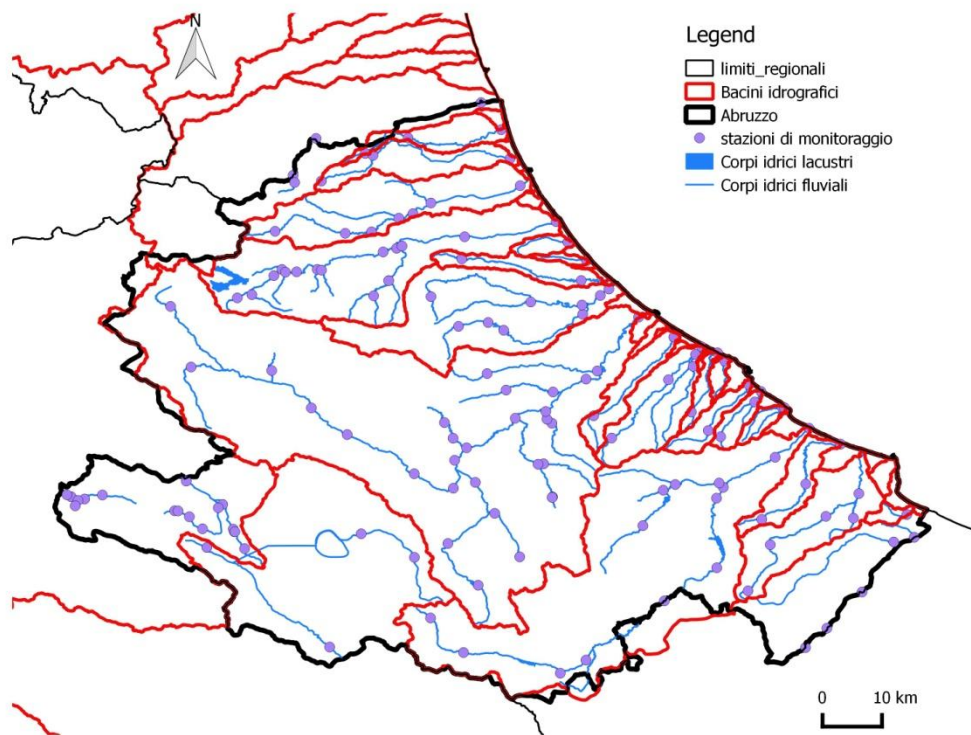


Fig. 2.1. Le stazioni di campionamento ed i bacini idrografici nella regione Abruzzo

3. Materiali e metodi

Lo schema del modello concettuale è descritto in Fig. 3.1.. Esso rappresenta le relazioni degli impatti (I) prodotti dalle pressioni (P) sulle matrici ambientali: sedimento (S); acqua (W); biota (B). Il modello proposto è stato definito PARBLEU che rappresenta l'acronimo di Pressures/impacts/responses Assessment into River Basins Leading to Environmental Uses.

Nel presente lavoro, solo un sottomodello di PARBLEU è stato sviluppato ed investigato e precisamente quello relativo alle interazioni degli impatti con le matrici acqua e biota. L'approccio modellistico utilizzato risulta pertanto utile per determinare le particolari relazioni del sistema descritto in Fig. 3.1. (vedi oltre), nel momento in cui un database più completo e strutturato sarà disponibile.

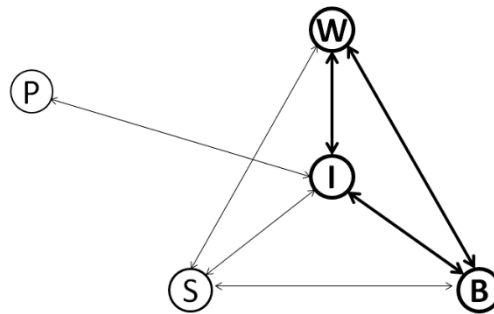


Fig. 3.1. Lo schema concettuale del modello che rappresenta le relazioni degli impatti (I) prodotti dalle pressioni (P) sulle matrici ambientali (sedimento S, acqua W, biota B). Le linee più spesse indicano le relazioni sviluppate nel sottomodello PARBLEU.

Al fine di rendere la complessità delle relazioni tra impatti e biota (dovuta a fenomeni di interazione tra le specie e tra le comunità e anche tra gli stessi elementi abiotici), difficile (se non addirittura impossibile) da rappresentare attraverso l'elaborazione di un modello deterministico, si è scelto di approssimare il comportamento del sistema attraverso le equazioni empiriche di regressione utilizzando i parametri disponibili dal monitoraggio che entrano nella procedura di classificazione dello stato ecologico. E' opportuno sottolineare di nuovo che interesse dell'Autorità è l'identificazione di una strategia di contrasto alle pressioni a grande scala simulando (a livello di singola stazione di monitoraggio) gli stati ecologici attraverso il controllo degli impatti sull'acqua, prodotti dalle pressioni. Tutte le misure previste

dalla WFD hanno incidenza sugli elementi chimici e fisici nella convinzione che un ristabilimento di questi vicino ai valori naturali induca un miglioramento dell'ecologia fluviale.

E' invece nell'interesse delle singole Regioni, che assumono tale strategia come comune riferimento all'interno del proprio territorio amministrato, sviluppare, laddove ritenuto necessario a livello di singolo corpo idrico (o di aggregati di questi), una procedura più spinta di definizione degli scenari a patto di disporre di un maggiore e più strutturato set di dati.

L'applicazione delle tecniche di regressione multivariata permette di rappresentare in forma semplificata ma completa la relazione esistente tra i due set di variabili (biota e impatti misurati), prescindendo dalle leggi fisiche che governano i processi. Solo apparentemente si perde la fisicità della realtà indagata (ad esempio: la capacità autodepurativa del corso d'acqua o i fenomeni di trasporto degli inquinanti conservativi) in quanto questa risulta intrinsecamente compresa nelle variabili da correlare. La scelta di queste ultime è il problema centrale dei modelli empirici.

La regressione lineare multivariata studia la variazione in media della variabile dipendente Y rappresentata nel modello considerato dal biota, rispetto a più variabili esplicative (regressori) rappresentate dagli impatti; l' i -esima osservazione Y è formalmente determinata dalla combinazione di una componente sistematica, espressione del valore atteso condizionato di Y dati i valori delle variabili esplicative, e di una componente stocastica, espressione degli effetti non controllati, più in generale di tutti quei fenomeni casuali che inevitabilmente intervengono nell'occorrenza di un evento reale. E' bene sottolineare che proprio la componente stocastica è in grado di produrre in forma oggettiva (e contestualmente al risultato) il desiderato livello d'incertezza del risultato delle simulazioni, laddove nei modelli fisicamente-basati la valutazione dell'incertezza deve essere condotta attraverso ulteriori procedure computazionali. La valutazione dell'incertezza non è secondaria rispetto al risultato in quanto concettualmente collegata al principio di precauzione.

Indipendentemente dalle tecniche usate per verificare l'affidabilità del modello, la scelta dei regressori e della variabile dipendente deve comunque avvenire in base a considerazioni fisicamente-basate.

Anche il modello empirico ha bisogno di numerosi dati ma, a differenza del modello fisicamente basato, l'elevata numerosità si riferisce alle grandezze (biota ed impatti; pressioni

ed impatti) prese in considerazione direttamente osservabili e misurabili. Per esempio, in relazione ai carichi inquinanti diffusi prodotti dall'attività agricola (nutrienti e fitofarmaci), mentre un modello fisicamente basato richiede il loro valore in ingresso al corpo idrico stimabile (attraverso un ulteriore modello fisicamente-basato) ma non osservabile né misurabile (se non a patto di complesse e costose operazioni di campagna), il modello empirico può essere alimentato con i dati della superficie agricola utilizzata che può essere effettivamente osservata e misurata. In altri termini a parità di condizioni (disponibilità dei dati ed estensione del dominio di calcolo) il modello empirico risulta più efficiente in quanto fornisce come risultato aggiuntivo il livello di incertezza.

La contropartita è – come già accennato – la perdita di risoluzione: se il modello fisicamente basato fornisce un risultato apprezzabile localmente, il modello empirico fornisce un risultato “mediamente” valido nel dominio spaziale (o temporale) indagato.

Lo schema concettuale del sotto-modello ovvero la specificazione iniziale è rappresentato in fig. 3.2.

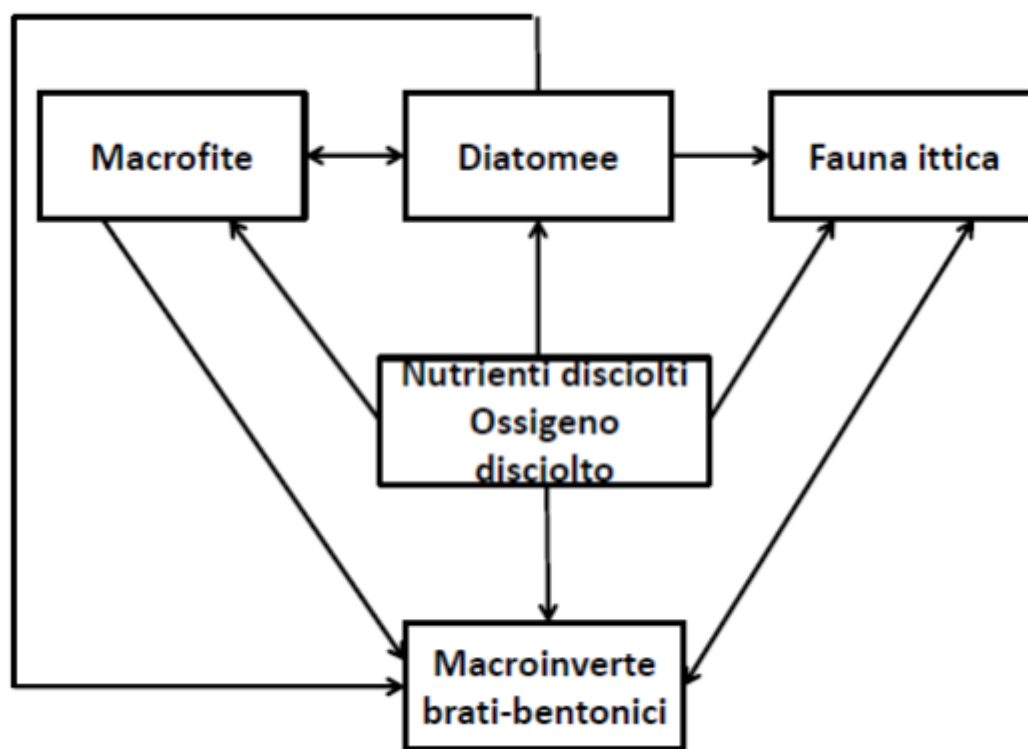


Fig. 3.2. Schema del sotto-modello concettuale ovvero la specificazione iniziale elaborata usando altri schemi implementati nei modelli meccanicistici (vedi Fig. 3.3.) o trovati in letteratura (Covich et al.,1999 – in particolare vedi oltre Fig. 3.4.).

Lo schema 3.2. è stato ottenuto dall'analisi di altri schemi usati per sviluppare processi ed equazioni implementati nei modelli meccanicistici (vedi Fig. 3.3.) o trovato in letteratura (esempio: Covich et al.,1999 – in particolare vedi oltre Fig. 3.4.).

Si è fatto soprattutto riferimento allo schema di Covich che è stato semplificato nel modo seguente: il detrito organico, zooplankton, N_2 , CO_2 , CH_4 , H_2S sono stati eliminati, non avendo dati a disposizione per rappresentarli. Lo schema è riferito agli ecosistemi lacuali. Tuttavia, pur trattandosi di taxa diversi, il sistema delle relazioni incrociate tra gli elementi abiotici e biotici rimane invariato: infatti la WFD prevede per la classificazione gli stessi indici biotici calcolati appunto con riferimento ai diversi taxa coinvolti.

Dallo schema di Covich, è possibile intuire il ruolo chiave dei macro-invertebrati bentonici nel ciclo dei nutrienti, sebbene descritto attraverso la trasformazione del detrito organico da loro operata; essendo il detrito materiale organico morto costituito da materia vivente morta (animale e vegetale) e dagli scarti fecali, l'influenza degli altri gruppi sui macroinvertebrati è evidenziata attraverso le frecce indicate nello schema relativo alla specificazione iniziale del modello.

Ma i macroinvertebrati-bentonici sono una fonte di cibo importante per i pesci; per tali motivi nasce l'esigenza di indicare una doppia freccia tra i due gruppi.

Il link delle macrofite con i pesci ed i macroinvertebrati è rappresentato nello schema tratto dal sito dell'EPA: le macrofite sono un'importante risorsa di cibo per numerosi organismi, fornendo sia materiale organico morto (come precedentemente evidenziato) che vivo (Thomaz and Cunha, 2010). Inoltre essi offrono un habitat agli organismi acquatici; le caratteristiche di tale habitat possono influire sulla distribuzione, densità e diversità delle specie dei macroinvertebrati (Carpenter and Lodge, 1986); inoltre, esso è uno dei principale fattori che struttura la comunità dei pesci nell'esplorazione delle fonti di cibo e nella deposizione delle uova.

Varie ricerche hanno dimostrato l'influenza delle macrofite sulla composizione tassonomica e la produzione di biomassa delle microalghe perfitiche (qui diatomee) (Fernandes, 2016). Ma le diatomee, essendo produttori primari alla base della catena trofica, sono importanti nel determinare l'equilibrio dei livelli trofici superiori, per cui qualsiasi alterazione nella composizione e struttura della comunità diatomica può potenzialmente ripercuotersi sui livelli trofici successivi. Nasce la necessità di indicare una doppia freccia tra diatomee e le piante

acquatiche per far riferimento alle relazioni di competizione e cooperazione che si instaurano tra loro e di indicare i links delle diatomee con i restanti gruppi;

Il sottomodello utilizza i rapporti di qualità ecologica (RQE) ed i 4 componenti del LIMeco. I RQE sono definiti nell'allegato V della WFD come il rapporto tra il valore del parametro biologico osservato per un dato corpo idrico superficiale ed il valore atteso sotto le condizioni di riferimento. Essi sono espressi come un valore numerico tra zero ed uno, con valori tendenti ad uno o superanti l'unità, per rappresentare lo stato ecologico "elevato" e con valori che tendono a zero per rappresentare uno stato ecologico "cattivo". La WFD afferma che la necessità di esprimere risultati attraverso i RQE nasce dall'esigenza di assicurare una comparabilità tra metodi di valutazione diversi ossia di fornire una scala comune di qualità ecologica. Il concetto di RQE incorpora i requisiti chiave della WFD per la classificazione dello stato ecologico: tipologia, condizioni riferimento e limite delle classi. Si rimanda all'Allegato V della WFD.

Durante la procedura di specificazione è stata introdotta, comunque, una variabile *dummy* (dicotomica) definita come IDRO-ECOREGIONE, per verificarne la significatività statistica. Essa è stata codificata in questo modo: 0 se il corpo idrico appartiene all'idro-ecoregione Costa Adriatica (cod. 12); 1 se il corpo idrico appartiene all'idro-ecoregione Appennino Meridionale (cod. 18) o Appennino Centrale (cod. 13). La percentuale di corpi idrici inserita nell'Idro-ecoregione 18 è molto bassa e quindi la sua influenza sui risultati finali ritenuta non significativa. Inoltre, il ricorso ai RQE annulla di fatto la significatività di tale variabile

Allo stesso modo, l'uso dei RQE riduce anche l'influenza degli altri indici che non sono disponibili come l'alterazione del flusso di regime idrologico e la qualità idromorfologica. L'interruzione, infatti, della continuità fluviale o l'alterazione del regime idrologico possono condizionare negativamente le comunità biologiche e quindi influire sul raggiungimento/mantenimento degli obiettivi di qualità ecologica. Inoltre indici come l'alterazione del flusso di regime idrologico e la qualità idromorfologica sono importanti solo per controllare la valutazione di stato ecologico "elevato" (si sottolinea ancora una volta che il modello ipotizzato è funzionale alla procedura di classificazione della WFD inserita nel processo che deve individuare le misure più idonee a conseguire un determinato obiettivo e non ha l'obiettivo di rappresentare il "reale" funzionamento degli ecosistemi fluviali). Gli

indici di inquinamento chimico non sono stati considerati tanto rivelanti dal momento che essi non indicavano (dai dati del monitoraggio) una situazione critica.

La legislazione italiana ha, invece, stabilito come indice per la qualità chimico-fisica il LIMeco a supporto degli elementi biologici. Il LIMeco prevede l'integrazione dei seguenti macrodescrittori: ossigeno disciolto (OD), espresso come percentuale di saturazione; nutrienti (azoto nitrico (N-NO₃), azoto ammoniacale (N-NH₄), fosforo totale (PT)).

Lo schema concettuale semplificato è stato trasformato in un sistema generale di regressioni e quindi sottoposto alla procedura di specificazione (si anticipa qui che dopo la procedura di specificazione non è stata condotta la procedura di validazione che avrebbe richiesto un ulteriore campione di stazioni di monitoraggio al fine di non ridurre quello utilizzato in sede di specificazione). Ciascuna regressione è espressa nel seguente modo:

$$RQE = \beta_0 + \beta_i * I_i + \beta_k * RQEs_k + \varepsilon$$

In cui:

- RQE è il rapporto di qualità ecologica scelta come variabile dipendente;
- RQE_k sono i restanti rapporti di qualità ecologica (considerati repressori secondo la complessità dell'ecosistema fluviale);
- I sono gli impatti costituiti dai componenti del LIMeco;
- β_0 è il coefficiente di intercetta;
- β_i e β_k sono i coefficienti associati ai repressori;
- ε è il termine d'errore.

Tutti i risultati discussi sono stati ottenuti dal "R Project for Statistical Computing version 2.11.1" (<http://www.r-project.org>). Tutti i comandi e le funzioni necessarie sono ben descritti in: "Principali Tecniche di regressione con R". Vito Ricci (2006), che può essere scaricato da: <http://cran.rproject.org/doc/contrib/Ricciregression-it.pdf>, e utilizzato nei termini imposti dalla Free Software Foundation (si veda il sito web <http://www.fsf.org/>).

I pacchetti statistici utilizzati e scaricati da CRAN (Comprehensive R Archive Network) sono stati i seguenti:

- car (Companion for Applied Regression) Package - John Fox e Sanford Weisberg- che contiene per lo più funzioni per la regressione applicata, modelli lineari e modelli lineari generalizzati , con particolare attenzione alle diagnostiche di regressione ed in modo particolare ai metodi diagnostici grafici;
- lmtest Package - Torsten Hothorn, Achim Zeileis, Giovanni Milla, David Mitchell - che fornisce utili strumenti diagnostici per modelli lineari;
- sandwich (Robust Covariance Matrix Estimators) Package- Thomas Lumley], Achim Zeileis che contiene funzioni per il calcolo degli errori standard robusti rispetto all'eteroschedasticità per serie temporali, per dati trasversali e longitudinali;

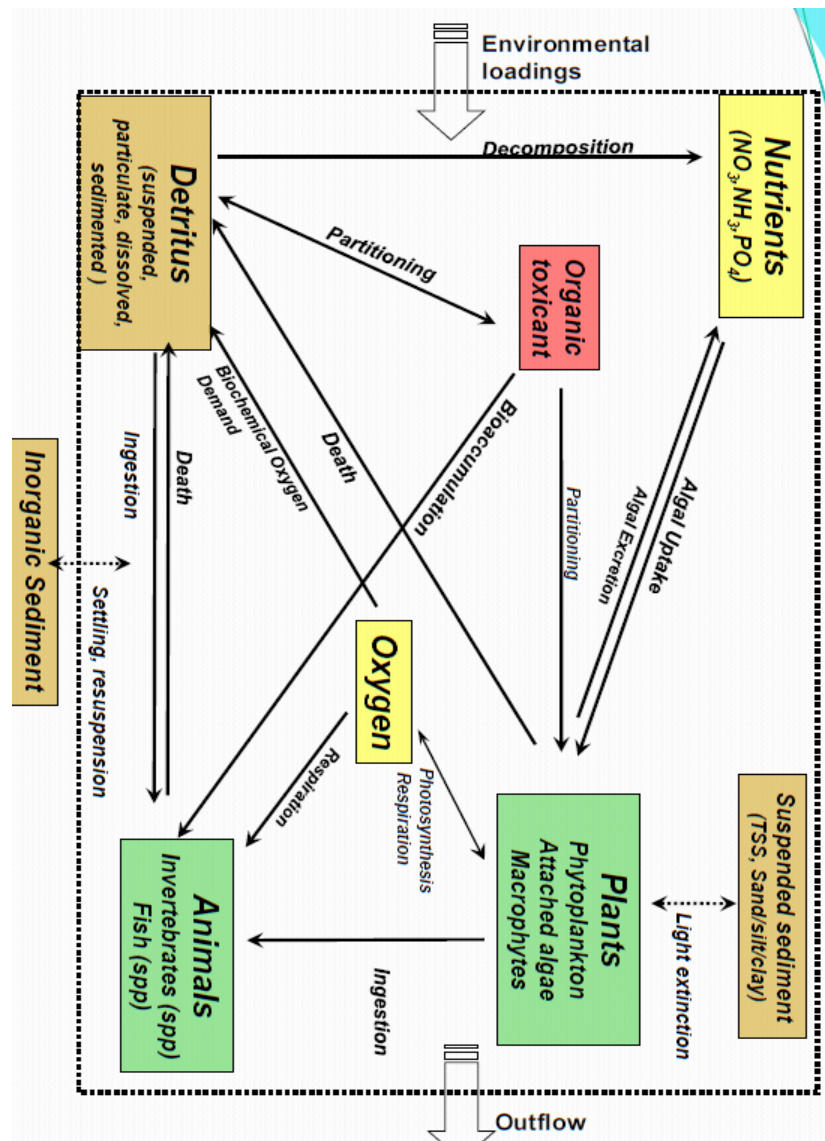


Fig. 3.3. Schema tratto da "<https://www.epa.gov/...models/water-quality>"

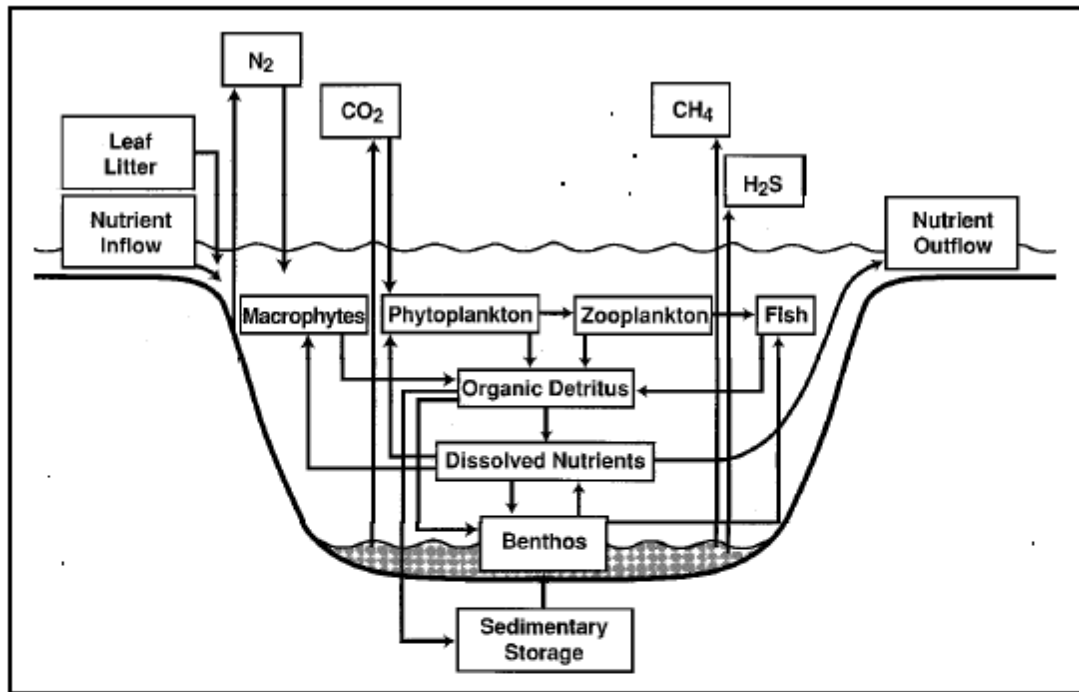


Fig. 3.4. Schema tratto da Covich et al. – 1999

3.1 Organizzazione del modello

Le diverse fasi affrontate nell'organizzazione del modello sono state:

- il recupero dei dati in ingresso;
- la costruzione del campione statistico;
- la scelta della variabile dipendente e dei repressori e della migliore forma funzionale ovvero la specificazione del modello;
- la stima dei parametri del modello con verifica delle principali assunzioni;
- calcolo dell'incertezza attraverso la stima degli intervalli di confidenza
- risoluzione e rappresentazione delle regressioni.

Le prime tre fasi costituiscono la fase di pre-processamento del modello. Ciascuna fase viene in seguito descritta in dettaglio.

3.1.1 I dati in ingresso

L'analisi è stata condotta sui corpi idrici fluviali che per numerosità e rappresentatività sono stati considerati il miglior "laboratorio" d'analisi. Gli elementi presi in considerazione sono quelli relativi allo stato di qualità ecologica in quanto lo stato chimico a livello distrettuale non rappresentava un significativo fattore di rischio (alcune criticità a scala locale sono risolvibili attraverso opportune misure nei piani di tutela delle acque).

Dei circa 500 fiumi appartenenti a strutture sub-distrettuali diverse, sono stati scelti quelli monitorati dalle 121 stazioni della Regione Abruzzo nel corso del 2010 (per omogeneità di osservazione e misura, attuazione di interventi e caratteristiche fisico-ambientali). I dati sono stati forniti dall'ARTA che è l'organo istituzionale responsabile dei monitoraggi ambientali e dei controlli all'interno della regione Abruzzo.

Le 121 stazioni di campionamento fluviali appartenenti a 111 corpi idrici furono identificate spazialmente. I dati degli elementi di qualità biologica espressi come "RQE" ed il LIMeco con i suoi 4 componenti (OD: ossigeno disciolto, N-NH₄: azoto ammoniacale, N-NO₃: azoto contenuto nei nitrati e TP: fosforo totale), associati a ciascuna stazione furono analizzati valutandone l'ordine di grandezza, valore massimo e valore minimo, rappresentati in GIS per capirne la distribuzione; i valori che non rispettavano l'ordine di grandezza non furono presi in considerazione perché ritenuti errori di misura. Anche il dato relativo alle pressioni espresso come numero di impianti di depurazione e carico, come attività produttive insistenti su ciascun corpo idrico, numero di siti contaminati, impianti idroelettrici fu analizzato ma ritenuto insoddisfacente ancora a fornire un'informazione completa. Informazioni sugli indici usati per la valutazione dello stato ecologico sono stati forniti dall'ARTA (secondo il decreto ministeriale n. 260 dell' 8 Novembre 2010).

3.1.1.1 Macroinvertebrati Bentonici

Lo STAR Intercalibration Common Metric Index (STAR_ICMi) è l'indice usato per la classificazione dello stato ecologico considerando i macroinvertebrati bentonici. Lo STAR_ICMi è stato calcolato usando il sistema Macropen aggiornato alla penultima versione del 2012;

3.1.1.2 Macrofite

L'indice "Macrophytes Biological Index for Rivers" (IBMR) è quello usato per le macrofite; esso è stato calcolato sulla base di fogli di calcolo preparati dall'Istituto Nazionale di Ricerca sulle Scienze e Tecnologie per l'Ambiente e l'Agricoltura francese, il CEMAGREF, che prevedono l'inserimento dei dati dei rilievi floristici e l'espressione del risultato dell'Indice;

3.1.1.3 Diatomee

L'indice "Intercalibration Common Metric Index" (ICMi) è usato per la valutazione dello stato ecologico attraverso le comunità diatomeiche considerate rappresentative del phytobenthos. Esso è stato elaborato sulla base di fogli di calcolo integrati dall'Agenzia sulla base delle liste di specie indicatrici e dei relativi coefficienti di sensibilità;

3.1.1.4 Fauna ittica

L'indice ISECI (Indice dello Stato Ecologico delle Comunità Ittiche) è quello utilizzato per la valutazione dello stato ecologico attraverso la fauna ittica. Il campionamento e la valutazione della componente biologica, sono stati effettuati completamente dal personale dei Distretti Provinciali ARTA; l'analisi dei dati popolazionistici (struttura e demografia delle popolazioni ittiche rilevate), necessaria per l'applicazione dell'indice, è stata effettuata utilizzando il software FISATII (fornito gratuitamente dalla FAO), mentre per il calcolo dell'indice ISECI è stato utilizzato un foglio di calcolo appositamente preparato dal personale tecnico dei Distretti di L'Aquila e Teramo.

3.1.2 La costruzione del campione statistico

Dalle 121 stazioni di monitoraggio sono stati tratti due campioni rappresentativi dei parametri sopramenzionati considerando i valori registrati durante l'anno 2010 e 2011 durante i quali le attività di monitoraggio furono completate.

Indipendentemente dal tipo di monitoraggio, è stato osservato che la maggior parte delle stazioni in Abruzzo erano state campionate nel 2010, dal punto di vista biologico. Invece il

monitoraggio chimico era stato effettuato su frequenza trimestrale durante ciascun anno del triennio 2010-2012.

Risultò necessario individuare un valore medio annuale di ciascun parametro chimico relativo al LIMeco associato a ciascuna stazione di campionamento e allo stesso anno in cui i RQE erano stati elaborati. Occorre ricordare, infatti che i RQE pur essendo ricavati da attività di monitoraggio svolte in un anno secondo le frequenze stabilite dalla direttiva, sono riportati come valori medi annuali.

La costruzione del campione è fortemente relazionata alla sua specificazione. In base a quest'ultima furono costruiti due campioni statistici: uno era costituito dalle stazioni dove tutti gli elementi sopramenzionati eccetto RQE-STAR-ICMi ed ISECI erano stati calcolati (108 stazioni); un altro era costituito dalle stazioni dove tutti gli elementi sopra indicati erano stati calcolati (88 stazioni). Ambedue i campioni fornivano un'informazione completa ed integrata. In questo studio, è stato preferito sacrificare il processo di validazione in favore di un grande campione.

3.1.3 La scelta della variabile dipendente e dei regressori: la specificazione del modello

La specificazione del modello consiste sia nella scelta delle variabili indipendenti da includere ed escludere che nell'individuazione della migliore forma funzionale. Esso rappresenta il primo e lo stadio più critico di tutti: sia la stima dei parametri che l'interpretazione dipende dalla corretta specificazione del modello (Allen, 1998).

La specificazione iniziale relativa alla scelta delle variabili indipendenti da includere nel modello di regressione, come evidenziato nell'introduzione, è stata effettuata quindi su delle basi teoriche.

Occorre premettere, però, come sottolineato da Liu (1960) che i limiti dei dati più che quelli teorici, sono principalmente la causa della tendenza persistente a semplificare i modelli di regressione.

Il sistema generale di regressioni investigate è il seguente:

1. $RQE_ICMi = \beta_0 + \beta_1 RQE_IBMR + \beta_2(N - NH_4) + \beta_3(N - NO_3) + \beta_4 PT + \beta_5 OD + \varepsilon_1;$
2. $RQE_IBMR = \beta'_0 + \beta'_1 RQE_ICMi + \beta'_2(N - NH_4) + \beta'_3(N - NO_3) + \beta'_4 PT + \beta'_5 OD + \varepsilon_2;$
3. $RQE_STAR - ICMi = \beta''_0 + \beta''_1 RQE_ICMi + \beta''_2(N - NH_4) + \beta''_3(N - NO_3) + \beta''_4 PT + \beta''_5 OD + \beta''_6 EQR_IBMR + \beta''_7 ISECI + \varepsilon_3;$
4. $ISECI = \beta'''_0 + \beta'''_1 RQE_ICMi + \beta'''_2(N - NH_4) + \beta'''_3(N - NO_3) + \beta'''_4 PT + \beta'''_5 OD + \beta'''_6 RQE_IBMR + \beta'''_7 RQE_STAR - ICMi + \varepsilon_4.$

In cui:

- RQE_ICMi è l'indice $ICMi$ espresso come RQE ;
- RQE_IBMR è l'indice $IBMR$ espresso come RQE ;
- $RQE_STAR - ICMi$ è l'indice $STAR_ICMi$ espresso come RQE ;
- $ISECI$ è l'indice per la fauna ittica (è interessante notare che il calcolo dell' $ISECI$ è basato su 5 indicatori chiave ed il calcolo di tali indicatori è basato sul confronto tra i valori misurati ed attesi nelle condizioni autoctone; così il valore numerico finale dell' $ISECI$ è un valore normalizzato tra zero ed uno ed è assimilabile ad un RQE);
- $\beta_0, \beta_1, \dots, \beta_5$ sono rispettivamente l'intercetta ed i coefficienti di regressione, nella regressione del RQE_ICMi ;
- $\beta'_0, \beta'_1, \dots, \beta'_5$ sono rispettivamente l'intercetta ed i coefficienti di regressione, nella regressione del RQE_IBMR ;
- $\beta''_0, \beta''_1, \dots, \beta''_7$ sono rispettivamente l'intercetta ed i coefficienti di regressione, nella regressione del $RQE_STAR - ICMi$;
- $\beta'''_0, \beta'''_1, \dots, \beta'''_7$ sono rispettivamente l'intercetta ed i coefficienti di regressione, nella regressione dell'indice $ISECI$;
- $\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4$ sono rispettivamente gli errori associati a ciascuna regressione;
- $(N - NH_4), (N - NO_3), OD$ e PT sono espressi come concentrazioni medie annuali.

La specificazione finale relativa alla scelta della forma funzionale più appropriata è stata raggiunta combinando le basi teoriche con quelle statistiche attraverso due differenti approcci.

Riguardo al primo approccio, lo step iniziale è stato quello di svolgere un'analisi preliminare alla regressione computando la matrice di correlazione e disegnando gli scatterplots nei quali ciascuna variabile era plottata contro ciascuna altra (saranno di seguito dettagliatamente

descritti). I giudizi qualitativi forniti dall'analisi degli scatterplots furono successivamente migliorati implementando un'analisi di regressione semplice per ciascuna coppia di variabili. Differenti forme funzionali furono comparate per ciascuna di queste.

Le seguenti forme funzionali furono scelte: lineare, lineare-logaritmica e polinomiale. Il confronto fu svolto usando solo le forme funzionali menzionate perché costituivano modelli lineari nei parametri.

L'introduzione di trasformazioni non lineari come quella di potenza o esponenziale avrebbe aggiunto una complessità maggiore (soprattutto nelle stime di regressione multipla) inutile ai fini modellistici. In tali casi la stima dei parametri dovrebbe essere effettuata attraverso o la trasformazione a sistemi lineari (metodi di linearizzazione), o minimi quadrati non lineari. La linearizzazione si basa sul cambiamento di variabili ma la riparametrizzazione potrebbe portare alla perdita del valore in senso fisico dei parametri trasformati. Inoltre non tutti i modelli possono essere resi lineari. Essa si basa sulla trasformazione della variabile dipendente. Un grave svantaggio è dato dal fatto che trasformando la Y si trasforma anche la distribuzione degli errori, introducendo distorsioni sulle assunzioni di base per l'inferenza statistica, che potrebbero risultare essere non più soddisfatte. I minimi quadrati non lineari, invece, rappresentano un problema di minimizzazione non lineare. Il principale svantaggio dei modelli non lineari è la mancanza di una soluzione analitica per la stima dei parametri. Tale svantaggio comporta il ricorso a metodi numerici; i problemi principali relativi a quest'ultimi sono legati alla difficoltà di convergenza della soluzione che può dipendere dalla cattiva scelta dei parametri iniziali o dalla non corretta selezione del miglior modello che si adatta bene ai dati. Inoltre, trattandosi di metodi iterativi, le stime ottenute sono solo un'approssimazione (accettabile!) dei valori reali.

Il secondo approccio era basato sul capire come ciascun RQE fosse simultaneamente relazionato alle variabili scelte come predittori usando la regressione multipla. Differenti tecniche furono usate per specificare ciascun regressione: il primo modo consisteva sul confronto di più modelli ottenuti combinando diverse forme funzionali delle variabili scelte come variabili indipendenti. Esso era basato sul confronto della significatività dei coefficienti di regressione, coefficiente di determinazione, significatività statistica totale con l'uso del test ANOVA, il risultato del test di Breush-Pagan e di Shapiro. Sebbene sia impossibile stimare tutte le possibili combinazioni delle variabili indipendenti, modificando le diverse forme

funzionali, tale approccio ha permesso di effettuare delle considerazioni. Esso è stato affiancato con un secondo approccio basato sui “partial residual plots” (detti anche “component+residual plots”) che rappresentano degli strumenti di supporto ai grafici dei residui. La formulazione finale, per ciascuna regressione del modello, è stata raggiunta usando una procedura di eliminazione backward.

Ciascuna tecnica menzionata è descritta di seguito in dettaglio.

3.1.3.1 La matrice di correlazione

La matrice di correlazione $A = (\alpha_{ij})$ è una simmetrica matrice in cui α_{ij} è il coefficiente di correlazione di Pearson del data set corrispondente alla i-esima e j-esima colonna di A.

E’ ben noto che il coefficiente di correlazione che assume valore pari ad 1 o a (-1) indica una forte relazione lineare con una pendenza positiva (o negativa) tra le due variabili coinvolte, mentre un coefficiente che assume valore zero non indica nessuna relazione lineare. Sebbene Chatfield et al. (1980) definiscono imprudente calcolare il coefficiente di correlazione per una coppia di variabili la cui relazione è ovviamente non lineare e sebbene per le variabili considerate potrebbe essere alta la probabilità di una relazione non lineare, tale matrice è stata computata in tale studio per avere una indicazione preliminare sulla forza del legame. D’altra parte è importante sottolineare che “correlazione non implica causalità” e anche una forte correlazione tra due variabili non implica che i valori dell’una siano effettivamente causa dei valori dell’altra.

3.1.3.2 Scatterplots

Uno scatterplot matrice è una griglia simmetrica quadratica di scatterplots bivariati. Ciascuna cella della griglia mostra uno scatterplot di due variabili. La variabile j è plottata contro la variabile i nella ij cella e le stesse variabili appaiono nella cella ji, con gli assi x ed y interscambiabili (Everitt and Hothorn, 2011). Gli scatterplots forniscono un sommario compatto delle informazioni sulle relazioni tra ciascuna coppia di variabili, permettendo di osservare simultaneamente i plots bivariati di più variabili. Esso risulta essere utile per studiare la natura delle relazioni tra ciascuna coppia di variabili e per diagnosticare la presenza di outliers e di clusters nei dati. Mentre la matrice di correlazione rappresenta solo la direzione e l’intensità della relazione lineare, lo scatterplot matrice aiuta l’analista a visualizzare possibili relazioni non lineari tra due variabili.

Occorre però aggiungere un'ulteriore considerazione: sebbene gli scatterplots forniscano un punto di vista iniziale delle potenziali

relazioni bivariate tra le varie coppie di variabili contenute nel campione di dati multivariato, essi non prendono in considerazione l'effetto di possibili interazioni tra le variabili (Ofungwu, 2011).

3.1.3.3 Regressione lineare semplice

Come precedentemente detto l'analisi di regressione lineare è stata utilizzata come primo approccio alla specificazione del modello: per ciascuna coppia di variabili sono stati elaborati tre modelli caratterizzati da diverse forme funzionali della variabile indipendente: lineare, polinomiale e logaritmica e confrontati attraverso il coefficiente di determinazione, il risultato del test-F, del test di Shapiro-Wilk e del test di Breush-Pagan.

E' generalmente noto che il coefficiente di determinazione R^2 è una misura della frazione della varianza in una variabile che è quantificata da un'altra. E' un indice del buon fitting della curva di regressione. Ma esso presenta i seguenti limiti: il suo valore incrementa sempre quando nuove variabili indipendenti sono aggiunte, anche se queste non forniscono un contributo significativo al modello. Così, poiché il coefficiente di determinazione aggiustato tiene conto del numero di variabili indipendenti nel modello (Montgomery and Peck, 1982), esso è stato considerato essere più utile per confrontare modelli caratterizzati da differenti forme funzionali.

Un altro approccio addizionale è stato usato per valutare il buon fitting. Esso è costituito dalla statistica test F. Essa è ottenuta dal rapporto tra la varianza spiegata dalla regressione e la varianza spiegata dai residui ed è una misura della significatività dell'intero modello.

Un'importante fase dell'analisi del buon fitting è rappresentata anche dall'esaminare sempre i residui di una regressione lineare per determinare se le assunzioni del modello fossero violate. Per ciascuna coppia di variabili, sono stati esaminati i grafici relativi ai residui contro i valori fittati e ciascuna variabile indipendente. E' generalmente noto che se i residui appaiono dispersi casualmente con nessun pattern, allora esiste l'evidenza di un buon adattamento della regressione stimata ai dati osservati.

Un plot dei residui può aiutare anche ad identificare eventuali outliers.

Esso, inoltre, può anche rivelare la mancanza di omogeneità della varianza degli errori ossia aiutare ad identificare l'eteroschedasticità. Se per esempio il plot residui contro i valori fittati assume una forma ad imbuto, con la parte terminale più piccola dell'imbuto disposta in corrispondenza dei valori predetti più piccoli e la parte più grande disposta in corrispondenza dei valori predetti maggiori, allora esso fornisce un'indicazione relativa all'incremento della varianza proporzionale ai valori della variabile risposta.

L'andamento dei grafici dei residui può fornire importanti informazioni; ma formali test statistici possono aiutare a confermare quanto dedotto dai grafici. Il test di Breush-Pagan proposto da Breusch e Pagan (1979) è, per esempio, consigliato per individuare eteroschedasticità. Il principale limite, però, di questo test è costituito dall'assunzione di normalità degli errori della regressione (Kaufman, 2013). Questa assunzione risulta essere troppo forte per il lavoro pratico. E' stata superata usando una versione studentizzata proposta da Koenker (Koenker, 1981).

E' noto che se il modello fittato è corretto, allora i residui dovrebbero avere approssimativamente una distribuzione normale. Ci sono numerosi test e plots diagnostici per valutare la normalità; il plot più comune è il Q-Q Plot; questo è un grafico che testa la conformità tra la distribuzione empirica e la data distribuzione teoretica. Un buon fitting della distribuzione teorica ai valori osservati dovrebbe essere indicato dal plot se i valori cadono sulla linea diagonale. In un Q-Q plot la presenza di curvature è indicativo di una distribuzione che si distacca da quella normale. La forma e la direzione di tali curvature è utile per diagnosticare la deviazione (Vittinghoff et al., 2005). Per esempio una curva concava verso l'alto evidenzia una dissimmetria a destra, mentre una curva concava verso il basso una dissimmetria a sinistra. Una curva a forma di S, invece, può essere interpretata come una distribuzione dei dati che è simmetrica ma con code più pesanti o leggere rispetto (dipendendo dalla direzione della S) ad una distribuzione normale.

Analisi di regressione lineare multipla

Sebbene la formulazione del modello di regressione lineare multipla sembra essere una semplice generalizzazione del modello di regressione lineare semplice, l'inclusione delle altre variabili indipendenti crea un nuovo concetto nell'interpretazione dei coefficienti di regressione; nel modello di regressione multipla, questi esprimono la variazione in media della variabile dipendente in corrispondenza di una variazione unitaria della relativa variabile

indipendente mantenendo costanti le altre variabili esplicative. Ciò risulta essere diverso dal computare differenti semplici regressioni lineari, cambiando di volta in volta una variabile, poiché ciascuna regressione ignora i possibili effetti delle altre variabili (Freund et al., 2006).

Il valori dei coefficienti di regressione non forniscono, però alcuna informazione sulla loro significatività nello spiegare la variazione nella variabile dipendente. Quindi, test di ipotesi diventano necessari per tale fine.

Nella regressione lineare multipla il t-test effettuato per ciascun coefficiente di regressione permette di valutare la significatività di ciascuno di essi e aiuta a decidere le variabili da lasciare o da eliminare dal modello. Nel modello lineare di regressione multipla, il test F permette di testare, invece, se tutti i coefficienti di regressione relativi a tutte le variabili nel modello di regressione siano simultaneamente uguali a zero.

Quindi, mentre il test-t è utile per valutare la significatività di ciascuna variabile indipendente, il test F è utile per individuare il contributo collettivo di tutte le variabili specificate nel modello.

La maggior parte dei software statistici riporta il valore del p-value come risultato dei test di ipotesi. Il p-value è definito da Levine et al. (2010) come la probabilità di ottenere un valore della statistica test uguale o più estremo del risultato campionario, nell'ipotesi che H_0 (ipotesi nulla) sia vera.

Adottando un coefficiente di significatività pari a 0,05, i coefficienti di regressione risultano essere significativi se i corrispondenti p-values sono minori o uguali a 0,05; le stesse considerazioni valgono per il p-value associato alla statistica-F.

Considerazioni analoghe a quelle fatte nel caso di regressione semplice valgono per il coefficiente di determinazione relativo alla regressione multipla. Esso misura la riduzione proporzionata nella variazione della variabile dipendente che è direttamente attribuibile al modello di regressione (Freund et al., 2006). Come nel caso di regressione lineare semplice, è stato preferito adottare il coefficiente di determinazione aggiustato.

Analoghe considerazioni svolte per la regressione lineare semplice valgono per i grafici dei residui e per l'interpretazione dei Q-Q Plot.

3.1.3.5. Il secondo approccio per la specificazione: “partial residual plots”

Nella fase di specificazione del modello sono stati utilizzati anche “partial residual plots” che sono degli utili strumenti di supporto ai grafici dei residui. Essi sono definiti da Gunst & Mason (1980) come:

$$r_i^* = Y_i - (\hat{Y}_i - \hat{\beta}_j X_{ij}) = r_i + \hat{\beta}_j X_{ij},$$

Le variabili sopra considerate sono tratte dal modello preliminare e completo di regressione lineare ai minimi quadrati; questo, a livello della *i*-esima unità statistica del campione considerato, è noto che si specifica come segue con *k* variabili indipendenti:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + e_i, i = 1, 2, \dots, n.$$

$\hat{Y}_i - \hat{\beta}_j X_{ij}$ rappresenta il valore stimato della *i*-esima variabile risposta usando tutti i predittori eccetto X_j ; conseguentemente, i residui parziali r_i^* per X_j si ottengono aggiungendo ai residui relativi ai minimi quadrati del modello completo r_i la componente lineare della relazione tra *Y* ed X_j .

L'idea base è che una componente non lineare non modellata della relazione tra *Y* e X_j dovrebbe apparire nei residui dei minimi quadrati; così il plot tra r_i^* contro X_j rivelerà la relazione parziale tra *Y* ed X_j e possibili scostamenti dalla linearità (Fox, 1991)

Il CCPR (component and component-plus-residual) plot è un affinamento del partial residual plot aggiungendo

$$\hat{\beta}_j X_{ij} \text{ contro } X_{ij}$$

Questo è una parte componente del plot e mostra dove la linea fittata giacerebbe.

3.1.3.6. Metodo Backward

Il metodo backward è una procedura step-by-step che inizia con il modello completo (con tutti i predittori *k*). Esso si basa su una ricerca iniziale che viene svolta per determinare se sono presenti delle variabili indipendenti non significative nel modello, osservando i *p*-value associati ai *t*-test di ciascuna variabile. Se nessun predittore non significativo (il cui *p*-value è più grande di una determinata soglia di significatività) è individuato, il processo backward

termina con il modello completo. Se, invece sono presenti predittori non significativi, quello con il più piccolo valore assoluto di t (o quello con il più grande p -value associato) è eliminato ed un nuovo modello è sviluppato con $k-1$ variabili indipendenti.

Il nuovo modello è nuovamente esaminato per determinare se esso contiene variabili indipendenti non significative. Se nuovamente esse sono presenti, quella con il più piccolo valore assoluto di t è eliminato ed un nuovo modello è sviluppato per lo step successivo. Questa procedura viene iterata finchè nel modello restano solo le variabili caratterizzate da t -values significativi.

3.1.4 Stima dei parametri con verifica delle assunzioni di base della regressione

Scelta la migliore forma funzionale e le variabili da includere in ciascuna regressione, i coefficienti di regressione stimati con il metodo dei minimi quadrati (in inglese Ordinary Least Squares, OLS) sono stati presi in considerazione. Senza dilungarsi sul metodo dei minimi quadrati per l'ampia letteratura in merito, occorre premettere che ciascuna regressione stimata appartiene ad un sistema di regressioni simultanee. La simultaneità sorge quando una o più variabili indipendenti è determinata congiuntamente con la variabile dipendente, attraverso un meccanismo di equilibrio. E' un'importante forma di endogeneità ossia uno o più variabili esplicative sono correlate con il termine di errore. Ciò determina che lo stimatore OLS sia distorto ed inconsistente.

Il ricorso alle regressioni simultanee è stato, però, necessario. Senza di esse infatti non si sarebbe potuto tener in conto della bidirezionalità delle relazioni tra i vari elementi biotici, fatto questo (causalità) assolutamente più importante della stessa casualità che contraddistingue le regressioni multivariate. La presenza di regressioni simultanee avrebbe richiesto di ricorrere a variabili strumentali o al metodo dei minimi quadrati generalizzati (GLS). Per le prime non si avevano ulteriori dati a disposizione per tutte le stazioni in grado di poter essere validamente usati. Per il secondo, tenuto conto della natura asintotica degli stimatori GLS, si è optato per la soluzione con il metodo delle successive iterazioni per convergenza, con un elementare controllo dei residui ottenuti di volta in volta.

Queste regressioni sono anche affette da multicollinearità. La presenza nei regressori dei due composti dell'azoto e dell'ossigeno disciolto è fonte, infatti, di multicollinearità a causa dei processi di nitrificazione. Tuttavia la continua alimentazione dall'esterno verso il corpo idrico

in modo separato delle tre sostanze (pressioni diffuse, pressioni puntuali, scambi acqua-aria, presenza di moti turbolenti nella corrente) è stata considerata prevalente sui processi di nitrificazione (anche se in una delle regressioni, come si vedrà in seguito, la procedura di specificazione ha fornito un miglior risultato considerando l'azoto totale). Sul problema della collinearità e della multicollinearità vedi oltre.

In ragione di tutto ciò, lo stimatore OLS è stato utilizzato in tale lavoro come stimatore preliminare da comparare con approcci e metodi statistici più avanzati negli sviluppi futuri di questo lavoro avendo a disposizione un maggior numero di stazioni di monitoraggio e un set di dati più ampio e strutturato in relazione ai processi bio-chimici che avvengono nei fiumi in presenza di immissioni di determinate sostanze inquinanti.

Poiché ciascuna regressione è stata risolta con gli OLS, è necessario valutare che le assunzioni di base siano rispettate, comprendendo le conseguenze di tali violazioni.

3.1.4.1. Omoschedasticità

Una delle principali assunzioni alla base della regressione è quella relativa alla costanza della varianza della distribuzione degli errori condizionata alla variabile indipendente ed in particolare della sua indipendenza dalle variabili esplicative. La violazione di tale assunzione è definita eteroschedasticità. Sebbene essa non causi distorsione o inconsistenza negli stimatori OLS, le varianze degli stimatori OLS, sono distorte. In altri termini gli stimatori OLS non sono più BLUE (Best Linear Unbiased Estimator) ossia tra tutti gli stimatori non distorti essi non forniscono più la stima con la più piccola varianza. Così, gli intervalli di confidenza e le statistiche t non sono più validi. Le usuali statistiche t non seguono la distribuzione t di Student in presenza di eteroschedasticità ed il problema non è risolto usando grandi campioni. La statistica F non segue più la distribuzione F di Fisher-Snedecor (Wooldridge, 2006).

Già nei paragrafi precedenti è stato accennato ai grafici e ai test utili per diagnosticare eteroschedasticità; nella regressione lineare multipla, un'idea della presenza di eteroschedasticità può, come già detto, essere testata da un plot dei residui in funzione delle variabili indipendenti; se la nuvola di punti è molto stretta nella parte iniziale e via via diventa più ampia, all'aumentare dei valori della variabile indipendente, allora il grafico sta suggerendo che la varianza condizionata degli errori data la variabile esplicativa sta crescendo; se la nuvola dei punti, invece assume una forma d'imbuto con la parte più ampia a sinistra,

allora il grafico suggerisce che la varianza condizionata degli errori data la variabile esplicativa si sta restringendo; ambedue i grafici riflettono eteroschedasticità.

Fra i test ampiamente diffusi per diagnosticare eteroschedasticità vi è il test di White e di Breush-Pagan. Il test di White si basa su una regressione ausiliaria relativa ai residui al quadrato (ottenuti dalla precedente regressione stimata) rispetto alle variabili indipendenti. Se non esiste eteroschedasticità, le variabili indipendenti non spiegano molto della variazione nei residui al quadrato. Se, invece, è presente eteroschedasticità, le variabili indipendenti spiegano una porzione significativa della variazione nei residui al quadrato. L'ipotesi nulla del test afferma che il termine d'errore al quadrato è incorrelato con le variabili indipendenti e si basa sull'assunzione di assenza di eteroschedasticità. Sotto l'ipotesi nulla è stato mostrato che nR^2 (dove n = dimensione campionaria e R^2 = coefficiente di determinazione al quadrato ottenuto dalla regressione ausiliaria) segue asintoticamente la distribuzione chi-quadrato con gradi di libertà pari al numero di regressori (non includendo il termine relativo alla costante) nel modello di regressione ausiliario. Il test si esegue comparando nR^2 con il valore critico della chi-quadrato. Se $nR^2 > \chi^2_{\alpha,5}$, allora c'è eteroschedasticità; se, invece $nR^2 < \chi^2_{\alpha,5}$, allora non c'è eteroschedasticità (Yan and Su, 2009). L'idea alla base del test di Breush-Pagan è la stessa del test di White; mentre quest'ultimo è generico e vale per ogni tipo di eteroschedasticità, il test di Breush-Pagan richiede la conoscenza sulla fonte di eteroschedasticità ossia sulle variabili che sono state riconosciute come essere responsabili della mancanza di costanza della varianza.

Svariate possono esserne le cause; sono di seguito elencate:

1. le unità campionarie provengono da sottopopolazioni diverse;
2. errori di misura;
3. uno o più regressori importanti sono stati omessi ovvero non compaiono nella forma più adatta; questo è un problema di non corretta specificazione del modello;
4. dissimmetria nella distribuzione di un regressore.

Esistono diversi metodi per correggere l'eteroschedasticità: errori standard consistenti all'eteroschedasticità e minimi quadrati pesati (in inglese weighted least squares, WLS) che è

una generalizzazione dei minimi quadrati generalizzati (in inglese Generalized Least Squares, GLS).

Il primo metodo si basa sempre sugli stimatori OLS (che sono ancora consistenti, non distorti ma inefficienti) e sul calcolo degli errori standard alternativi (robusti) che sono validi anche in presenza di eteroschedasticità e che permettono di costruire intervalli di confidenza e statistiche *t* robusti. Hayes & Cai (2007) mostrano le proprietà di ciascun stimatori degli errori standard consistente all'eteroschedasticità (in inglese HC: heteroskedasticity-consistent): HC0, HC1, HC2, HC3. In piccoli campioni, gli stimatori HC1, HC2, HC3 hanno proprietà di gran lunga superiori allo stimatore HC0. Ormai la maggior parte dei software statistici contengono delle funzioni per calcolarli rapidamente.

Il secondo metodo è una tecnica che pesa le osservazioni in modo proporzionale al reciproco della varianza degli errori relativa a quell'osservazione. Esso si basa sul principio di attribuire meno peso alle osservazioni con la più alta varianza d'errore.

Il problema principale nell'applicazione di tale metodo è la non facilità di conoscere ex ante il valore dei pesi (Nannicini, 2010). Poiché quindi spesso la forma dell'eteroschedasticità non è ovvia, si tende ad introdurre delle stime della funzione dei pesi. Questa trasformazione porta ad un nuovo stimatore definito feasible GLS (FGLS). In ogni caso, l'implementazione di un FGLS richiede sempre la conoscenza dettagliata della struttura dell'eteroschedasticità con il fine di proporre un modello trasformato omoschedastico. Un altro svantaggio di questo stimatore è che le sue proprietà campionarie finite sono sconosciute; è tuttora difficile, se non impossibile, derivare le proprietà campionarie finite come il valore atteso e la varianza di tale stimatore; occorre quindi affidarsi alle sue proprietà asintotiche per le inferenze statistiche (Kuan, 2004).

Poiché l'uso dello stimatore WLS dipende dalla conoscenza della struttura della varianza che raramente è nota, nella pratica è più comune usare gli errori standard consistenti all'eteroschedasticità che non richiedono assunzioni (Sosa, 2009).

3.1.4.2. Normalità della distribuzione degli errori

Un'altra assunzione fondamentale è la normalità del termine di errore che è equivalente a dire che la variabile risposta ha una distribuzione normale condizionata alla variabile indipendente (ossia per ciascuna combinazione dei valori delle variabili esplicative).

La non normalità non produce distorsione nei coefficienti stimati, né influisce sulla conclusione che lo stimatore OLS sia il miglior stimatore lineare non distorto sotto le assunzioni di Gauss-Markov. Ma l'inferenza esatta basata sulle statistiche t ed F richiede tale ipotesi e la violazione di tale assunzione può portare a p-value relativi al test F e ai test t non corretti e ad intervalli di confidenza non validi. Fortunatamente, in grandi campioni, nell'ipotesi di violazione di normalità della distribuzione degli errori, gli stimatori OLS seguono ancora (asintoticamente) una distribuzione normale per il teorema del limite centrale; i p-value relativi ai test di significatività ed i limiti degli intervalli di confidenza sono approssimativamente corretti dal momento che la loro derivazione si fonda sulla normalità degli stimatori OLS (Wooldridge, 2006). Quindi è possibile affermare che in grandi campioni, la violazione dell'assunzione di normalità non causa problemi.

La non normalità può essere rilevata dall'analisi dei grafici dei residui, istogrammi, Q-Q Plot dei residui e da formali test statistici.

Un primo approccio è costituito dagli istogrammi che possono aiutare ad individuare dissimmetrie, curtosi ed outlier.

La curtosi è la caratteristica della forma della distribuzione che misura lo spessore delle code di una distribuzione: il maggiore/minore spessore delle code della distribuzione determina un maggiore/minore appiattimento della forma della distribuzione.

Si dice, invece asimmetrica una distribuzione la cui forma non si presenta speculare rispetto alla posizione centrale. Si parla di:

- Asimmetria positiva quando la forma è caratterizzata da una coda allungata verso destra
- Asimmetria negativa quando la forma è caratterizzata da una coda allungata verso sinistra

Un outlier può essere definito come un'osservazione che appare essere inconsistente con i restanti dati. Un ulteriore grafico utilizzato per effettuare l'analisi sull'ipotesi di normalità è il Q-Q Plot. Questo grafico è ottenuto plottando i quantili della distribuzione empirica contro quelli teorici. Come già detto, deviazioni dalla diagonale a forma di S possono essere interpretate come una distribuzione di dati simmetrica ma con code più pesanti o più leggere (dipendendo dalla direzione della S). Come già evidenziato prima, la pesantezza delle code di una distribuzione in relazione all'addensamento intorno al valore centrale è definita come

curtosi. Invece, deviazioni dalla diagonale come una curva concava, può essere interpretata come una distribuzione di dati proveniente da una distribuzione asimmetrica. La direzione della dissimmetria dipende dalla direzione della curvatura sul plot: una curvatura verso l'alto è indicativa di una dissimmetria a destra; verso il basso di una dissimmetria a sinistra. Anche, il Q-Q Plot può aiutare a diagnosticare outlier.

Tra i test molto utilizzati vi è quello di Shapiro-Wilk. Esso si basa sull'ipotesi nulla affermando che il campione provenga da una distribuzione normale; precisamente, esso si fonda sul fatto che il coefficiente di correlazione del plot di probabilità normale è una misura della forza della relazione lineare tra i quantili empirici e normali. Se i dati seguono una distribuzione normale, allora il coefficiente di correlazione tende ad 1 mentre l'ipotesi nulla di normalità è rigettata se esso assume bassi valori.

Quando l'assunzione di normalità non è soddisfatta, la trasformazione della variabile dipendente è indicata come un possibile rimedio. Ma essa ha una priorità minore rispetto alle trasformazioni utili per semplificare le relazioni o per stabilizzare la varianza. Occorre ricordare che spesso, però, le trasformazioni per correggere eteroschedasticità hanno come effetto quello di migliorare anche la normalità (Rawlings et al., 1998). Un approccio molto diffuso è il metodo computazionale di Box and Cox (1964). Esso si basa sul determinare le stime dei parametri di una trasformazione di potenza per la variabile dipendente per individuare quale trasformazione dei dati può meglio normalizzare la loro distribuzione. E' noto che i valori di tali parametri coincidono con quelli che rendono massima la funzione di log-verosomiglianza dei dati. Secondo Jaccard et al. (1990), però, le trasformazioni cambiano l'unità fondamentale delle variabili al punto che la nuova metrica non ha più un reale significato. Jaccard et al. consigliano, allora, di usare un modello statistico che è più appropriato ai dati, piuttosto che manipolare i dati per conformarli al modello statistico di base. Altri metodi per approcciare alla non normalità sono costituiti, infatti, dagli errori standard robusti e dagli intervalli di confidenza robusti.

3.1.4.3. Collinearità

La collinearità rappresenta una violazione alle assunzioni classiche del modello di regressione. Si distingue tra collinearità perfetta ed imperfetta. La prima si realizza quando uno dei regressori è una funzione lineare esatta degli altri regressori. Tecnicamente, non è possibile stimare i parametri della regressione con gli OLS; In un caso del genere, infatti, lo stimatore

dei minimi quadrati $\hat{\beta} = (X'X)^{-1}X'Y$ (dove X = matrice dei regressori e Y = la variabile dipendente) non dà risultati attendibili. Se c'è una correlazione perfetta tra le due variabili, la matrice $(X'X)$ diventa singolare ha determinante uguale a 0 e perciò non esiste la matrice inversa. Non è possibile, praticamente, separare l'influenza su Y dei regressori che sono perfettamente correlati. Nelle applicazioni pratiche, è raro che si verifichi questa eventualità. Berry & Feldman (1985) affermano che nel caso di polinomiali, poiché le variabili non sono linearmente relazionate, l'analista non dovrebbe preoccuparsi della perfetta collinearità.

La multicollinearità imperfetta (quasi Multicollinearità) sorge quando c'è un'elevata correlazione tra due o più variabili indipendenti. Gli stimatori OLS possono ancora essere utilizzati e sono ancora BLUE. Le varianze e gli errori standard delle stime OLS dei coefficienti di regressione relativi alle variabili collineari cresceranno all'aumentare della correlazione tra le variabili indipendenti, inficeranno il valore della statistica test t dei coefficienti stimati facendoli decrescere; aumenterà la probabilità che l'ipotesi di non significatività dei coefficienti venga accettata. Questo può portare ad eliminare in modo errato una variabile che influenzerebbe la variabile dipendente. Gli intervalli di confidenza tenderanno ad essere più ampi. Un semplice modo per diagnosticare multicollinearità è l'uso dei variance inflativo factor, VIF, calcolato per ciascun regressore. L'indice VIF relativo a ciascun i -iesimo regressore è definito come

$$VIF_i = \frac{1}{(1 - R_i^2)}$$

In cui R_i^2 è il coefficiente di determinazione ottenuto da una regressione del predittore i esimo contro i restanti. Se R_i^2 tende ad 1, l' i esimo predittore è ben spiegato da una funzione lineare dei restanti e la sua presenza può considerarsi ridondante. Per valori dei VIF che eccedono 5, i regressori sono ritenuti collineari.

Graham (2003) menziona le diverse tecniche per correggere gli effetti della multicollinearità tra cui l'eliminazione di una variabile che è collineare, regressione sequenziale, le componenti principali della regressione, SEM (in inglese: Structural equation modeling). Graham (2003) conclude, in riferimento ai metodi per correggere la collinearità affermando che: *"Still, most of these procedures only help to stabilize the statistical analyses, making them less biased, less subjective, and more repeatable, but only the statistical collinearity will have been removed"*

from the data. The explanatory variables are still, by nature and in nature, correlated, whether or not functionally”.

3.1.5. Calcolo incertezza attraverso la stima degli intervalli di confidenza

Come già evidenziato precedentemente, nei modelli di regressione, l'incertezza strutturale è associata alla possibilità di una specificazione errata. Le tecniche successivamente descritte per il calcolo incertezza non prendono in considerazione l'errore introdotto dalla scelta di un modello di regressione inappropriato.

Quella che viene quantificata attraverso l'analisi di regressione lineare, di solito, è dovuta solo a due fonti:

1. Incertezza associata al termine di errore che può essere stimato dall'errore standard delle stime
2. Incertezza associata alle stime parametriche

E' noto che l'errore standard delle stime s_ε per n osservazioni con k variabili esplicative è calcolato in base alla formula successiva dai residui $(y_i - \hat{y}_i)$ in cui $\hat{y}_i$ è il valore stimato della variabile dipendente y relativo all'osservazione i -esima ed y_i è il valore osservato della variabile dipendente y relativo all'osservazione i -esima.

$$s_\varepsilon = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - k - 1}}$$

s_ε è utile per costruire gli intervalli di confidenza per la media della variabile dipendente.

L'incertezza, invece, associata alle stime parametriche è valutata attraverso gli intervalli di confidenza per ciascun parametro della regressione. Essi, per il generico parametro β_k sono così definiti :

$$b_k - t_{\frac{\alpha}{2}} S(b_k) \leq \beta_k \leq b_k + t_{\frac{\alpha}{2}} S(b_k)$$

in cui:

b_k = rappresenta il coefficiente di regressione campionario usato come stimatore del corrispondente parametro della popolazione β_k

$t_{\frac{\alpha}{2}}$ = valore critico della statistica t. Infatti, per la costruzione dell'intervallo di confidenza si utilizza uno stimatore che si distribuisce come una v.c t di Student con n-p-1 gradi di libertà (in cui n=numero di osservazioni e p=numero di variabili esplicative).
 $S(b_k)$ = errore standard del coefficiente di regressione b_k .

3.2.6. Risoluzione e rappresentazione delle regressioni

La risoluzione delle regressioni stimate è stata svolta in modo iterativo usando un metodo che convergeva alla soluzione attraverso successive simulazioni con controllo dei residui. I risultati sono stati riportati con una stima puntuale di ciascun RQE per ciascuna stazione, insieme con l'associato intervallo di confidenza e con il gap espresso in percentuale tra i valori osservati e simulati. Per meglio coglierne la specificità, è stato ritenuto opportuno studiare le interrelazioni esistenti in termini spaziali: i risultati sono stati analizzati e visualizzati su mappa in ambiente GIS.

4. Risultati

Il capitolo è stato organizzato rispettivamente in tre paragrafi intitolati Analisi Preliminare, Specificazione del modello e Stima dei parametri con verifica delle assunzioni.

4.1. Analisi preliminare

Il paragrafo comprende due sottoparagrafi intitolati rispettivamente Matrice di correlazione e Scatterplots.

4.1.1. Matrice di correlazione

La tabella 4.1 presenta la matrice di correlazione per le variabili considerate nella prima e seconda regressione. Essa mostra che sebbene le variabili N-NH₄, N-NO₃, PT ed OD siano ciascuna correlate con gli indici biologici rappresentati da RQE_ICMi e RQE_IBMR, alcune sono anche correlate tra di loro come N-NH₄ ed N-NO₃ con il PT e, meno, N-NH₄ con N-NO₃ (come visto sopra). Questa interconnessione è espressa da un coefficiente di correlazione positivo significando che ad alti valori di azoto corrispondono alti valori di fosforo ed alte concentrazioni di N-NO₃ corrispondono alte concentrazioni di N-NH₄. Essa evidenzia anche una interconnessione positiva tra RQE_ICMi e RQE_IBMR. Sono evidenziati in rosso, i valori dei coefficienti di correlazione più alti.

	RQE_IBMR	RQE_ICMi	N.NH4	N.NO3	PT	OD
RQE_IBMR	1,0000	0,4847	-0,2735	-0,3158	-0,3333	0,3474
RQE_ICMi	0,4847	1,0000	-0,4252	-0,3875	-0,3785	0,3568
N.NH4	-0,2735	-0,4252	1,0000	0,4990	0,6055	-0,2571
N.NO3	-0,3158	-0,3875	0,4990	1,0000	0,6407	-0,2503
PT	-0,3333	-0,3785	0,6055	0,6407	1,0000	-0,2384
OD	0,3474	0,3568	-0,2571	-0,2503	-0,2384	1,0000

Tabella 4.1. Matrice di correlazione inerente alla prima e seconda regressione. I valori dei coefficienti di correlazione più alti sono evidenziati in rosso.

La tabella 4.2. Invece fornisce la matrice di correlazione relativa alla terza e quarta regressione che mostra un'accentuata correlazione del fosforo con i nitrati mentre una minore con l'azoto ammoniacale. Una discreta correlazione dell'ISECI e del RQE-STAR_ICMi tra loro e con i nutrienti è da attendersi.

	RQE_ STAR_ICMi	RQE_IBMR	RQE_ICMi	N-NH4	N-NO3	PT	OD	ISECI
RQE_STAR_ICMi	1,0000	0,5126	0,3824	-0,5920	-0,5953	-0,5087	0,3355	0,6042
RQE_IBMR	0,5126	1,0000	0,4186	-0,2289	-0,3309	-0,2666	0,2687	0,4280
RQE_ICMi	0,3824	0,4186	1,0000	-0,3824	-0,3654	-0,3620	0,3625	0,1647
N-NH4	-0,5920	-0,2289	-0,3824	1,0000	0,3896	0,5731	-0,2351	-0,3298
N-NO3	-0,5953	-0,3309	-0,3654	0,3896	1,0000	0,7448	-0,2235	-0,3442
PT	-0,5087	-0,2666	-0,3620	0,5731	0,7448	1,0000	-0,1838	-0,3032
OD	0,3355	0,2687	0,3625	-0,2351	-0,2235	-0,1838	1,0000	0,1485
ISECI	0,6042	0,4280	0,1647	-0,3298	-0,3442	-0,3032	0,1485	1,0000

Tabella 4.2. Matrice di correlazione inerente alla terza e quarta regressione. I valori dei coefficienti di correlazione più alti sono evidenziati in rosso.

4.1.2. Scatterplots

Gli scatterplots relativi alla prima e seconda regressione (Fig 4.1) suggerirebbero un pattern lineare crescente tra RQE-IBMR e RQE-ICMi e curvilineo decrescente per le seguenti coppie: RQE-IBMR ed N-NH₄, RQE-IBMR e N-NO₃, RQE-IBMR e PT, RQE-ICMi ed N-NH₄, RQE-ICMi ed N-NO₃, RQE-ICMi ed PT e curvilineo crescente per RQE-IBMR ed OD, per RQE-ICMi ed OD. In effetti gli scatter-plots tra ciascun RQE e parametri chimici come l'azoto ed il fosforo evidenzierebbero soprattutto un effetto esponenziale. L'analisi dei residui, però, successivamente svolta ha permesso all'analista di focalizzare l'attenzione sul forte sbilanciamento dei grafici rispetto all'asse delle x e sulla considerazione che anche una trasformazione lineare-logaritmica potrebbe approssimare bene i dati osservati conferendo ai residui un pattern casuale.

E' evidente che anche gli scatterplot restanti mostrano delle relazioni: quelle più rilevanti sono tra OD ed N-NO₃, e PT e N-NO₃ e PT ed N-NH₄. Quest'ultimi due mostrano un addensamento di punti nel lato inferiore del diagramma che dovrebbe essere investigato con delle relazioni logaritmiche. Lo scatter plot relativo all'OD e N-NO₃-evidenzia un andamento decrescente; del resto la reazione di nitrificazione avviene a consumo di ossigeno, con ossidazione dell'ammoniaca a nitrati. Tali relazioni non sono state prese in considerazione nello sviluppo

delle regressioni. Come già precedentemente indicato, nel campo dei dati ambientali, la collinearità è intrinseca.

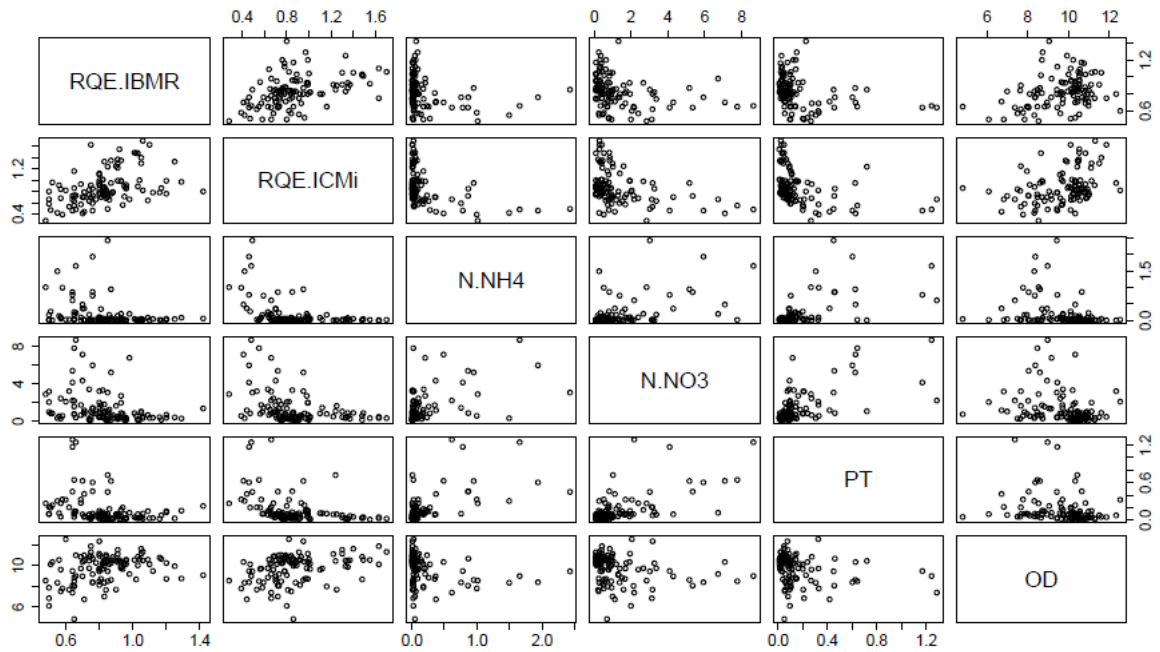


Figura 4.1. Scatter-plots relativi alla prima e seconda regressione

Il secondo scatter-plot (Fig 4.2) è stato costruito per analizzare soprattutto la terza e quarta regressione. Per la regressione relativa al RQE-STAR_ICMi, gli scatter-plots suggerirebbero un andamento curvilineo crescente con RQE-IBMR e RQE-ICMi, un andamento curvilineo decrescente con N-NH₄, N-NO₃, approssimativamente crescente lineare con l'OD e l'ISECI, lineare decrescente con il PT. Per quanto riguarda il RQE-STAR_ICMi, anche in tal caso è evidente un effetto esponenziale con le variabili costituite dall'azoto ma a differenza del RQE-IBMR e RQE-ICMi una minore incertezza nei dati. L'ISECI evidenzia una mancanza di correlazione con le variabili costituite dall'azoto e fosforo e RQE-ICMi, mentre è possibile individuare un pattern lineare crescente con RQE-STAR-ICMi, RQE-IBMR e OD.

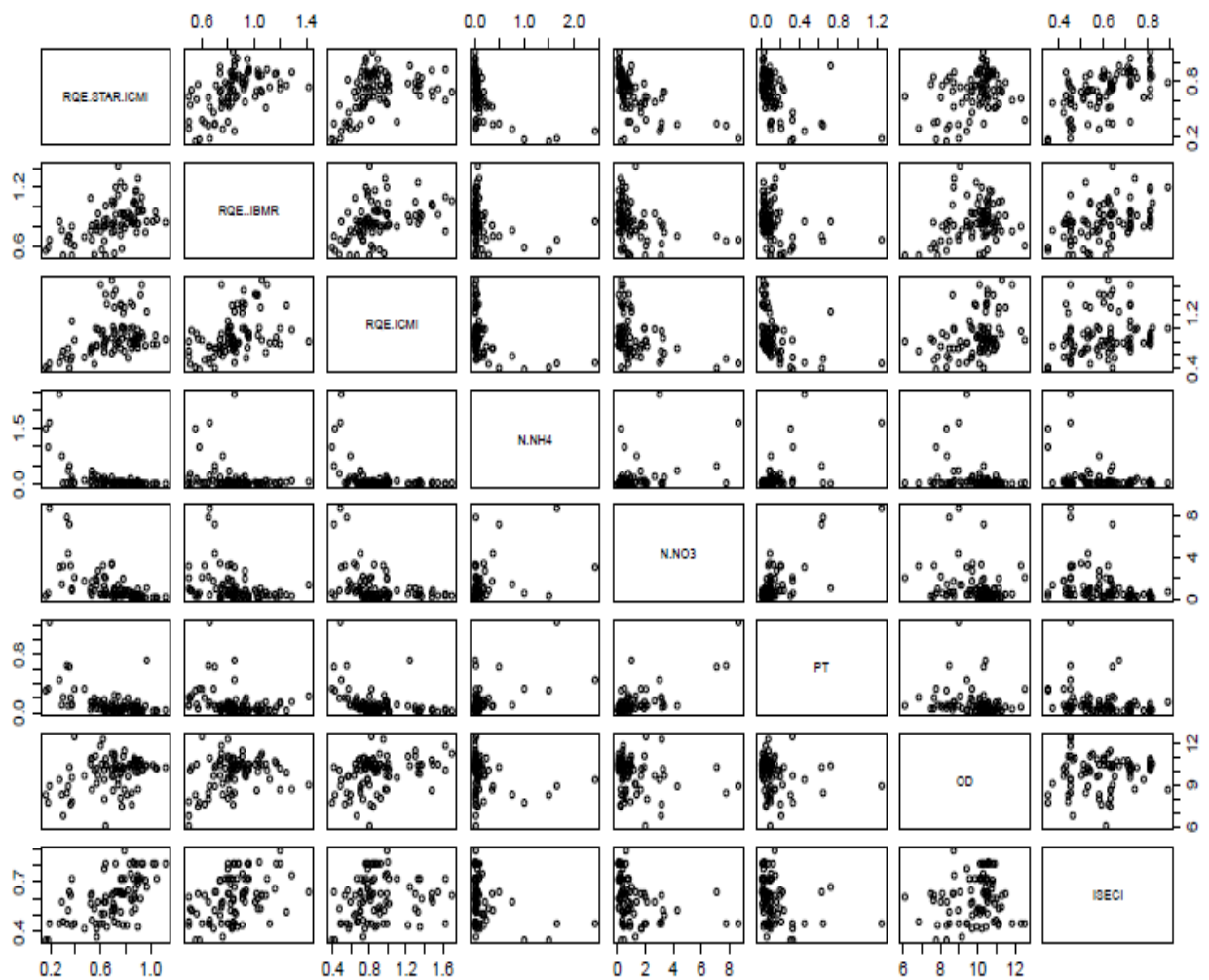


Figura 4.2. Scatter-plots relativi alla terza e quarta regressione

4.2. Specificazione del modello

La specificazione del modello, come già evidenziato, consiste nella scelta delle variabili da inserire nel modello e nella scelta di un'appropriata forma funzionale. Il paragrafo è stato organizzato in due sottoparagrafi principali: Analisi di regressione semplice e Regressione lineare multipla.

4.2.1. Analisi di regressione semplice

I valori del coefficiente di determinazione aggiustato, i risultati del test statistico F per la significatività globale del modello, del test di Shapiro-Wilk per la verifica della normalità della distribuzione degli errori, del test di Breusch-Pagan per la verifica dell'ipotesi di omoschedasticità sono stati riportati per ciascuna coppia di variabili in tabelle, per effettuare il

confronto tra le forme funzionali scelte. Tale confronto risulta utile per valutare la forma funzionale che consente il miglior adattamento della regressione ai dati osservati e che è stata evidenziata in giallo. Oltre ai formali test statistici, è stata effettuata anche un'analisi grafica che molto spesso consente di rilevare eventuali informazioni non evidenziate dall'analisi numerica. Quando i test non sono statisticamente significativi, sono indicati in rosso.

I risultati sono riportati per ciascuna variabile dipendente rappresentata dai RQE.

4.2.1.1. RQE-IBMR

Per le coppie di variabili aventi come variabile dipendente il RQE-IBMR (da tabella 4.3. a tabella 4.7., da figura 4.3. a 4.17.), la migliore trasformazione sembra essere la logaritmica, tranne per l'OD. Per tale coppia, è stata ritenuta essere più rappresentativa una trasformazione polinomiale (tabella 4.7., figura 4.17.).

I grafici dei residui relativi alla regressione lineare tra il RQE-IBMR e concentrazioni di nutrienti sono sbilanciati rispetto all'asse delle x (figura 4.6., figura 4.9., figura 4.12.); ciò non significa che il modello lineare sia errato. Ma è evidente che una trasformazione logaritmica della variabile indipendente in tal caso, fa in modo che i residui si distribuiscono in maniera casuale in ciascun plot considerato, fatta eccezione per un punto che è disposto troppo lontano dalla linea di riferimento orizzontale in corrispondenza dello zero e che è stato ritenuto essere un outlier (figura 4.7., figura 4.10., figura 4.13.).

Nessun plot dei residui evidenzia particolari pattern fatta eccezione per la singola osservazione sopra menzionata la cui rappresentazione non è consistente con quella delle altre 107 osservazioni.

Sebbene neanche nel plot dei residui relativo alla regressione lineare dell'OD con il RQE-IBMR, sia osservabile qualche pattern indicativo di una relazione non lineare (figura 4.17.), la trasformazione polinomiale è stata ritenuta essere più rappresentativa.

Nessun grafico dei residui relativo al RQE-IBMR evidenzia la violazione dell'assunzione di omoschedasticità (da figura 4.3. a 4.17.): non è possibile individuare alcun pattern ad imbuto che indicherebbe la mancanza di costanza nella distribuzione della varianza dei residui. Ciò è confermato dal risultato del test di Breush-Pagan.

Tutti i Q-Q Plots (da figura 4.3. a 4.17.) mostrano la presenza di un'osservazione anomala corrispondente alla stazione numero 66; essa sembra non essere consistente con la distribuzione dei restanti dati ed è disposta lontano dagli altri punti nella coda destra. Ma occorre essere cauti nel considerare un punto come un outlier; in termini statistici, un test potrebbe identificare un valore come outlier, ma potrebbe anche esserci nessuna ragione fisica attribuibile a tale comportamento. Per questo motivo ulteriori analisi sono state successivamente svolte (vedi dopo).

I Q-Q Plots non mostrano importanti deviazioni dalla normalità; però soprattutto i Q-Q Plots relativi alla relazione tra RQE-IBMR e RQE-ICMi (da figura 4.3 a 4.5) suggeriscono una piccola dissimmetria a destra, essendo caratterizzati da una lieve concavità verso l'alto con una coda sinistra estremamente compatta ed una coda destra diffusa dovunque. Mentre quelli relativi alla relazione tra RQE-IBMR e OD suggeriscono una coda pesante a destra (da figura 4.15 a 4.17).

4.2.1.2. RQE-ICMi

Per le coppie di variabili aventi come variabile dipendente il RQE-ICMi (da tabella 4.8 a 4.12, da figura 4.18 a 4.32), la migliore trasformazione sembra essere la logaritmica per tutte le variabili. Riferendoci alla relazione con il RQE-IBMR (Tabella 4.9, da figura 4.21 a figura 4.23), infatti, sebbene la trasformazione polinomiale sembrerebbe non essere affetta da eteroschedasticità ed avere il valore più alto di R^2 , essa è caratterizzata dal valore più basso della statistica F rispetto alle altre (tabella 4.9.) ed è stata preferita allora una trasformazione logaritmica.

In ciascuna coppia di variabili, la normalità non è mai rispettata. Tutti i Q-Q Plots aventi come variabile dipendente RQE-ICMi mostrano lo stesso pattern: ambedue le code sono disposte sopra la diagonale, e il Q-Q plot assume una forma concava. Questo è indicativo di una forte dissimmetria a destra (da figura 4.18. a figura 4.32.).

Anche in questo caso come per il RQE-IBMR, i grafici dei residui relativi alla regressione lineare tra il RQE-ICMi e concentrazioni di nutrienti sono sbilanciati rispetto all'asse delle x (figura 4.18., figura 4.24., figura 4.27.) e risulta evidente che una trasformazione logaritmica della variabile indipendente (figura 4.19., figura 4.25., figura 4.28.), fa in modo che i residui si distribuiscono in maniera casuale in ciascun plot considerato.

4.2.1.3. RQE-STAR-ICMi

Anche per il RQE-STAR-ICMi, la trasformazione lineare-logaritmica sembrerebbe conferire ai grafici dei residui contro l'N-NH₄, N-NO₃ e PT un pattern casuale (figura 4.46., figura 4.49., figura 4.52.). Per l'ISECI e l'OD, una trasformazione logaritmica non risulta essere necessaria, poiché già una regressione lineare sembrerebbe approssimare bene i dati (tabella 4.15., tabella 4.16.). Sebbene lo scatterplot contro il RQE-IBMR suggeriva un possibile effetto quadratico (figura 4.2.), dal confronto con le altre forme funzionali, anche la lineare sembrerebbe approssimare bene i dati, essendo caratterizzata dal valore più alto del risultato del test-F e dalla soddisfazione delle assunzioni di base della regressione (tabella 4.13.). Inoltre, dal grafico dei residui relativo alla regressione lineare contro il RQE-IBMR, non è possibile discernere un pattern curvilineo (figura 4.33.). Comunque, occorre evidenziare che la trasformazione polinomiale, è caratterizzata, invece, dal valore più alto di R^2 ma anche dalla violazione dell'assunzione di omoschedasticità (tabella 4.13). Queste considerazioni erano deducibili anche dall'osservazione dello scatter-plot contro RQE-IBMR, precedentemente rappresentato (figura 4.2.). Esso mostra infatti un possibile effetto quadratico ma anche una distribuzione eteroschedastica. Invece, per quanto riguarda, la regressione contro il RQE-ICMi, la migliore trasformazione sembrerebbe essere la polinomiale, essendo caratterizzata dai valori più alti di R^2 , statistica F e dalla soddisfazione delle ipotesi di base della regressione (tabella 4.14.); occorre, però, evidenziare che il grafico dei residui della regressione lineare contro il RQE-ICMi suggerirebbe la presenza di due clusters, più che di un effetto quadratico (figura 4.36.).

Nei Q-Q Plots relativi alla relazione tra RQE-STAR-ICMi e RQE-ICMi, tra RQE-STAR-ICMi e RQE-IBMR, i punti che tendono ad allinearsi lungo la bisettrice del grafico suggeriscono che i residui seguono una distribuzione normale (da figura 4.33 a 4.38); i risultati del test di Shapiro-Wilk confermano ciò, essendo statisticamente significativi (tabella 4.13. e tabella 4.14.); la maggior parte dei Q-Q Plots relativi alle restanti relazioni, invece mostrano una curva concava verso il basso suggerendo una lieve dissimmetria a sinistra (da figura 4.39 a figura 4.53).

4.2.1.4. ISECI

Le relazioni lineari visualizzate negli scatterplots tra ISECI e RQE-STAR-ICMi, ISECI e RQE-IBMR sembrano approssimare bene i dati, come mostrano i risultati ottenuti dallo

sviluppo delle regressioni lineari semplici (tabella 4.20., tabella 4.21., figura 4.54., figura 4.57.). Invece, le regressioni lineari tra ISECI e RQE-ICMi, ISECI ed OD non sono statisticamente significative e sono caratterizzate da bassi valori di R^2 (tabella 4.22, tabella 4.25.). Per queste due relazioni, la trasformazione polinomiale è significativa, sebbene al limite. Questa mancanza di significatività, soprattutto nella regressione tra ISECI ed OD, potrebbe dipendere dalla presenza dell'intercetta (tabella 4.25.). Anche per l'ISECI, la trasformazione lineare-logaritmica sembrerebbe conferire ai grafici dei residui contro l' $N-NH_4$, $N-NO_3$ e PT un pattern casuale (figura 4.64., figura 4.67., figura 4.73.). Nei Q-Q Plots contro RQE-STAR-ICMi, i punti che tendono ad allinearsi lungo la bisettrice del grafico suggeriscono che i residui seguono una distribuzione normale (da figura 4.54. a figura 4.56.); i risultati del test di Shapiro-Wilk confermano ciò, essendo statisticamente significativi (tabella 4.20.); la maggior parte dei Q-Q Plots relativi alle restanti relazioni (da figura 4.57. a figura 4.74.), invece, mostrano una S capovolta; questo pattern è indicativo di code leggere ossia di code più brevi rispetto ad una distribuzione normale.

	lineare	logaritmica	polinomiale
R2 corretto	0,2068	0,2257	0,2179
Statistica F	27,47	32,2	15,9
p_value_F	8,19E-07	1,22E-07	9,28E-07
p_value_Shapiro	0,00073	0,00108	0,003
p_value_Breush_Pagan	0,7458	0,8039	0,2075

Tabella 4.3. Confronto dei risultati relativi alla regressione lineare, polinomiale e lineare-logaritmica con il fine di esaminare la relazione tra RQE-IBMR e RQE-ICMi. In giallo la trasformazione che inizialmente è stata considerata approssimare bene i dati.

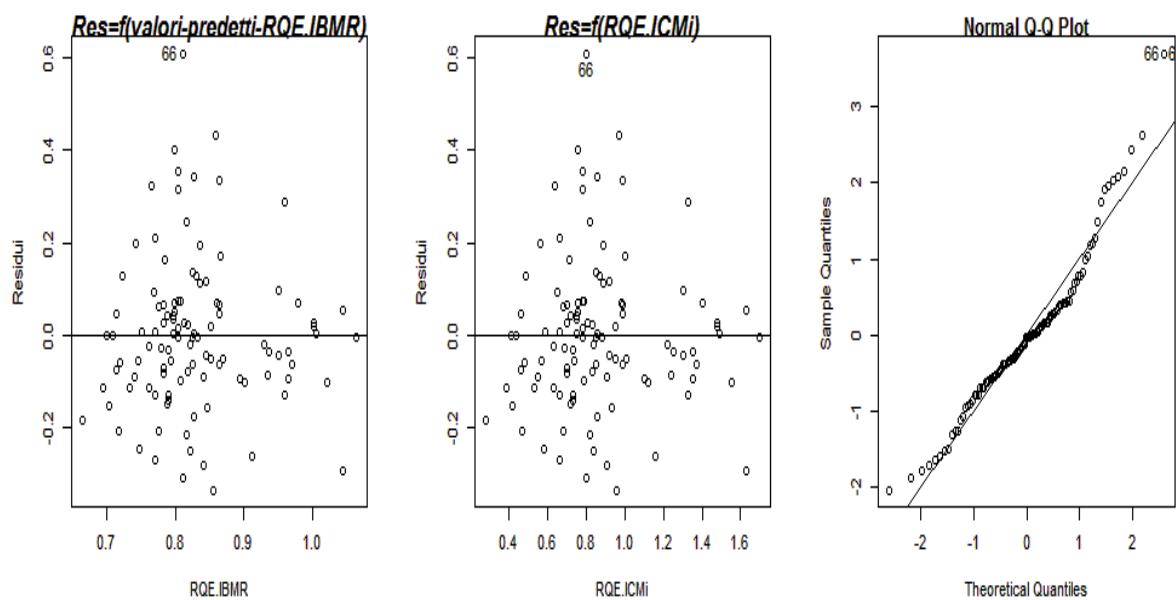


Figura 4.3. Grafici residui e Normal Q-Q Plot relativi alla regressione-lineare tra RQE-IBMR e RQE-ICMi

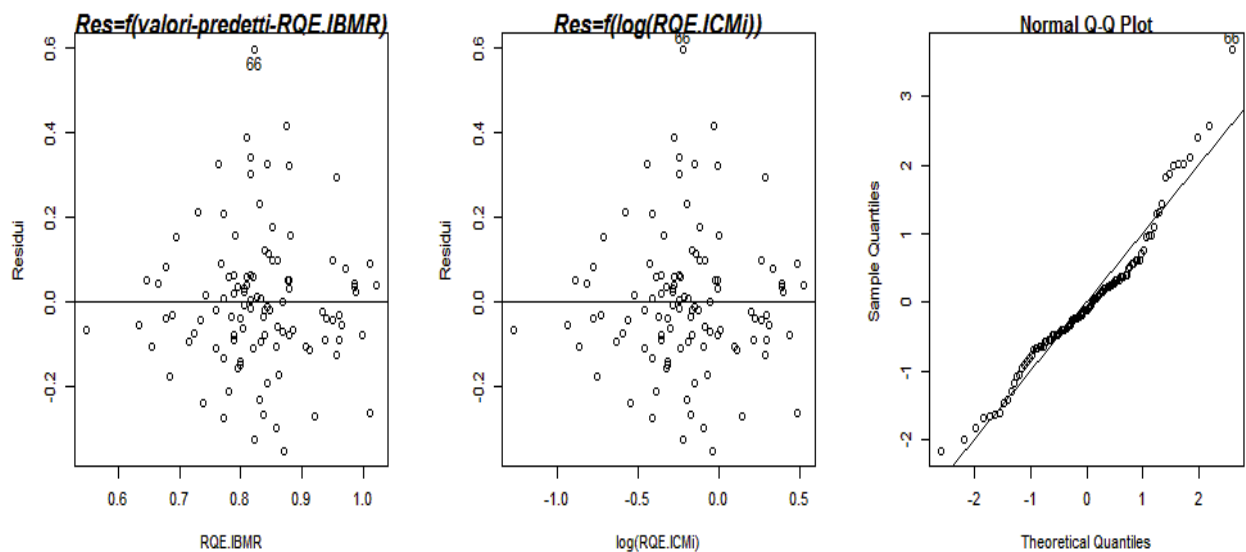


Figura 4.4. Grafici residui e Normal Q-Q Plot relativi alla regressione lineare-logaritmica tra RQE-IBMR e RQE-ICMi

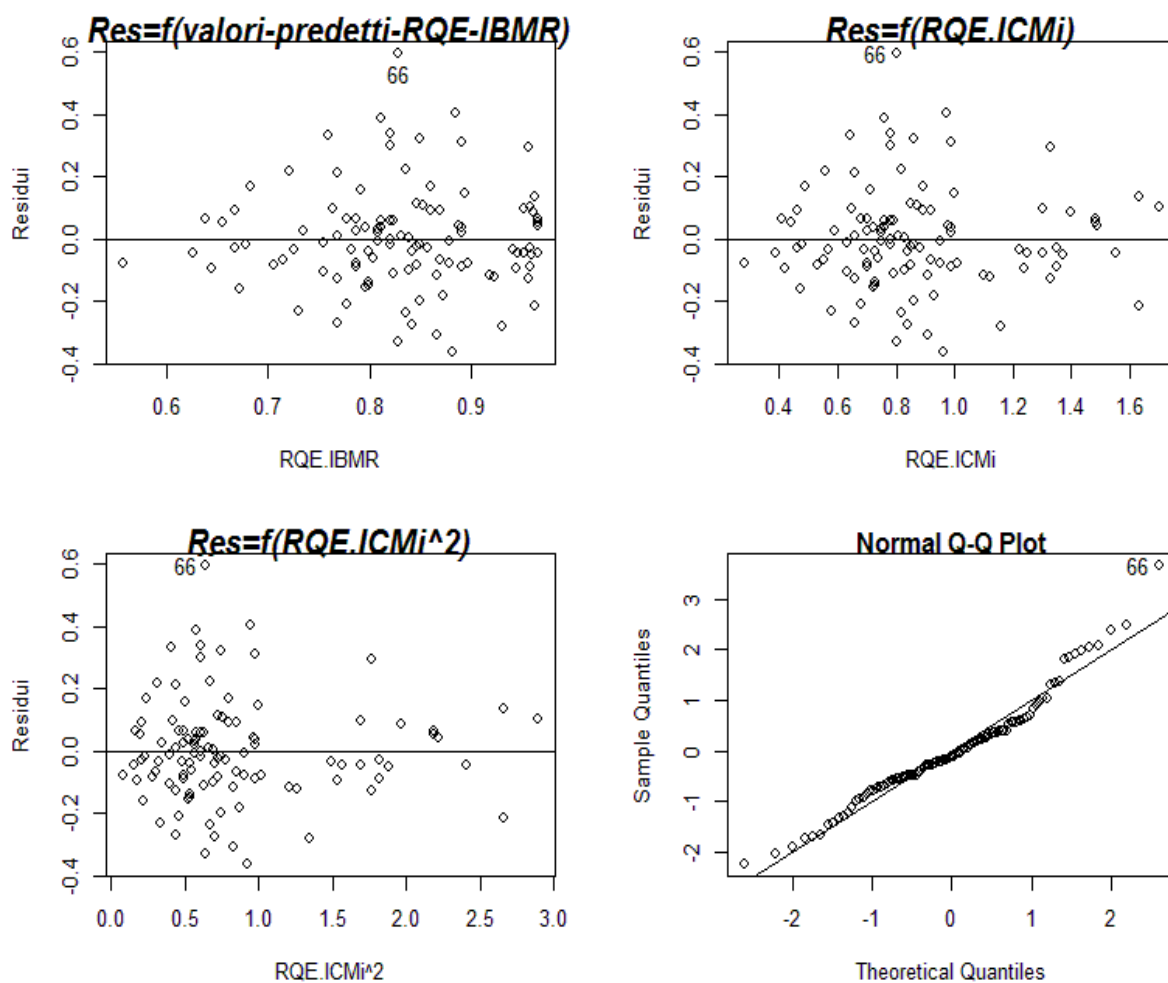


Figura 4.5. Grafici residui e Normal Q-Q Plot per regressione polinomiale di secondo grado tra RQE-IBMR e RQE-ICMi

	lineare	logaritmica	polinomiale
R2 corretto	0,063	0,086	0,124
Statistica F	8,283	11,000	8,544
p_value_F	4,80E-03	1,20E-03	3,70E-04
p_value_Shapiro	0,13	0,04	0,05
p_value_Breush_Pagan	0,869	0,962	0,324

Tabella 4.4. Confronto dei risultati relativi alla regressione lineare, polinomiale e lineare-logaritmica con il fine di esaminare la relazione tra RQE-IBMR e $N-NH_4$. In giallo, la trasformazione che inizialmente è stata considerata approssimare bene i dati.

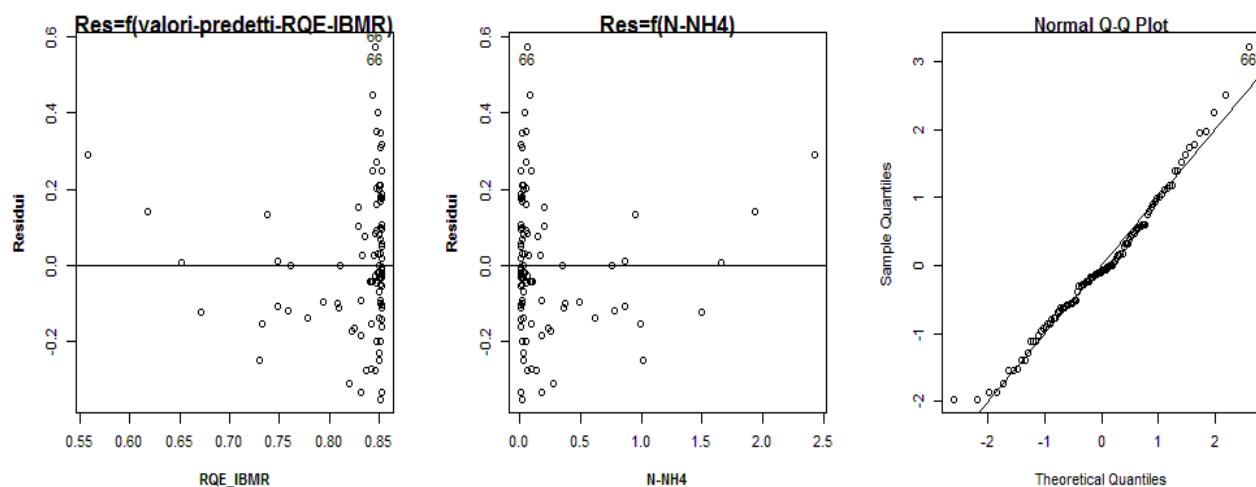


Figura 4.6. Grafici residui e Normal Q-Q Plot per regressione-lineare tra RQE-IBMR e N-NH₄

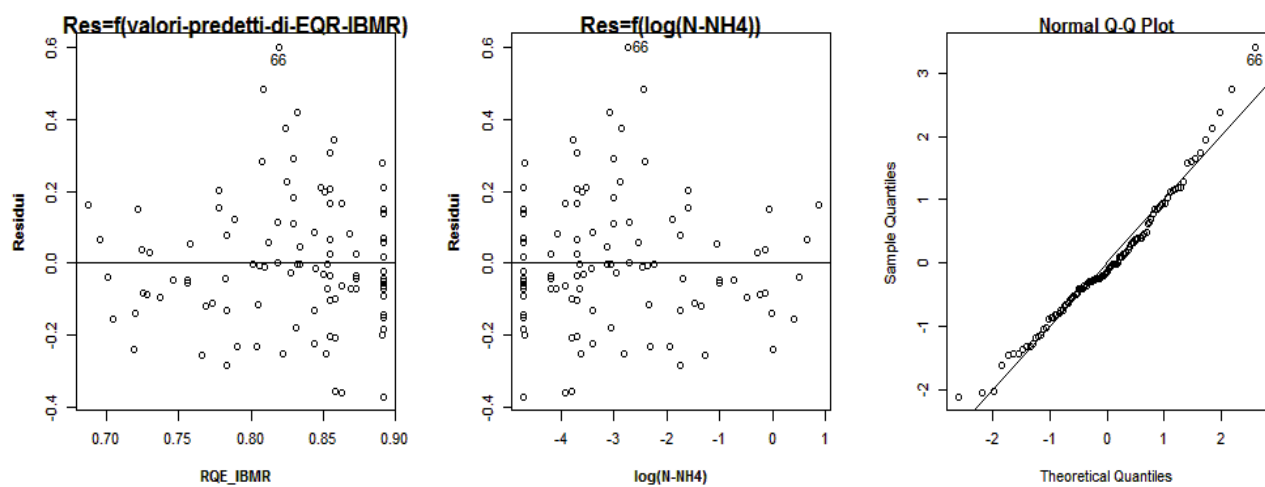


Figura 4.7. Grafici residui e Normal Q-Q Plot per regressione-lineare-logaritmica tra RQE-IBMR e N-NH₄

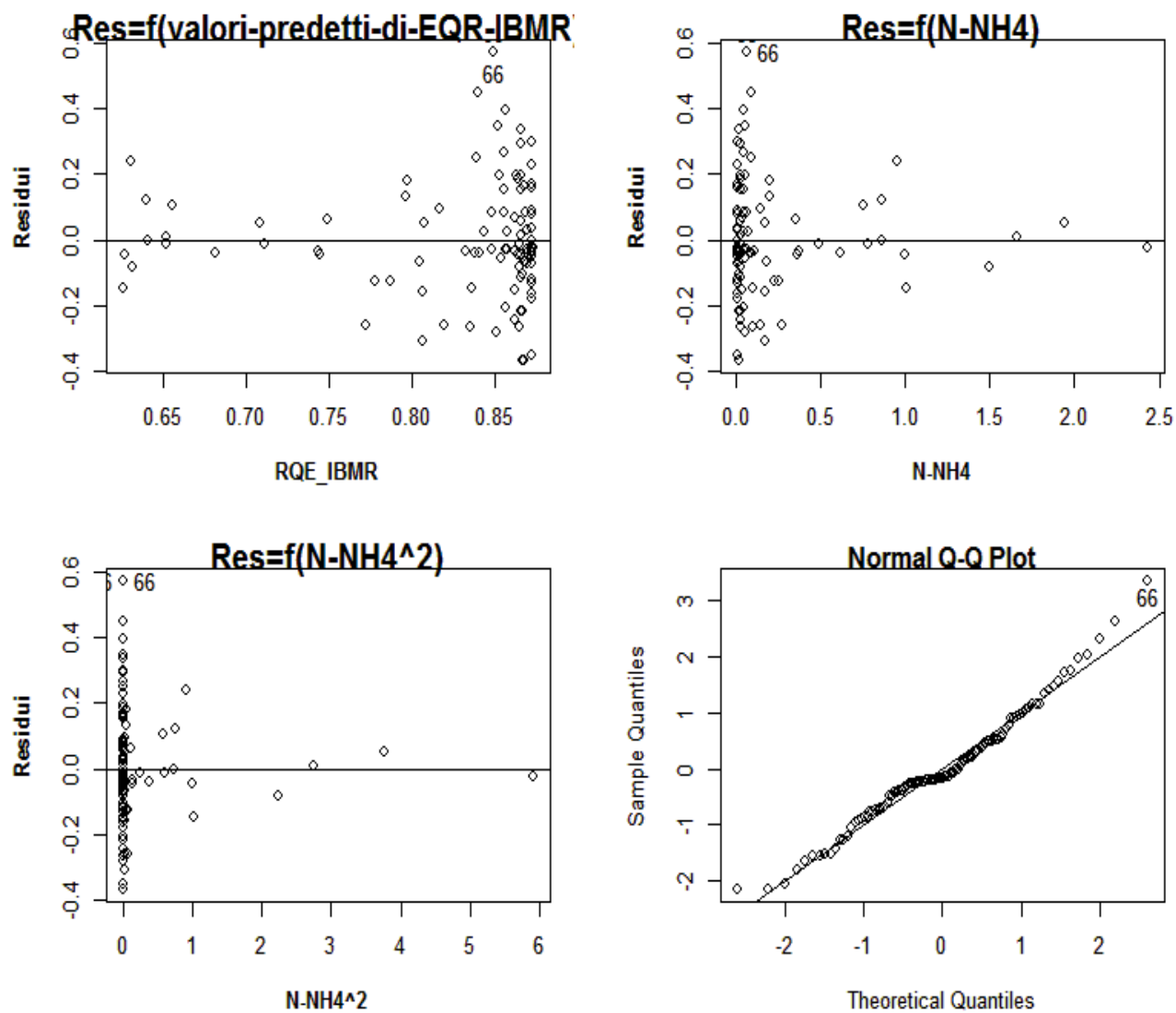


Figura 4.8. Grafici residui e Normal Q-Q Plot per regressione polinomiale tra RQE-IBMR e N-NH₄

	lineare	logaritmica	polinomiale
R2 corretto	0,082	0,134	0,122
Statistica F	10,532	17,499	8,432
p_value_F	1,57E-03	5,94E-05	4,02E-04
p_value_Shapiro	0,08	0,03	0,08
p_value_Breush_Pagan	0,488	0,771	0,776

Tabella 4.5. Confronto dei risultati relativi alla regressione lineare, polinomiale e lineare-logaritmica con il fine di esaminare la relazione tra RQE-IBMR e N-NO₃. In giallo, la trasformazione che inizialmente è stata considerata approssimare bene i dati.

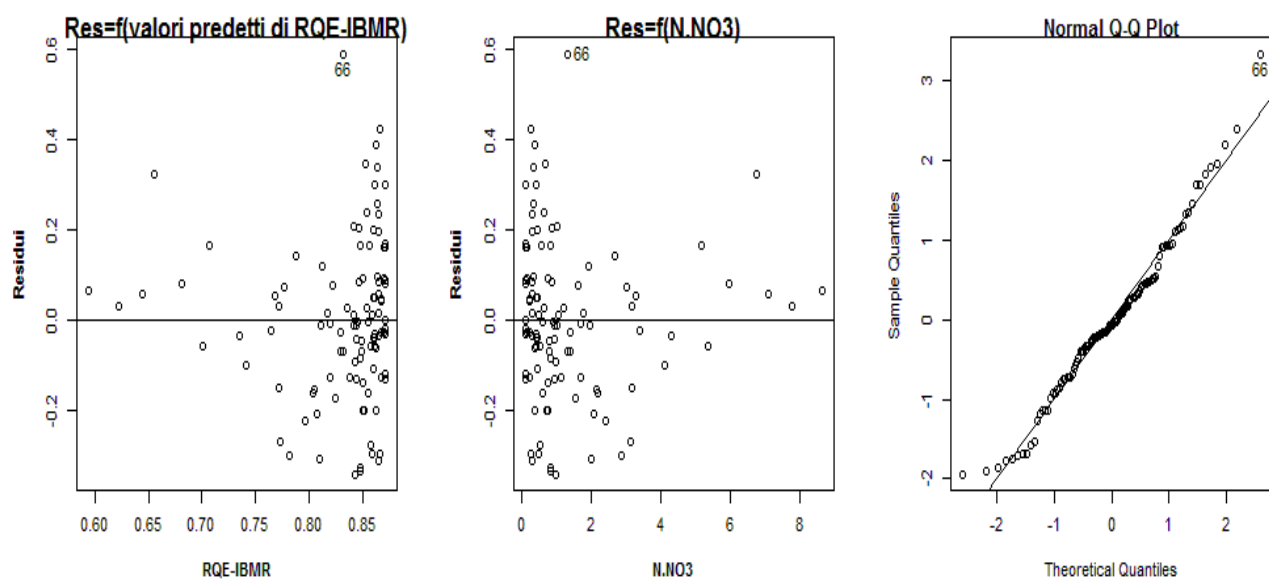


Figura 4.9. Grafici residui e Normal Q-Q Plot per regressione-lineare tra RQE-IBMR e N-NO₃

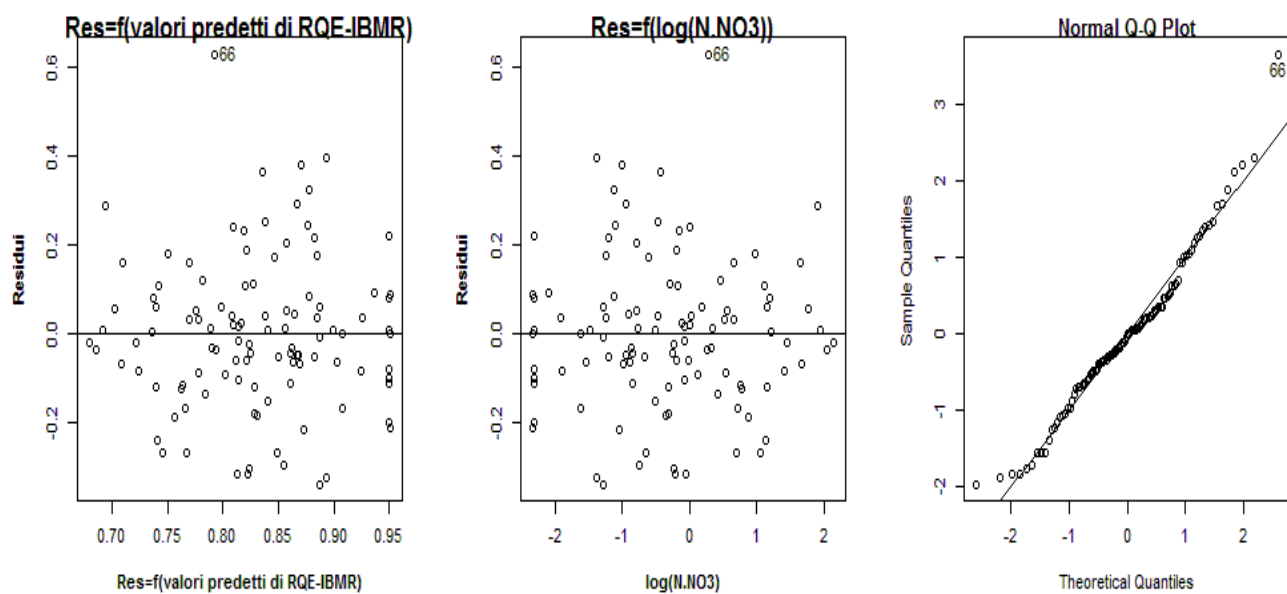


Figura 4.10. Grafici residui e Normal Q-Q Plot per regressione-lineare-logaritmica tra RQE-IBMR e N-NO₃

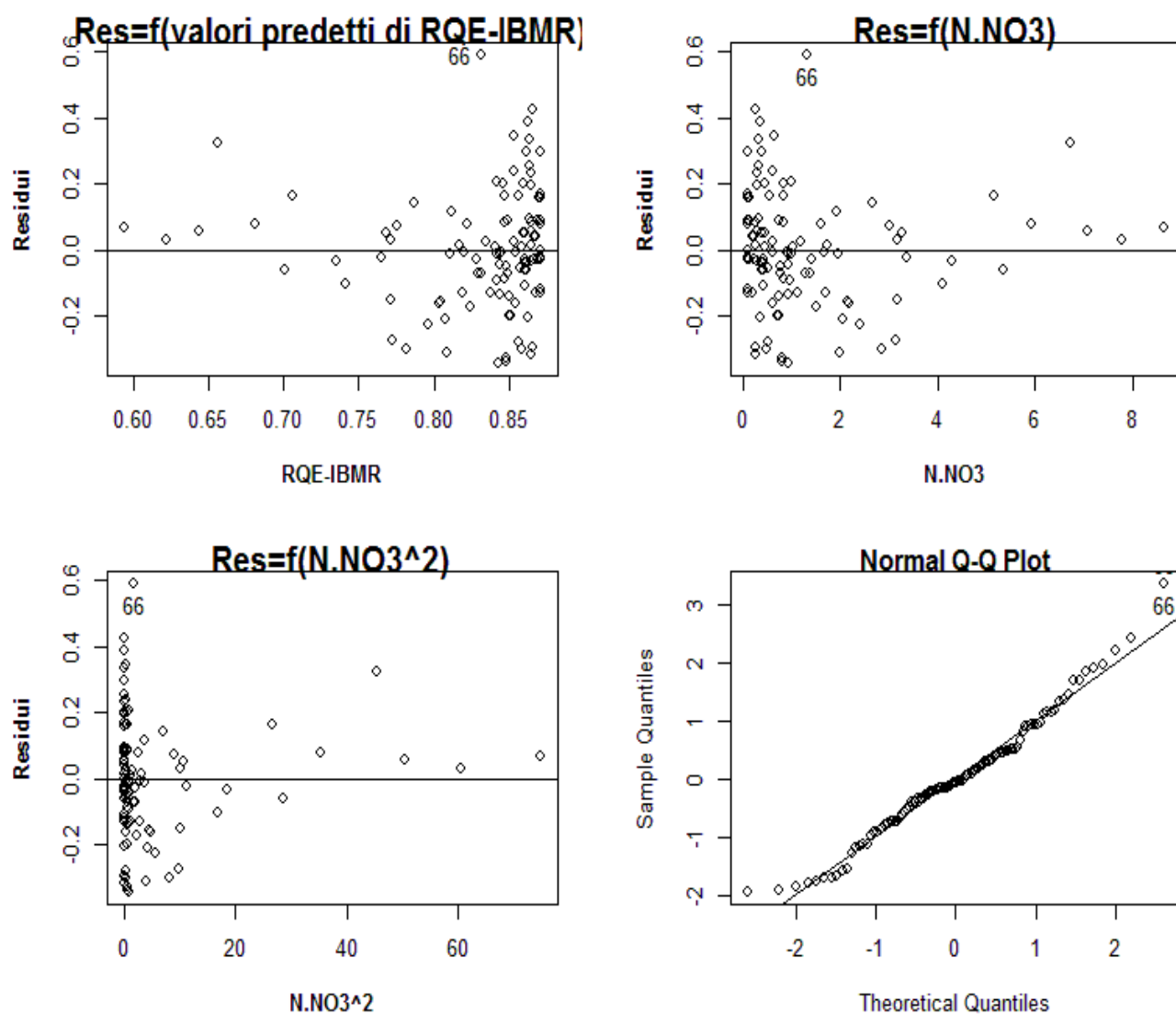


Figura 4.11. Grafici residui e Normal Q-Q Plot per regressione-polinomiale tra RQE-IBMR e N-NO₃

	lineare	logaritmica	polinomiale
R2 corretto	0,087	0,128	0,109
Statistica F	11,191	16,768	7,553
p_value_F	1,14E-03	8,00E-05	8,60E-04
p_value_Shapiro	0,070	0,002	0,011
p_value_Breush_Pagan	0,739	0,285	0,145

Tabella 4.6. Confronto dei risultati relativi alla regressione lineare, polinomiale e lineare-logaritmica con il fine di esaminare la relazione tra RQE-IBMR e PT. In giallo la trasformazione che inizialmente è stata considerata approssimare bene i dati.

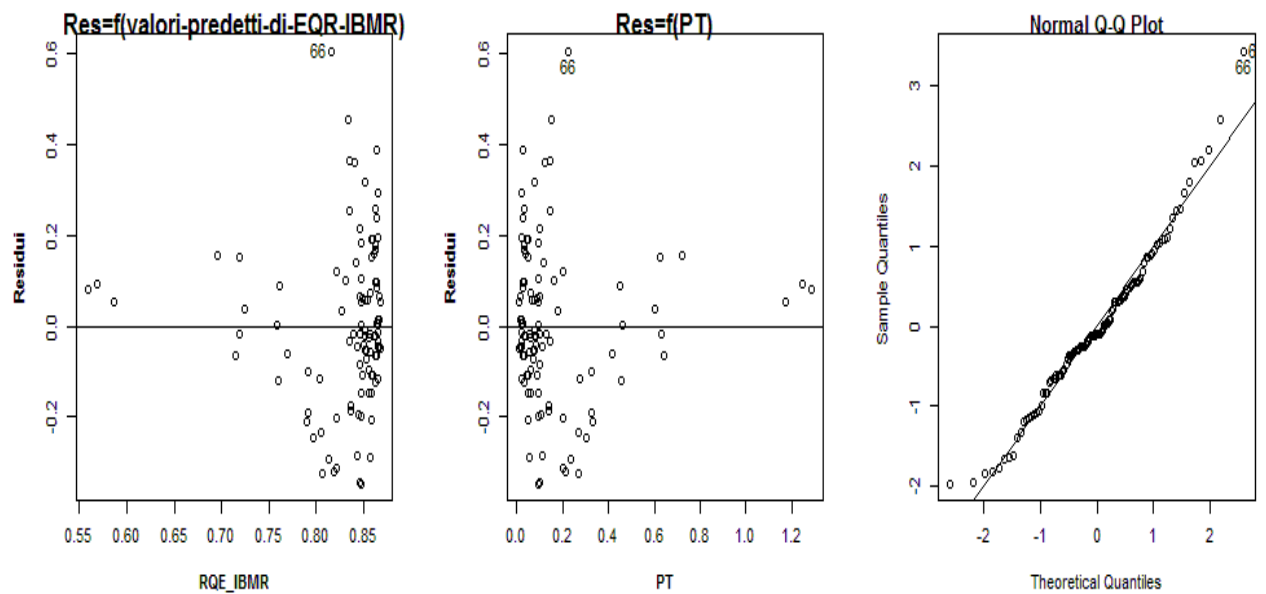


Figura 4.12. Grafici residui e Normal Q-Q Plot per regressione-lineare tra RQE-IBMR e PT

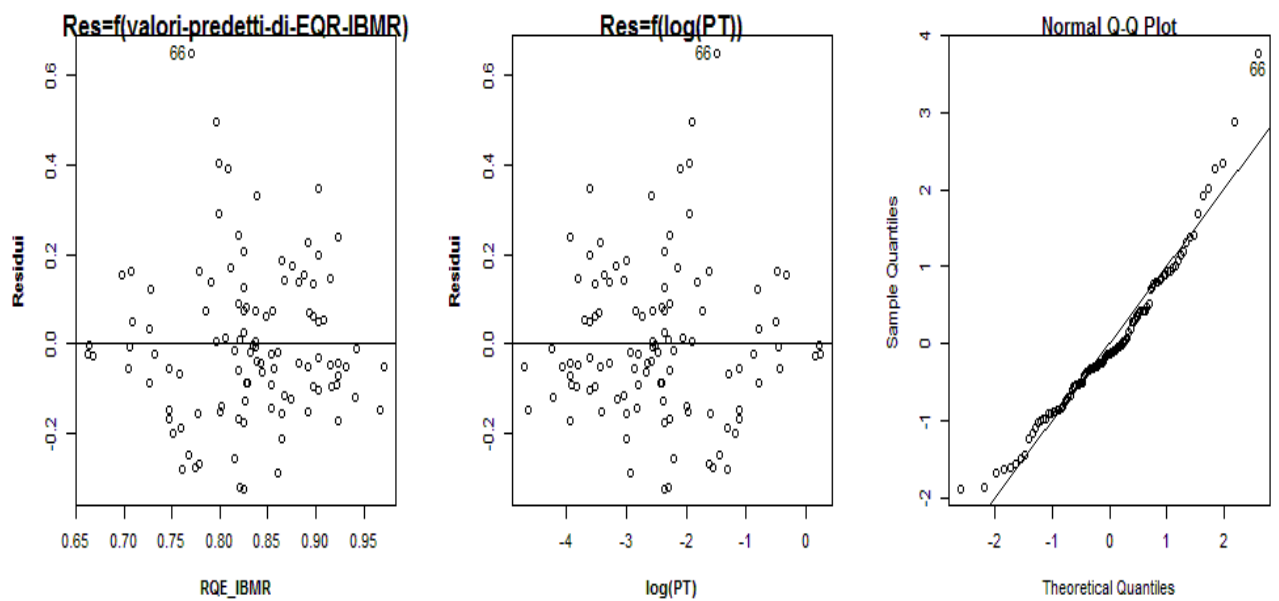


Figura 4.13. Grafici residui e Normal Q-Q Plot per regressione-lineare-logaritmica tra RQE-IBMR e PT

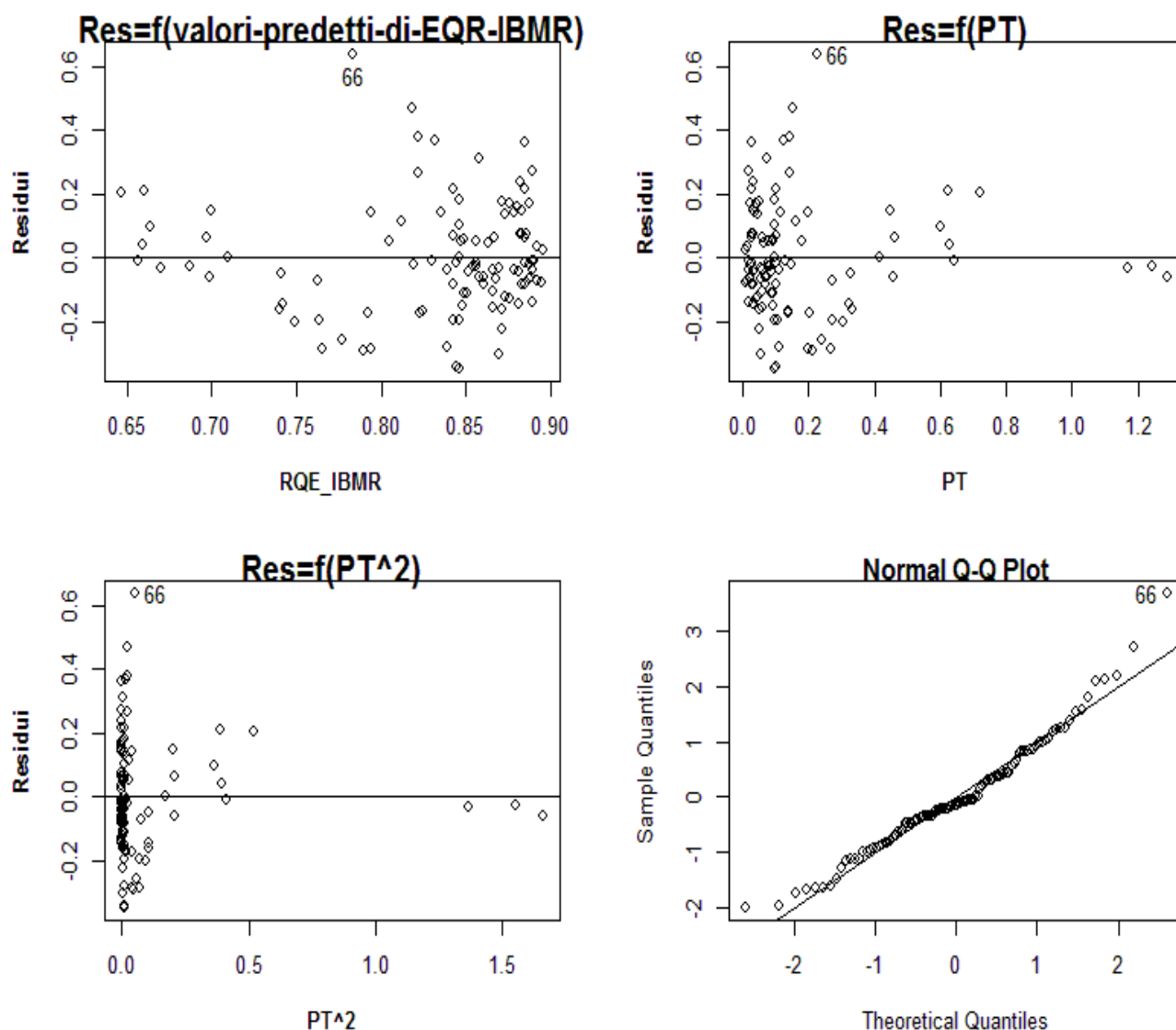


Figura 4.14. Grafici residui e Normal Q-Q Plot per regressione-polinomiale tra RQE-IBMR e PT

	lineare	logaritmica	polinomiale
R2 corretto	0,091	0,101	0,116
Statistica F	11,747	13,016	8,048
p_value_F	8,69E-04	4,73E-04	5,60E-04
p_value_Shapiro	0,015	0,015	0,033
p_value_Breush_Pagan	0,909	0,952	0,537

Tabella 4.7. Confronto dei risultati relativi alla regressione lineare, polinomiale e lineare-logaritmica con il fine di esaminare la relazione tra RQE-IBMR e OD. In giallo la trasformazione che inizialmente è stata considerata approssimare bene i dati.

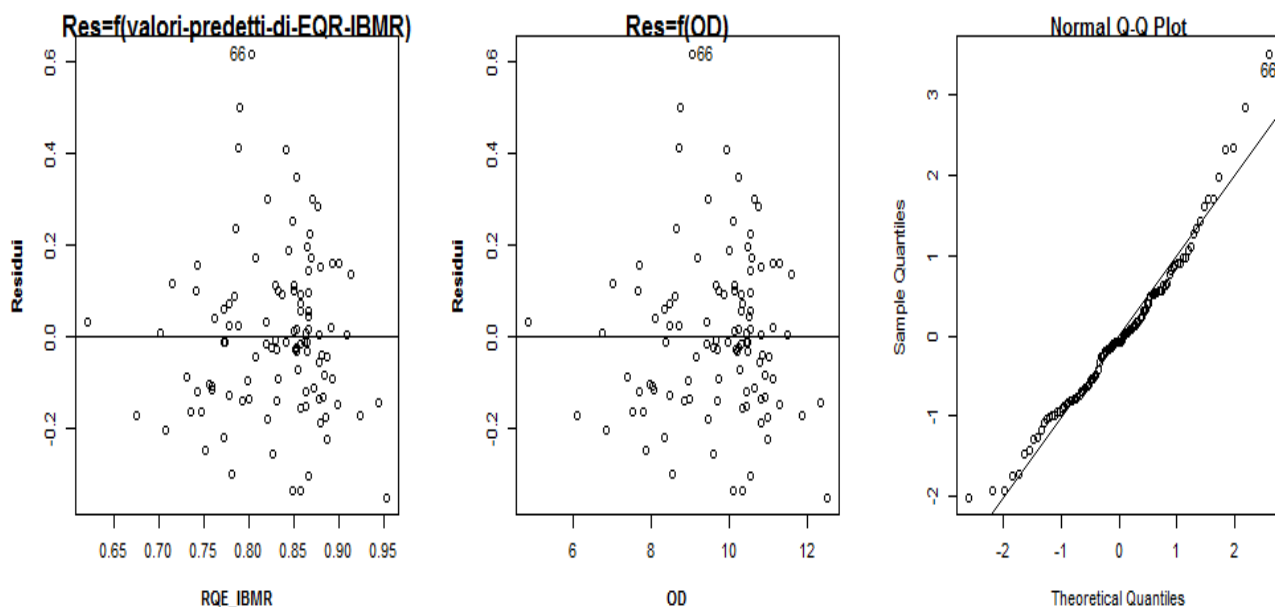


Figura 4.15. Grafici residui e Normal Q-Q Plot per regressione-lineare tra RQE-IBMR e OD

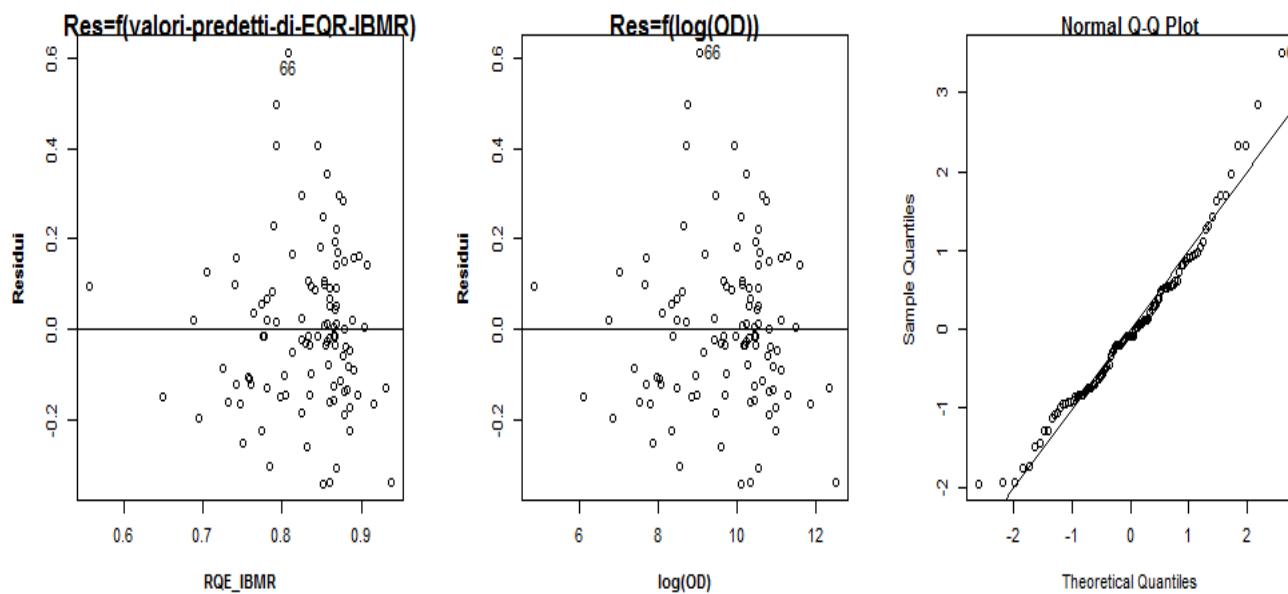


Figura 4.16. Grafici residui e Normal Q-Q Plot per regressione-lineare-logaritmica tra RQE-IBMR e OD

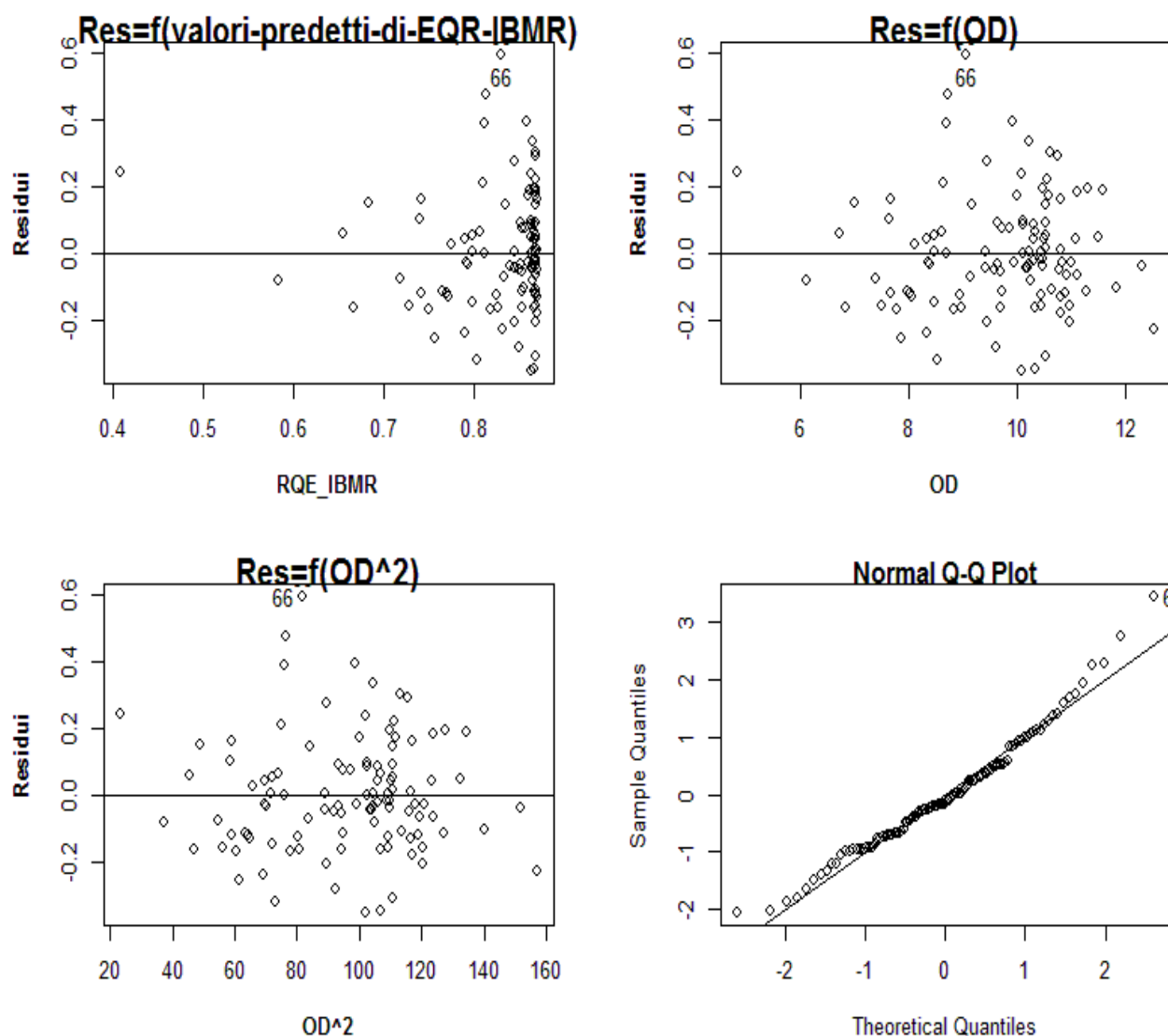


Figura 4.17. Grafici residui e Normal Q-Q Plot per regressione-polinomiale tra RQE-IBMR e OD.

	lineare	logaritmica	polinomiale
R2 corretto	0,172	0,211	0,206
Statistica F	23,250	29,630	14,910
p_value_F	4,77E-06	3,39E-07	2,00E-06
p_value_Shapiro	1,610E-06	2,830E-07	6,570E-07
p_value_Breush_Pagan	0,359	0,092	0,295

Tabella 4.8. Confronto dei risultati relativi alla regressione lineare, polinomiale e lineare-logaritmica con il fine di esaminare la relazione tra RQE-ICMi e N-NH₄. In giallo la trasformazione che inizialmente è stata considerata approssimare bene i dati.

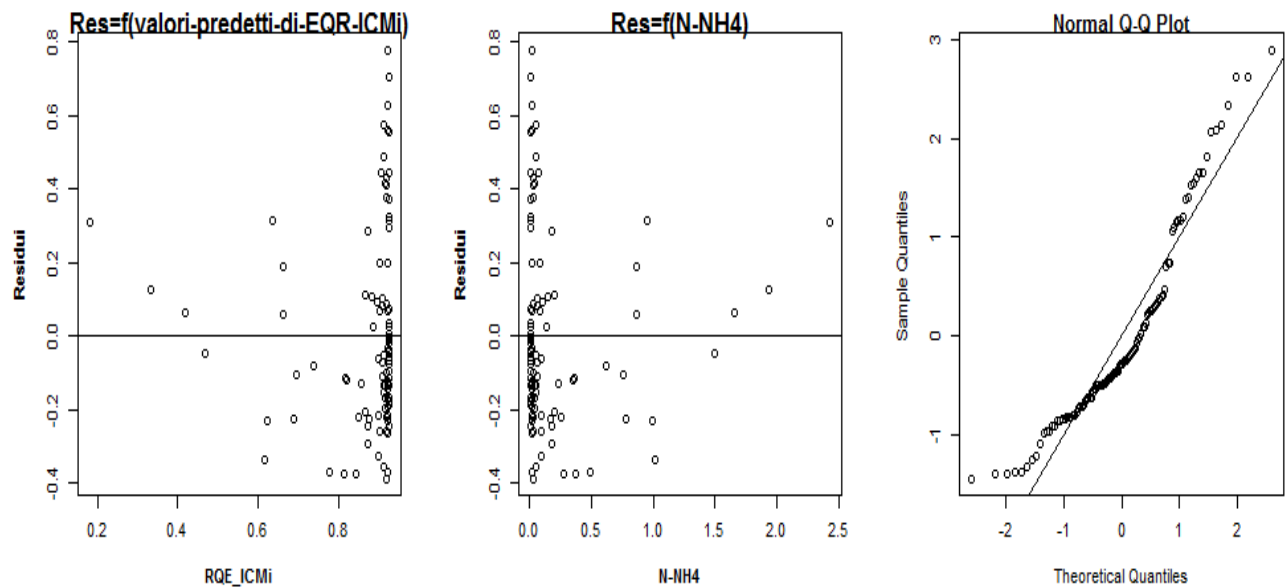


Figura 4.18. Grafici residui e Normal Q-Q Plot per regressione-lineare tra RQE-ICMi e N-NH₄

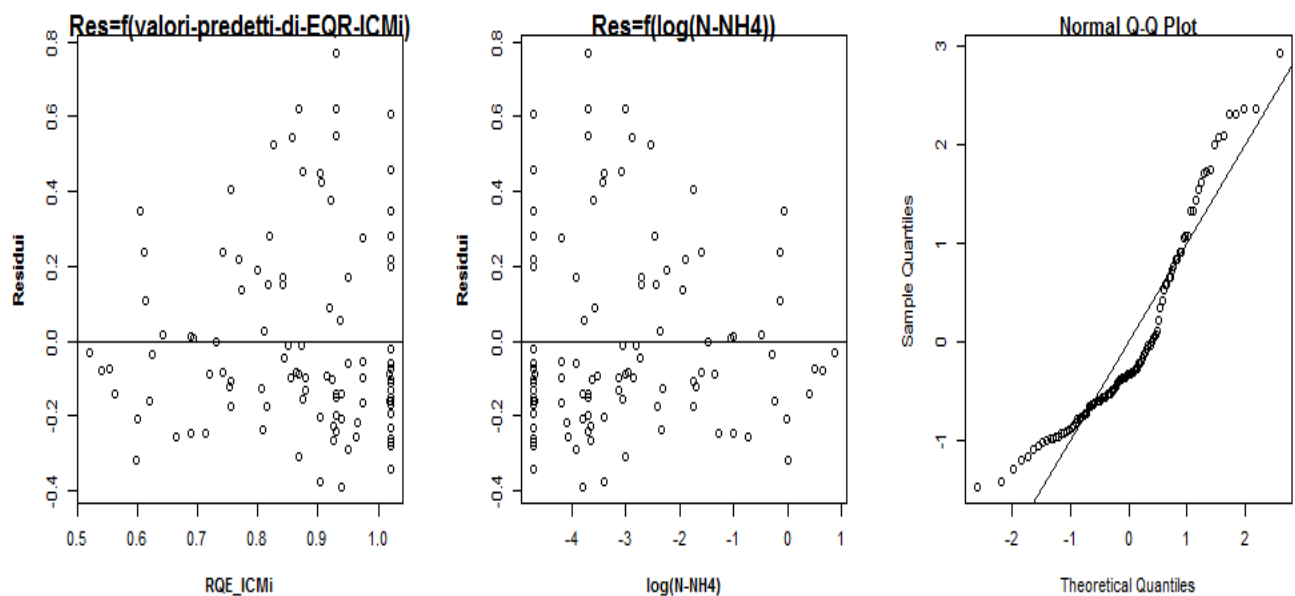


Figura 4.19. Grafici residui e Normal Q-Q Plot per regressione-lineare-logaritmica tra RQE-ICMi e N-NH₄

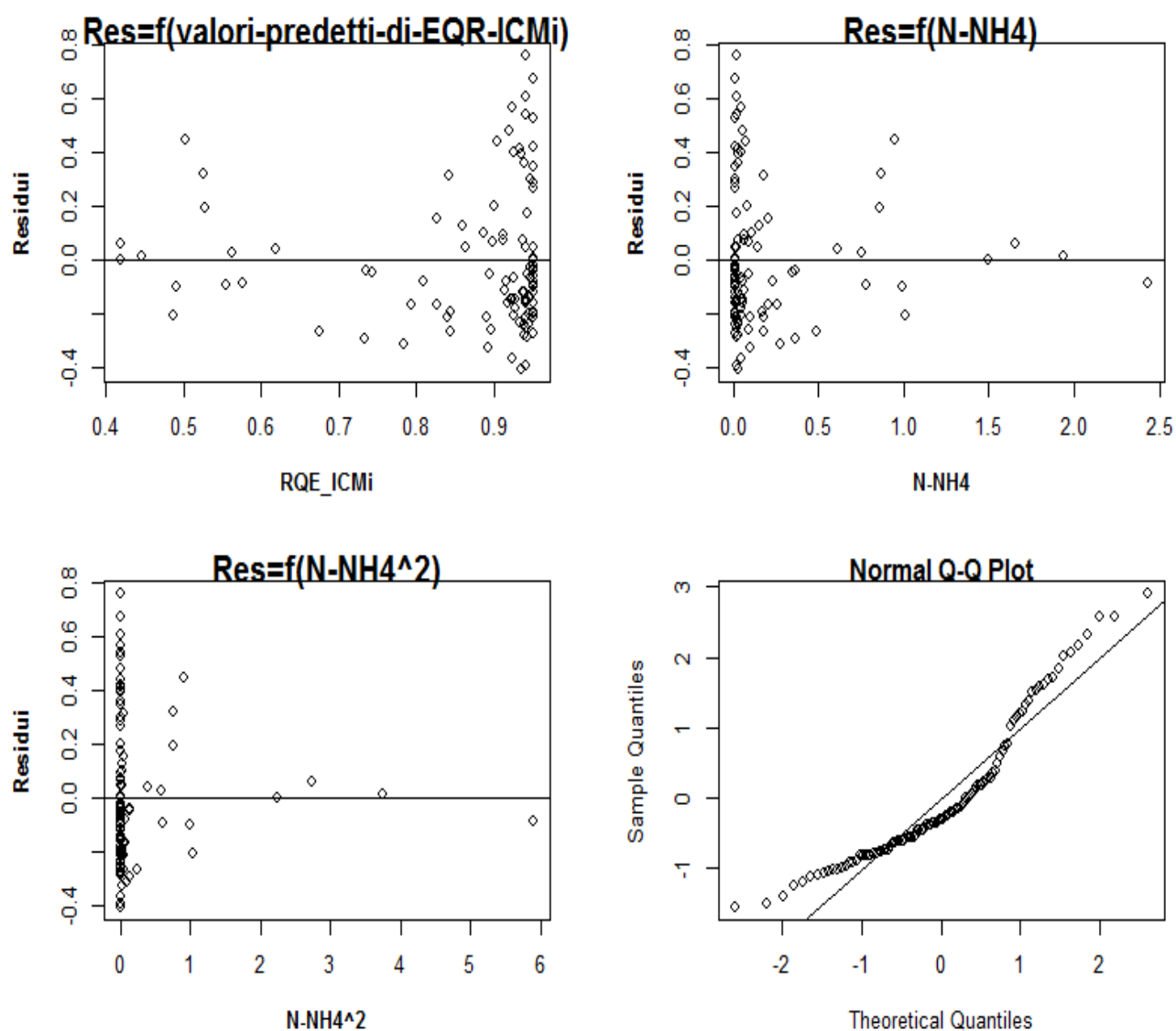


Figura 4.20. Grafici residui e Normal Q-Q Plot per regressione-polinomiale tra RQE-ICMi e N-NH₄.

	lineare	logaritmica	polinomiale
R2 corretto	0,198	0,216	0,228
Statistica F	27,470	30,420	16,840
p_value_F	8,19E-07	2,48E-07	4,54E-07
p_value_Shapiro	1,500E-03	4,710E-04	2,180E-04
p_value_Breush_Pagan	0,005	0,014	0,146

Tabella 4.9. Confronto dei risultati relativi alla regressione lineare, polinomiale e lineare-logaritmica con il fine di esaminare la relazione tra RQE-ICMi e RQE-IBMR. In giallo la trasformazione che inizialmente è stata considerata approssimare bene i dati.

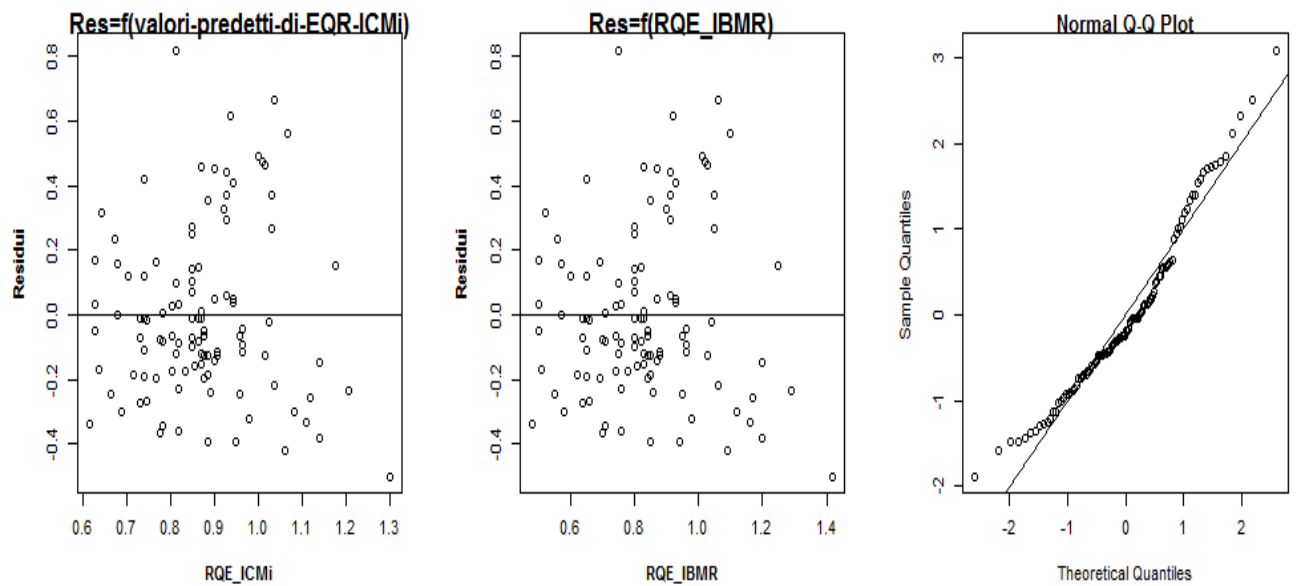


Figura 4.21. Grafici residui e Normal Q-Q Plot per regressione-lineare tra RQE-ICMi e RQE-IBMR

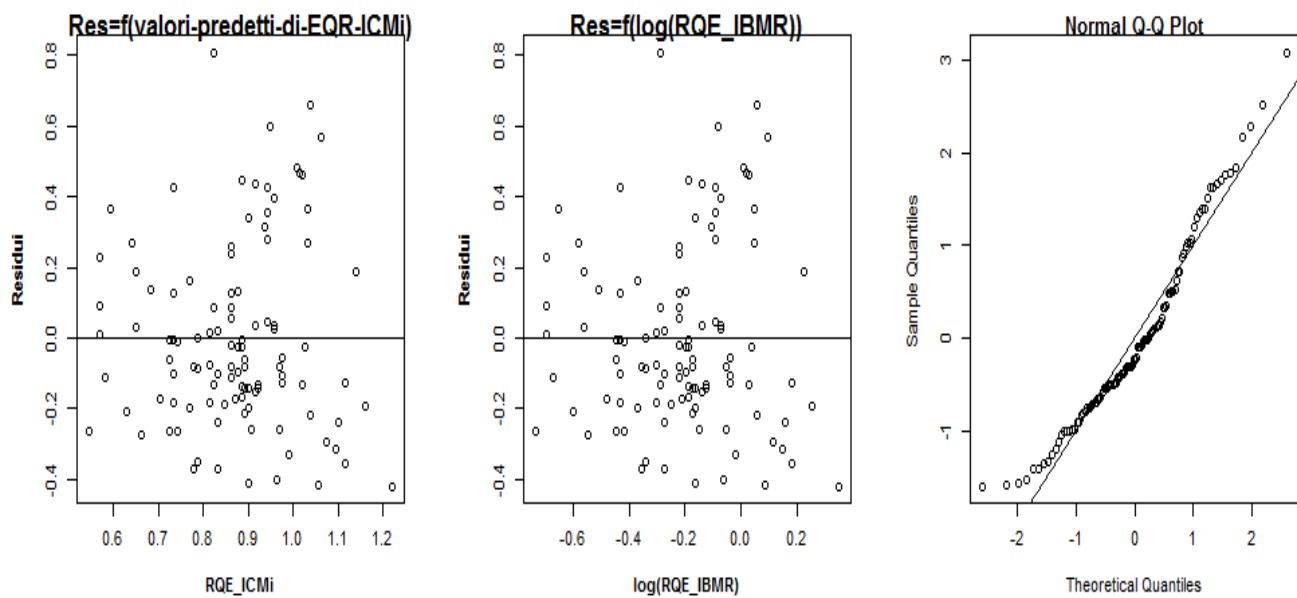


Figura 4.22. Grafici residui e Normal Q-Q Plot per regressione-lineare-logaritmica tra RQE-ICMi e RQE-IBMR

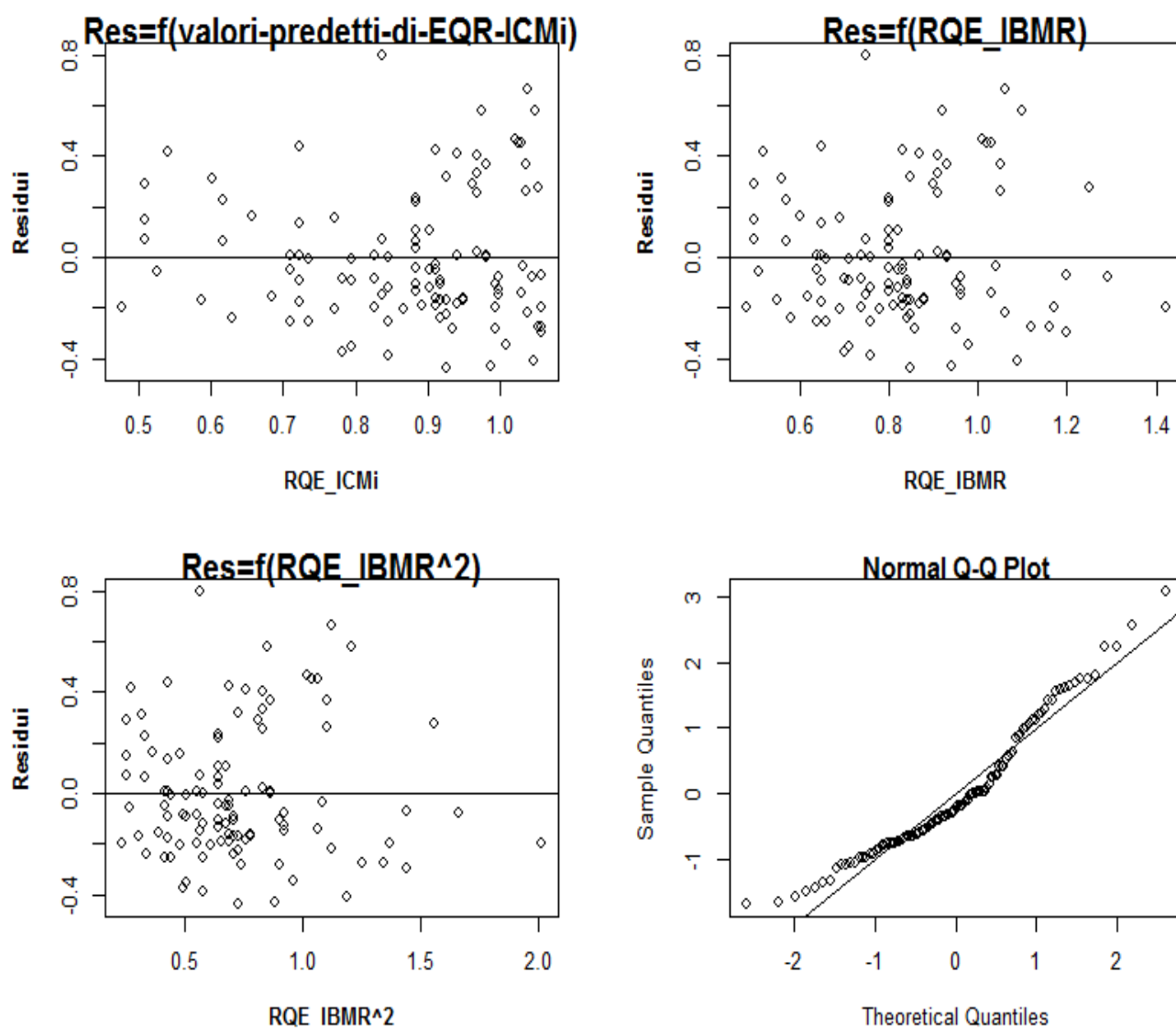


Figura 4.23. Grafici residui e Normal Q-Q Plot per regressione-polinomiale tra RQE-ICMi e RQE-IBMR.

	lineare	logaritmica	polinomiale
R2 corretto	0,142	0,116	0,228
Statistica F	18,720	15,030	16,840
p_value_F	3,48E-05	1,83E-04	4,54E-07
p_value_Shapiro	4,650E-05	7,430E-06	2,300E-05
p_value_Breush_Pagan	0,063	0,108	0,120

Tabella 4.10. Confronto dei risultati relativi alla regressione lineare, polinomiale e lineare-logaritmica con il fine di esaminare la relazione tra RQE-ICMi e N-NO₃. In giallo la trasformazione che inizialmente è stata considerata approssimare bene i dati.

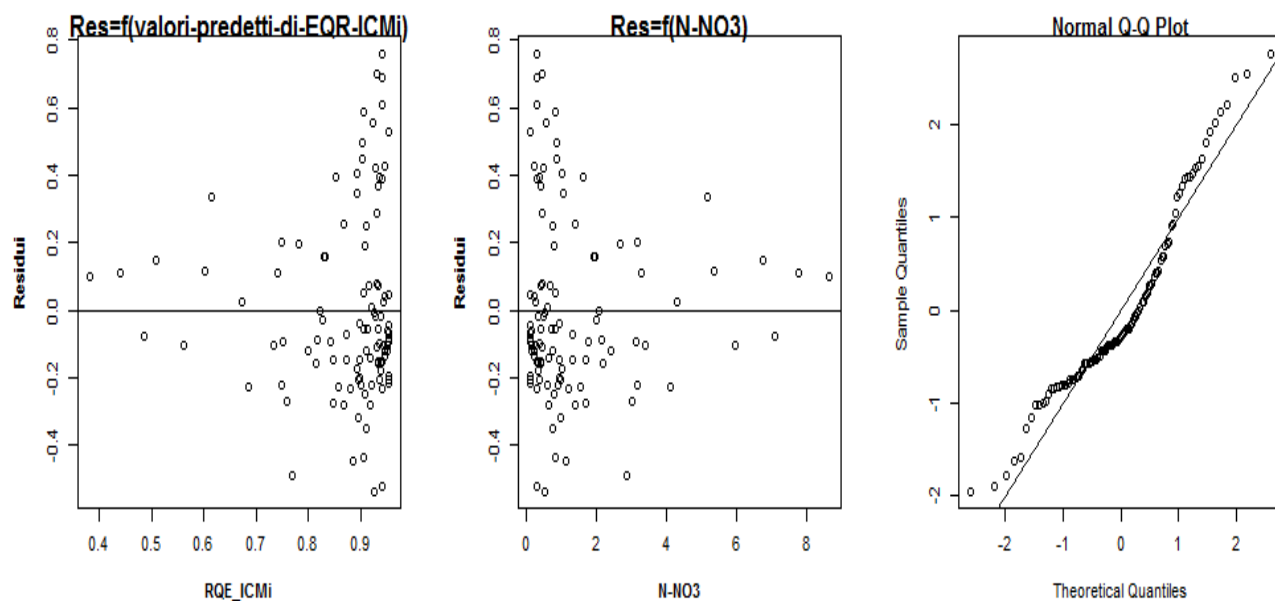


Figura 4.24. Grafici residui e Normal Q-Q Plot per regressione-lineare tra RQE-ICMi e N-NO₃

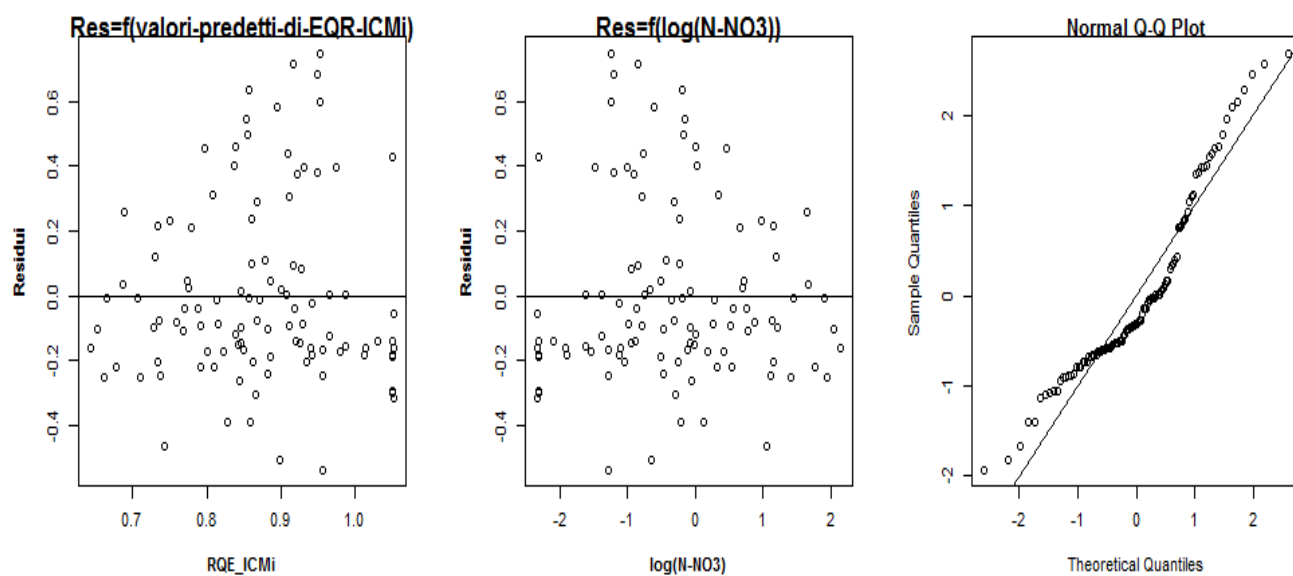


Figura 4.25. Grafici residui e Normal Q-Q Plot per regressione-lineare-logaritmica tra RQE-ICMi e N-NO₃

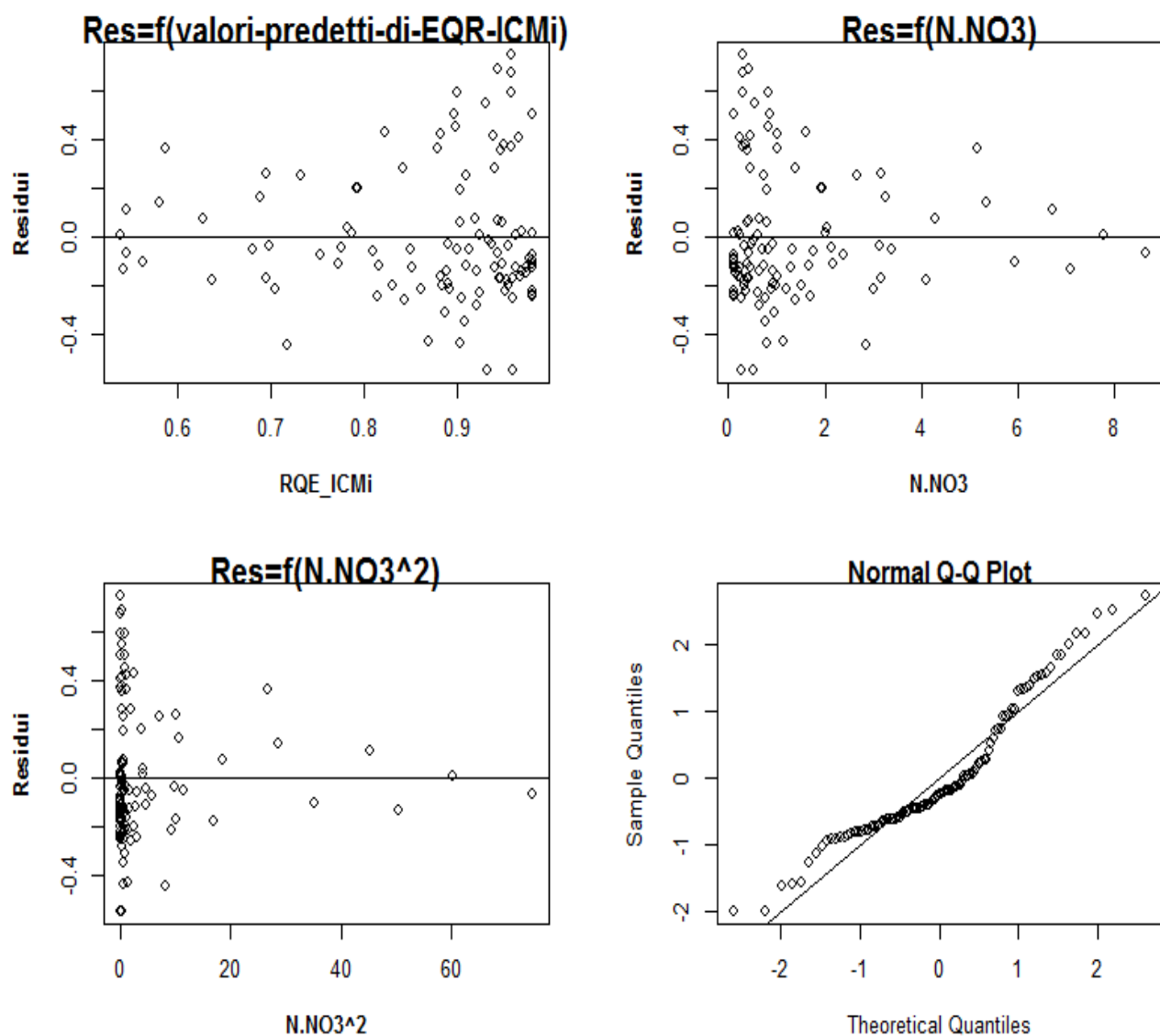


Figura 4.26. Grafici residui e Normal Q-Q Plot per regressione-polinomiale tra RQE-ICMi e N-NO_3

	lineare	logaritmica	polinomiale
R2 corretto	0,135	0,247	0,197
Statistica F	17,760	36,070	14,150
p_value_F	5,28E-05	2,71E-08	3,61E-07
p_value_Shapiro	4,040E-06	1,860E-06	2,530E-07
p_value_Breush_Pagan	0,736	0,083	0,914

Tabella 4.11. Confronto dei risultati relativi alla regressione lineare, polinomiale e lineare-logaritmica con il fine di esaminare la relazione tra RQE-ICMi e PT. In giallo, la trasformazione che inizialmente è stata considerata approssimare bene i dati.

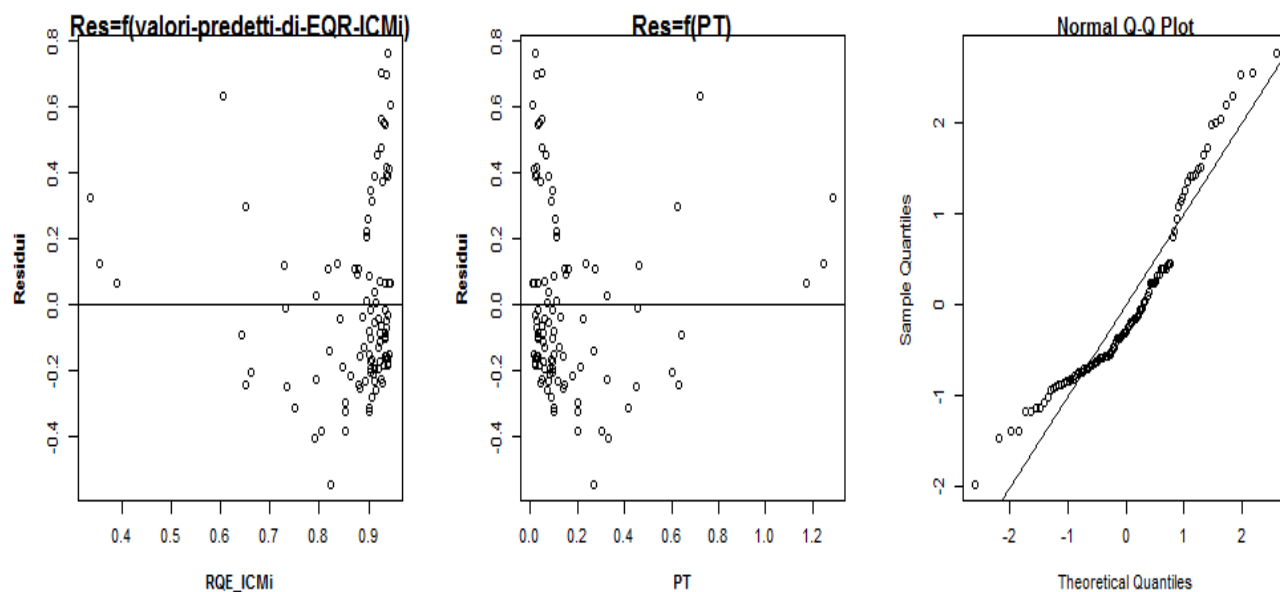


Figura 4.27. Grafici residui e Normal Q-Q Plot per regressione-lineare tra RQE-ICMi e PT

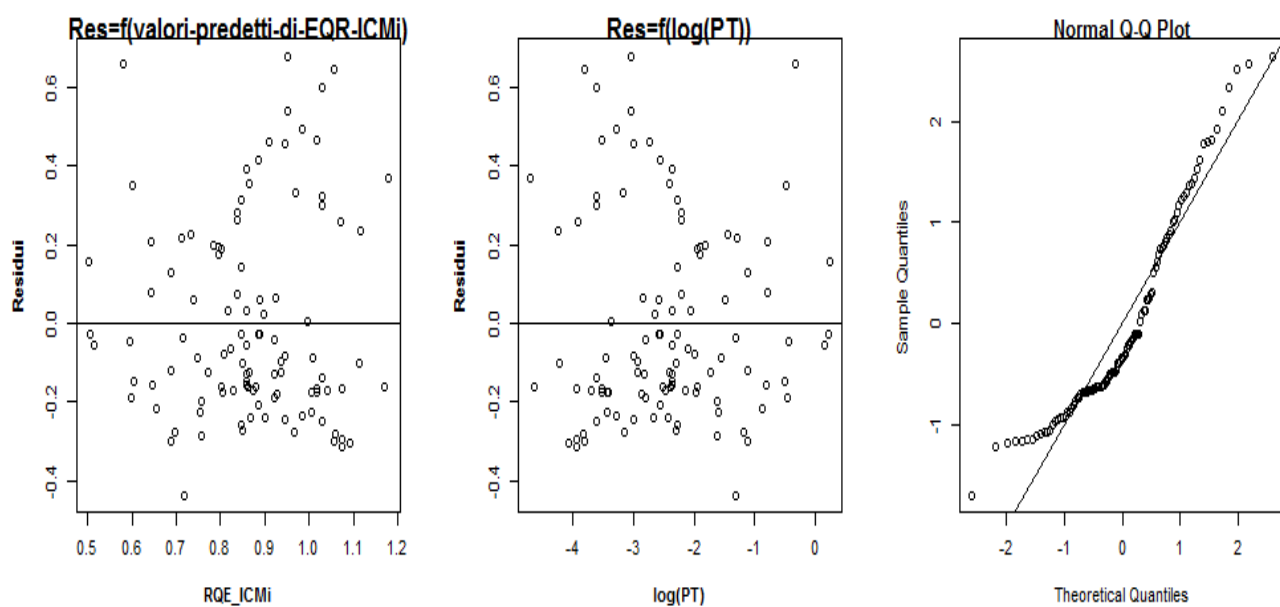


Figura 4.28. Grafici residui e Normal Q-Q Plot per regressione-lineare-logaritmica tra RQE-ICMi e PT

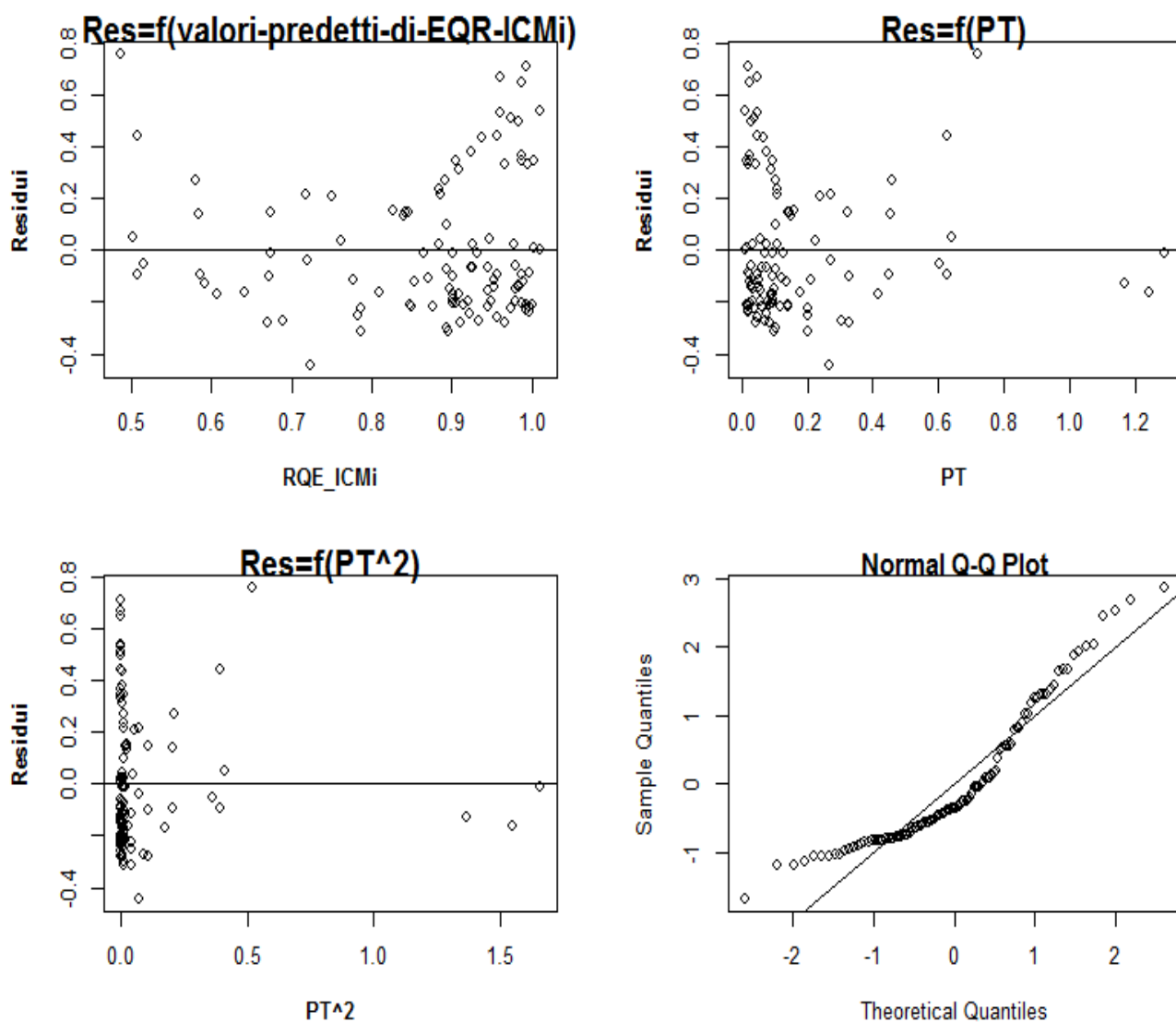


Figura 4.29. Grafici residui e Normal Q-Q Plot per regressione-polinomiale tra RQE-ICMi e PT

	lineare	logaritmica	polinomiale
R2 corretto	0,120	0,103	0,140
Statistica F	15,520	13,280	9,675
p_value_F	1,46E-04	4,18E-04	1,39E-04
p_value_Shapiro	1,130E-04	8,200E-05	2,010E-04
p_value_Breush_Pagan	0,332	0,497	0,295

Tabella 4.12. Confronto dei risultati relativi alla regressione lineare, polinomiale e lineare-logaritmica con il fine di esaminare la relazione tra RQE-ICMi e OD. In giallo la trasformazione che inizialmente è stata considerata approssimare bene i dati.

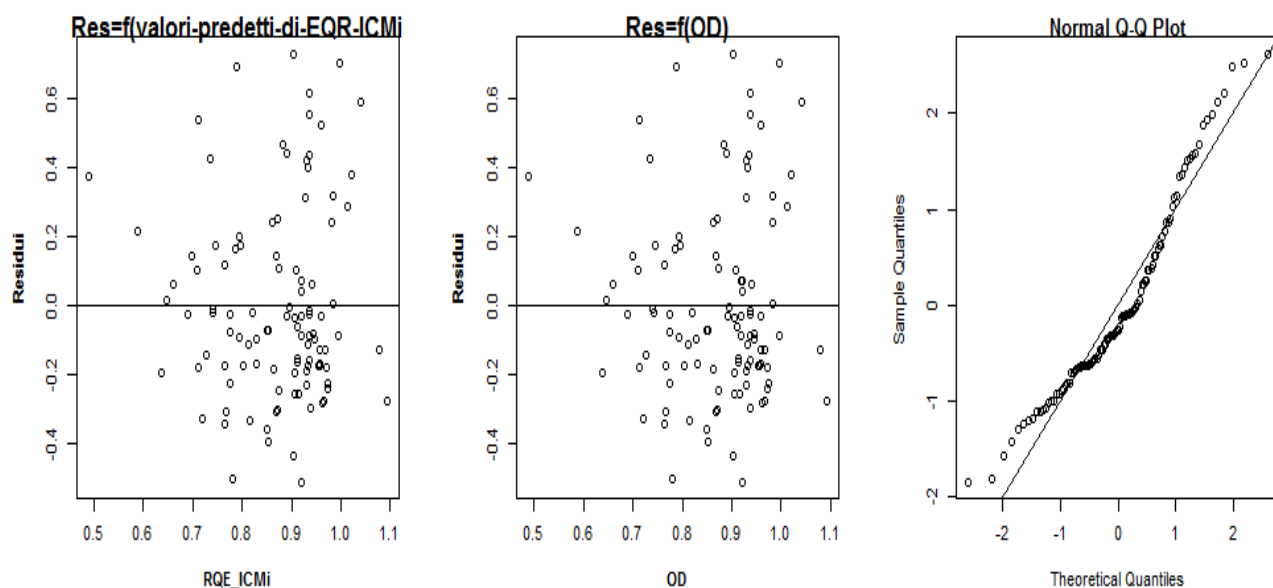


Figura 4.30. Grafici residui e Normal Q-Q Plot per regressione-lineare tra RQE-ICMi e OD.

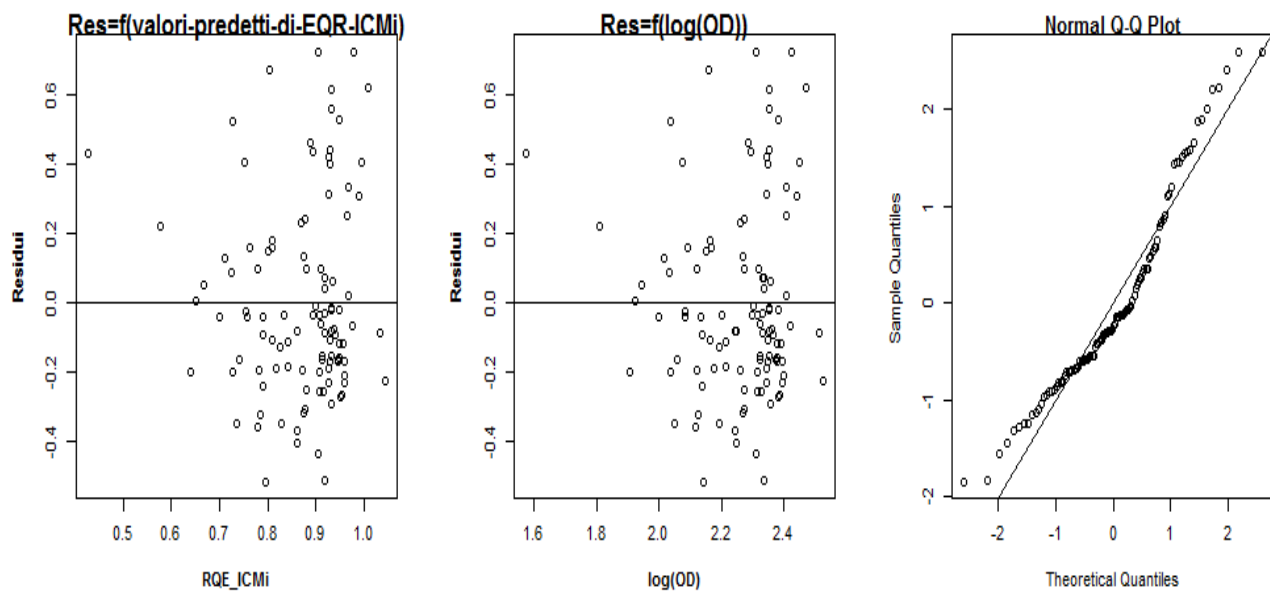


Figura 4.31. Grafici residui e Normal Q-Q Plot per regressione-lineare-logaritmica tra RQE-ICMi e OD

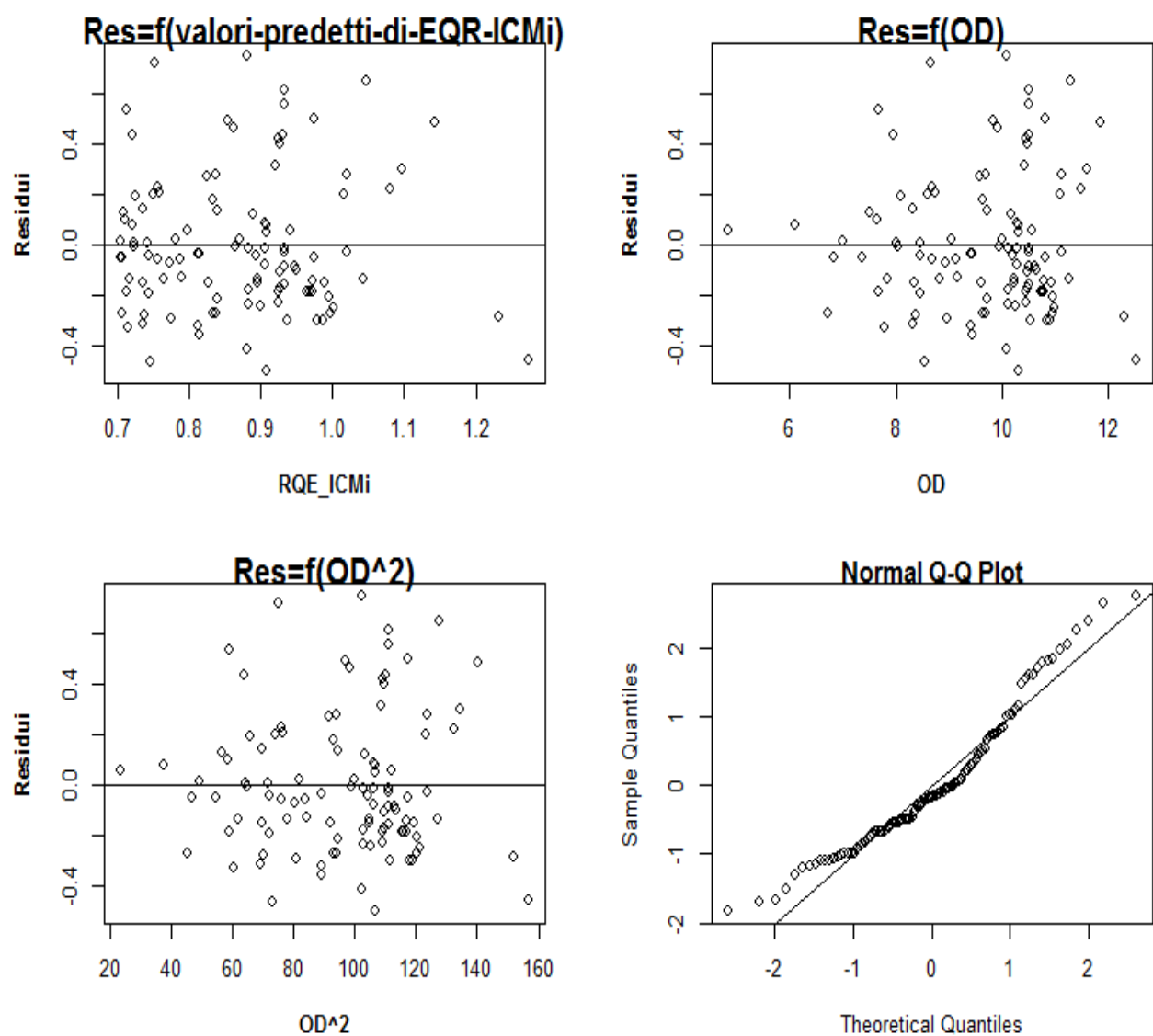


Figura 4.32. Grafici residui e Normal Q-Q Plot per regressione-polinomiale tra RQE-ICMi e OD.

	lineare	logaritmica	polinomiale
R2 corretto	0,254	0,289	0,322
Statistica F	30,657	36,278	21,671
p_value_F	3,29E-04	4,11E-05	2,48E-05
p_value_Shapiro	0,538	0,705	0,661
p_value_Breush_Pagan	0,110	0,054	0,023

Tabella 4.13. Confronto dei risultati relativi alla regressione lineare, polinomiale e lineare-logaritmica con il fine di esaminare la relazione tra RQE-STAR_ICMi e RQE_IBMR. In giallo, la trasformazione che inizialmente è stata considerata approssimare bene i dati.

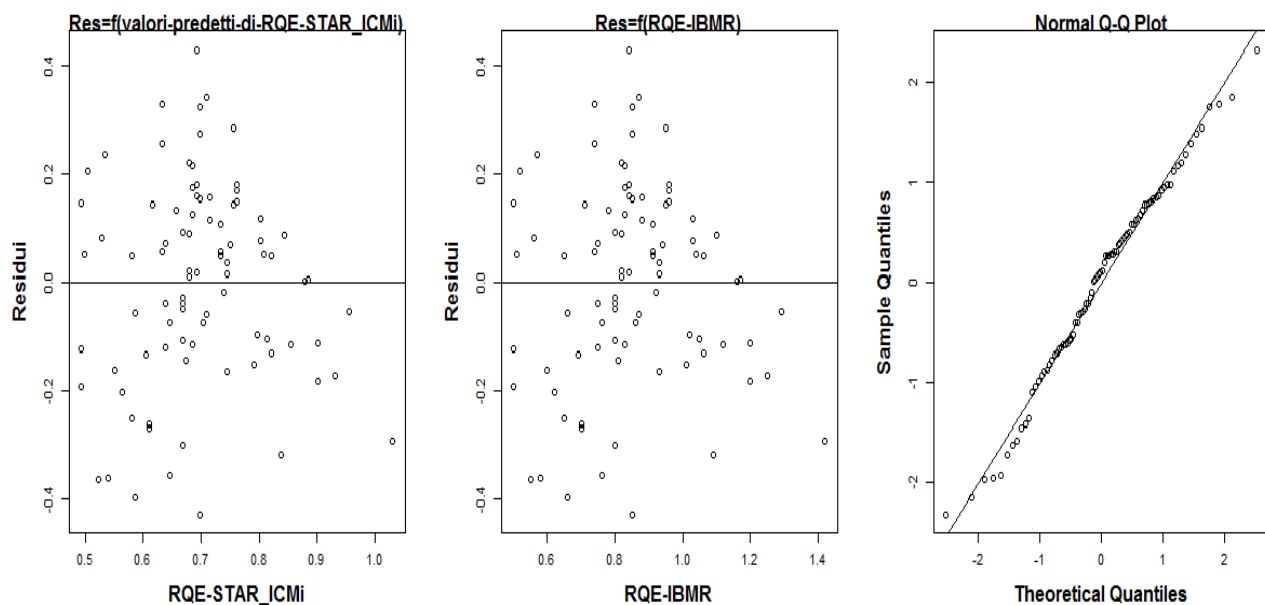


Figura 4.33. Grafici residui e Normal Q-Q Plot per regressione-lineare tra RQE-STAR_ICMi e RQE_IBMR.

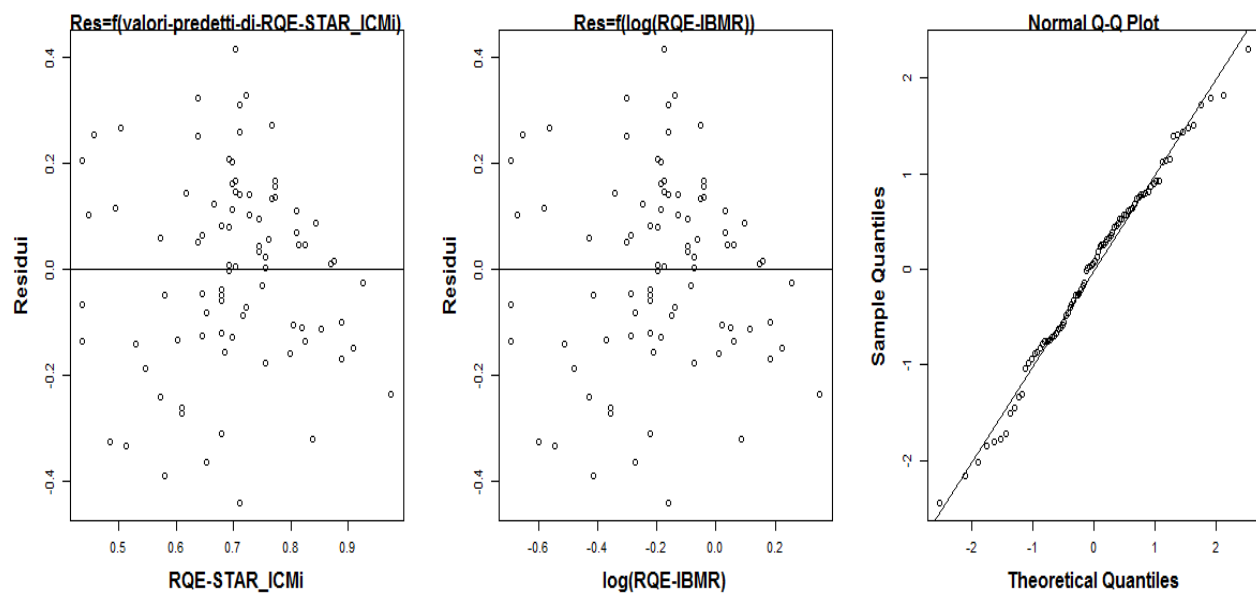


Figura 4.34. Grafici residui e Normal Q-Q Plot per regressione-lineare-logaritmica tra RQE-STAR_ICMi e RQE_IBMR.

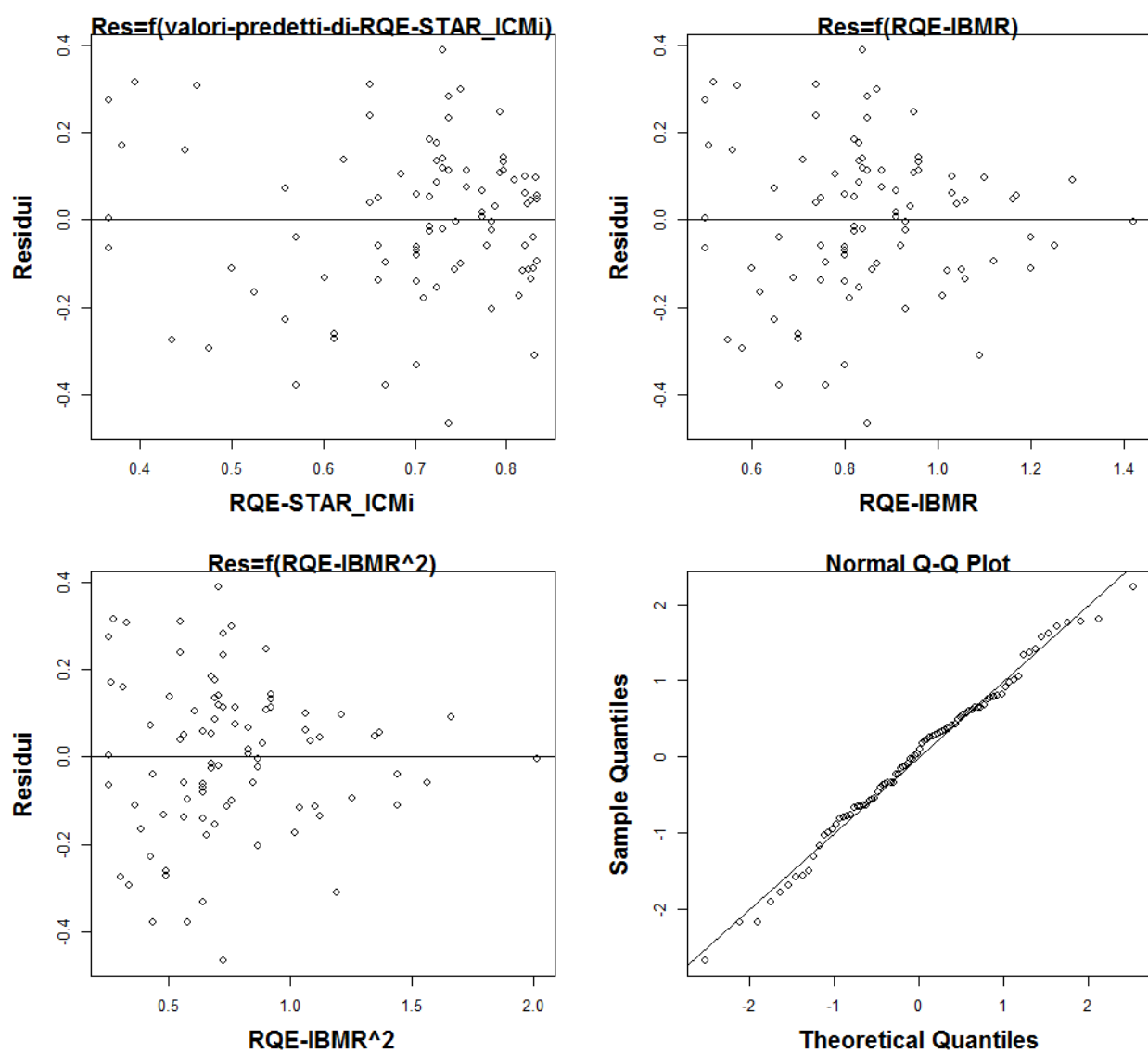


Figura 4.35. Grafici residui e Normal Q-Q Plot per regressione-polinomiale tra RQE-STAR_ICMi e RQE_IBMR.

	lineare	logaritmica	polinomiale
R2 corretto	0,136	0,226	0,375
Statistica F	14,726	26,380	27,100
p_value_F	2,37E-04	1,73E-03	7,85E-07
p_value_Shapiro	0,234	0,454	0,593
p_value_Breush_Pagan	0,009	0,086	0,582

Tabella 4.14. Confronto dei risultati relativi alla regressione lineare, polinomiale e lineare-logaritmica con il fine di esaminare la relazione tra RQE-STAR_ICMi e RQE_ICMi. In giallo, la trasformazione che inizialmente è stata considerata approssimare bene i dati.

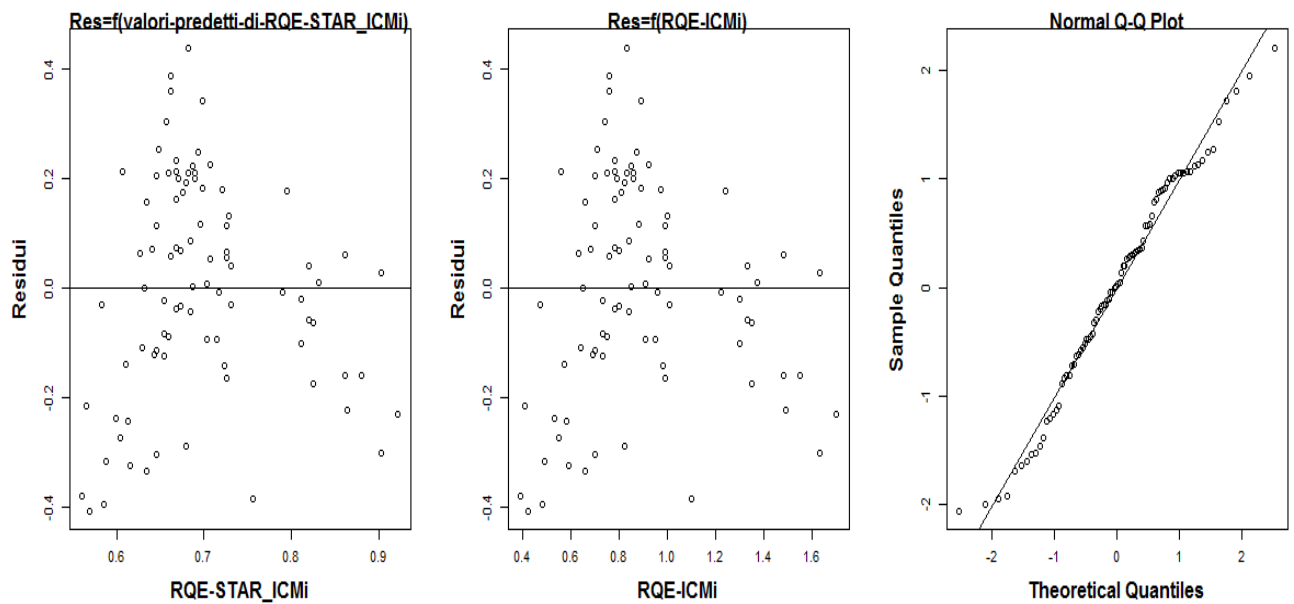


Figura 4.36. Grafici residui e Normal Q-Q Plot per regressione-lineare tra RQE-STAR_ICMi e RQE_ICMi.

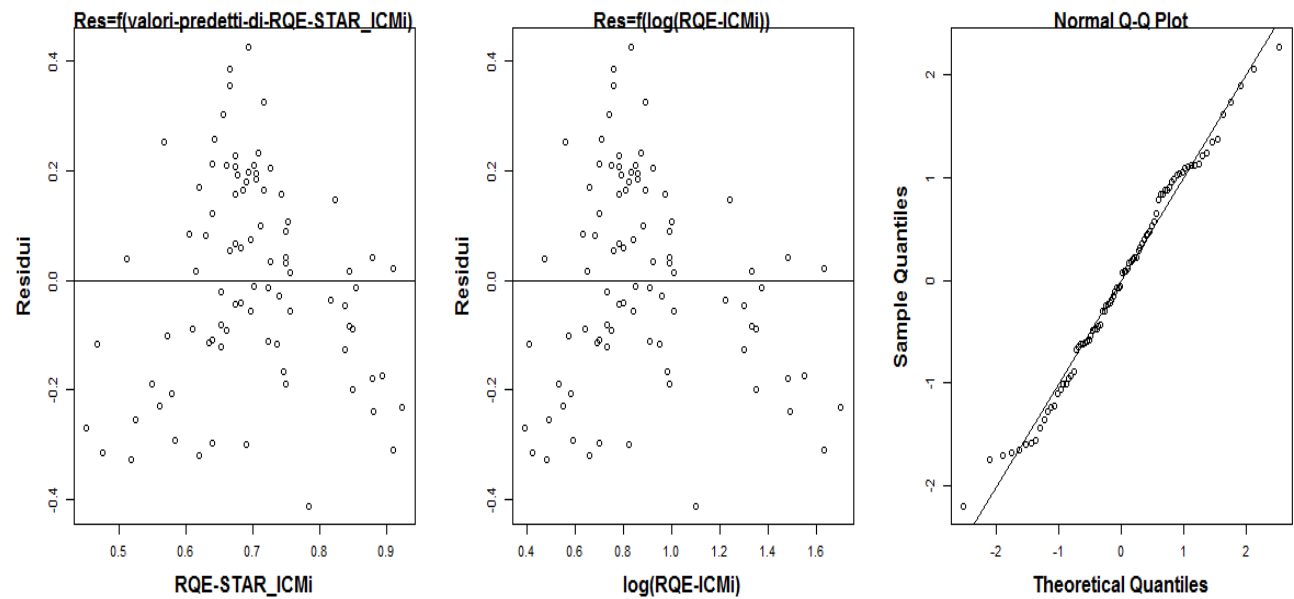


Figura 4.37. Grafici residui e Normal Q-Q Plot per regressione-lineare-logaritmica tra RQE-STAR_ICMi e RQE_ICMi.

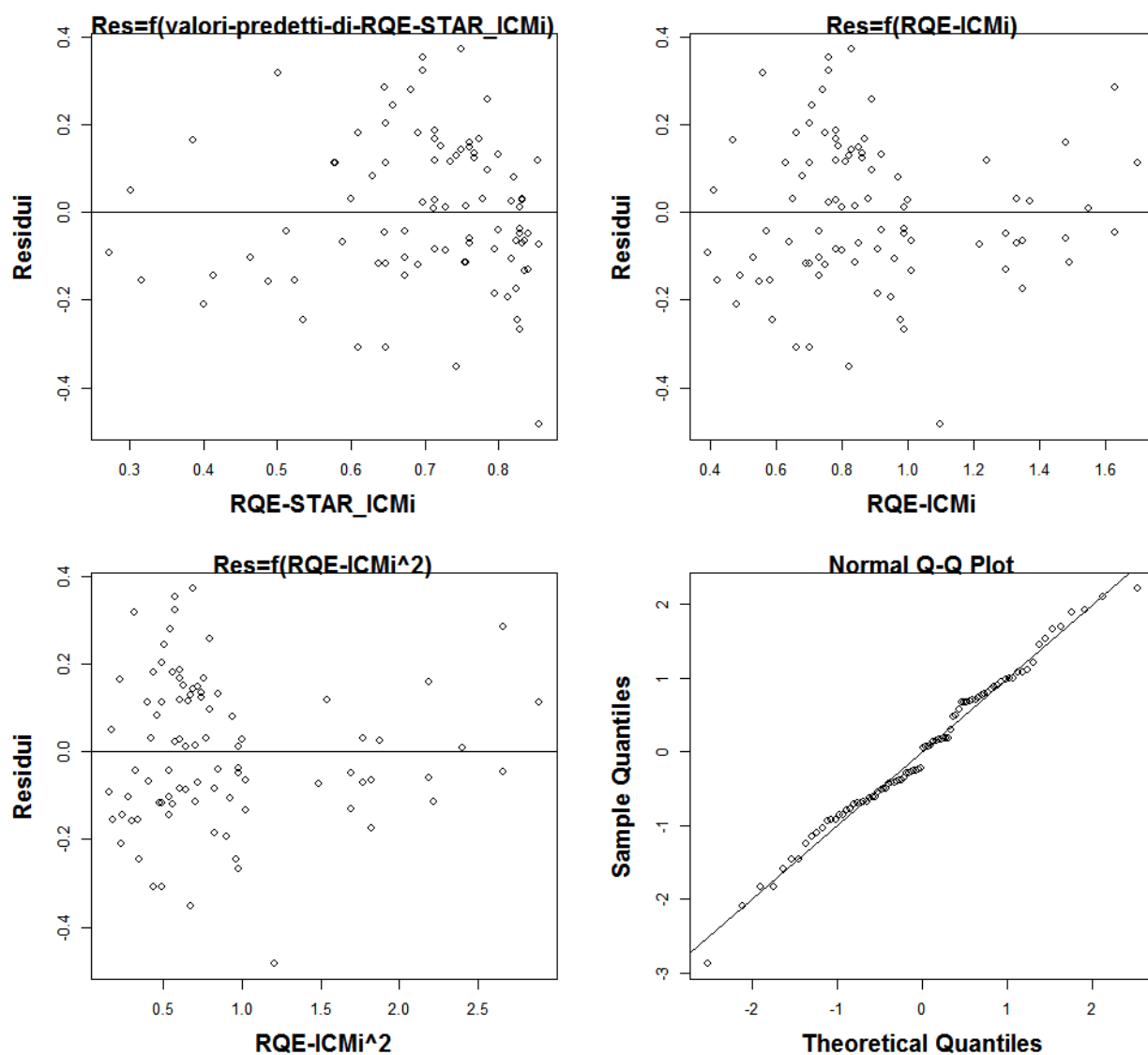


Figura 4.38. Grafici residui e Normal Q-Q Plot per regressione-polinomiale tra RQE-STAR_ICMi e RQE_ICMi

	lineare	logaritmica	polinomiale
R2 corretto	0,358	0,370	0,363
Statistica F	49,448	52,017	25,730
p_value_F	4,58E-10	2,01E-07	1,82E-06
p_value_Shapiro	0,010	0,025	0,036
p_value_Breush_Pagan	0,121	0,146	0,352

Tabella 4.15. Confronto dei risultati relativi alla regressione lineare, polinomiale e lineare-logaritmica con il fine di esaminare la relazione tra RQE-STAR_ICMi e ISECI. In giallo, la trasformazione che inizialmente è stata considerata approssimare bene i dati.

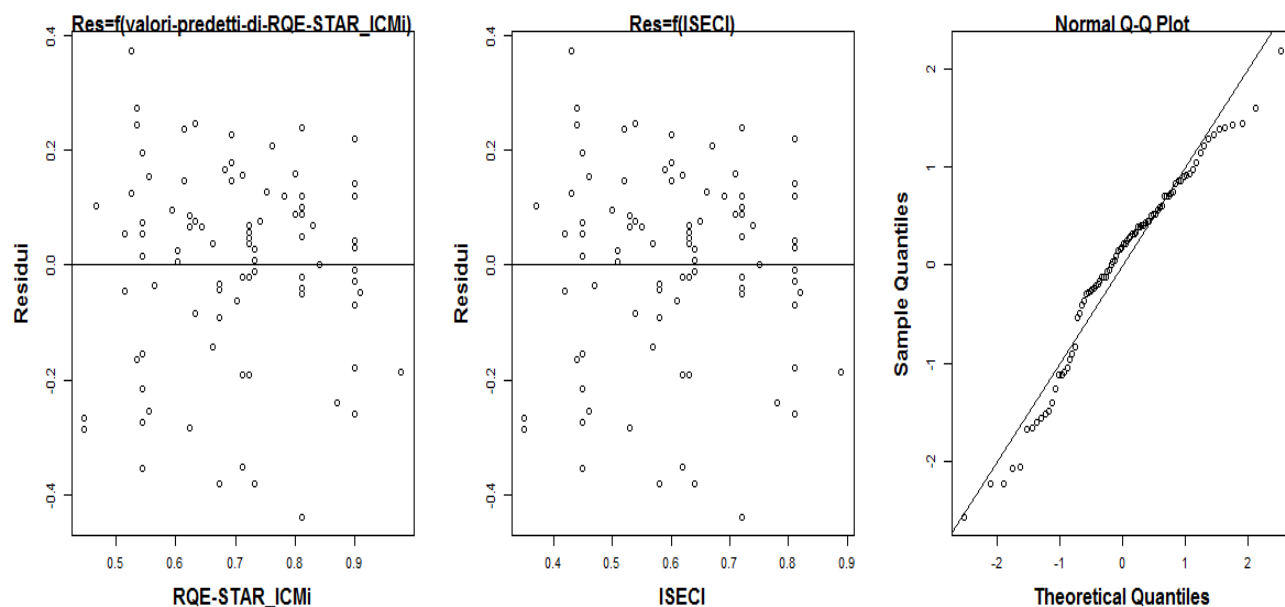


Figura 4.39. Grafici residui e Normal Q-Q Plot per regressione-lineare tra RQE-STAR_ICMi e ISECI

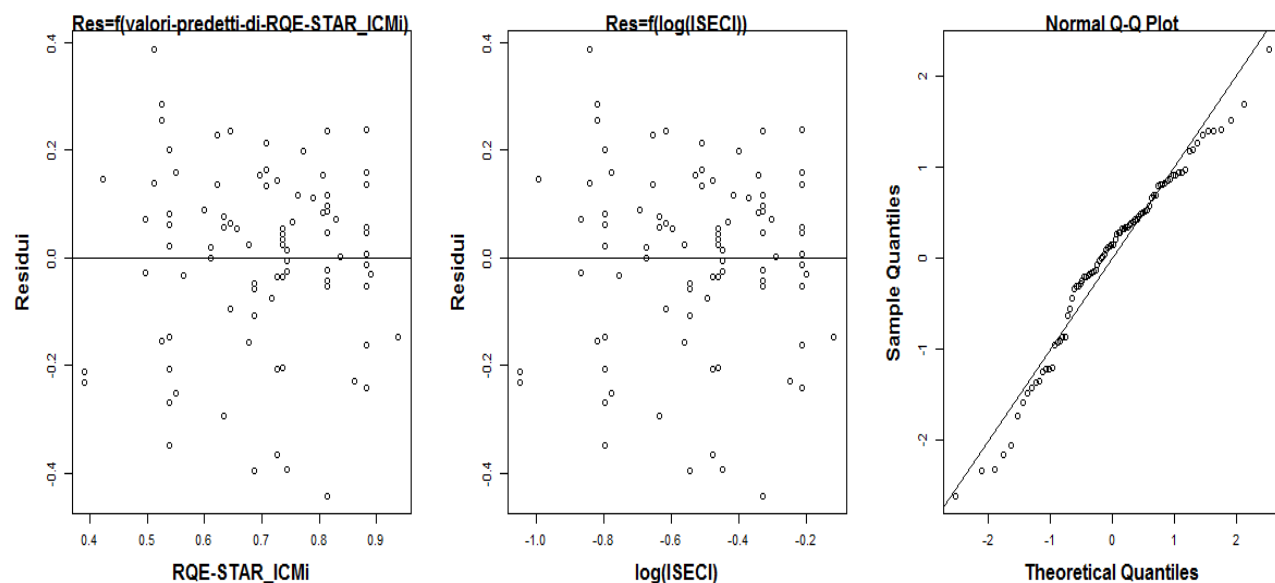


Figura 4.40. Grafici residui e Normal Q-Q Plot per regressione-lineare-logaritmica tra RQE-STAR_ICMi ed ISECI.

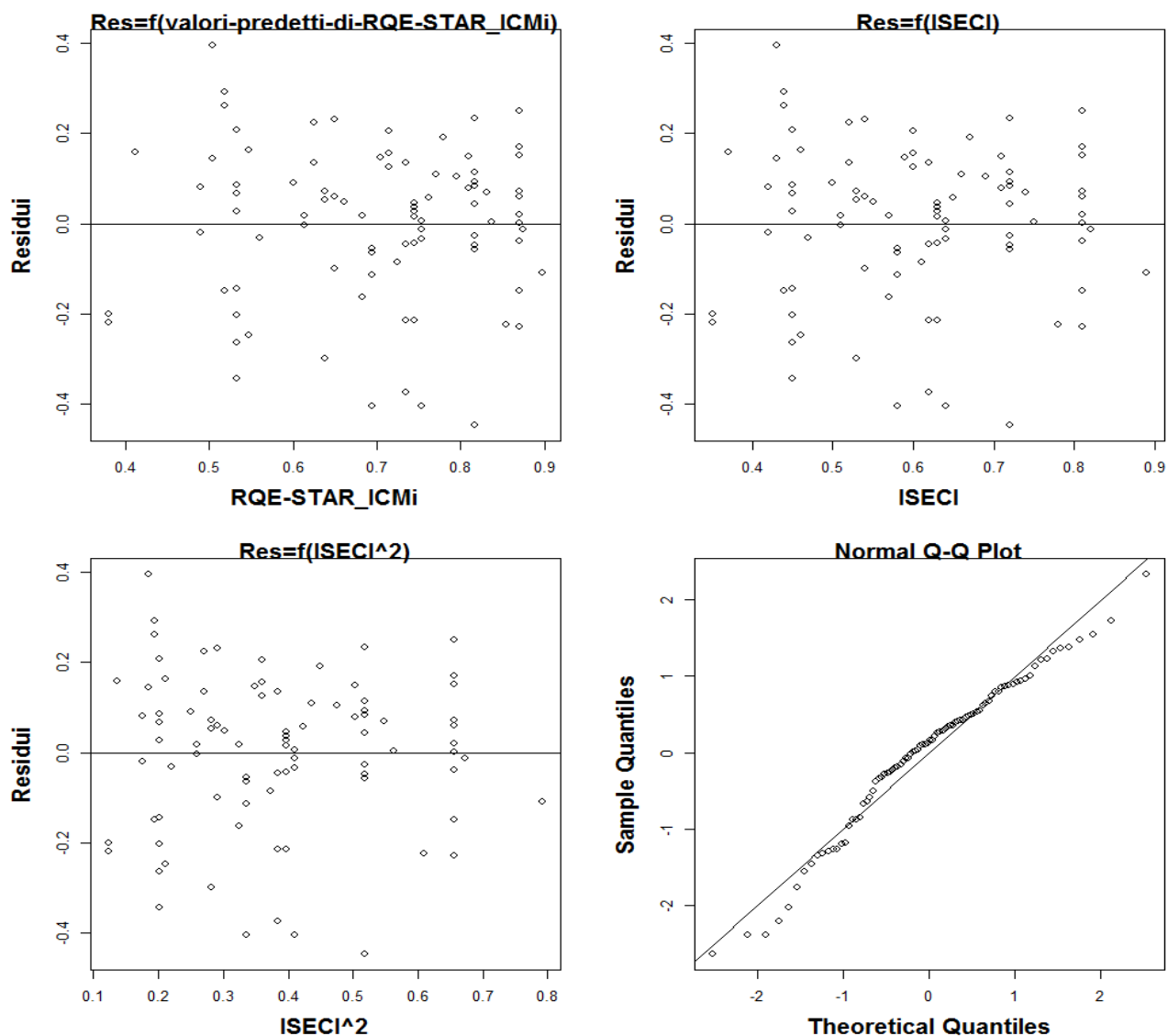


Figura 4.41. Grafici residui e Normal Q-Q Plot per regressione-polinomiale tra RQE-STAR_ICMi e ISECI

	lineare	logaritmica	polinomiale
R2 corretto	0,102	0,110	0,125
Statistica F	10,907	11,711	7,233
p_value_F	1,40E-03	9,50E-04	1,30E-03
p_value_Shapiro	0,027	0,031	0,124
p_value_Breush_Pagan	0,148	0,075	0,006

Tabella 4.16. Confronto dei risultati relativi alla regressione lineare, polinomiale e lineare-logaritmica con il fine di esaminare la relazione tra RQE-STAR_ICMi e OD. In giallo, la trasformazione che inizialmente è stata considerata approssimare bene i dati.

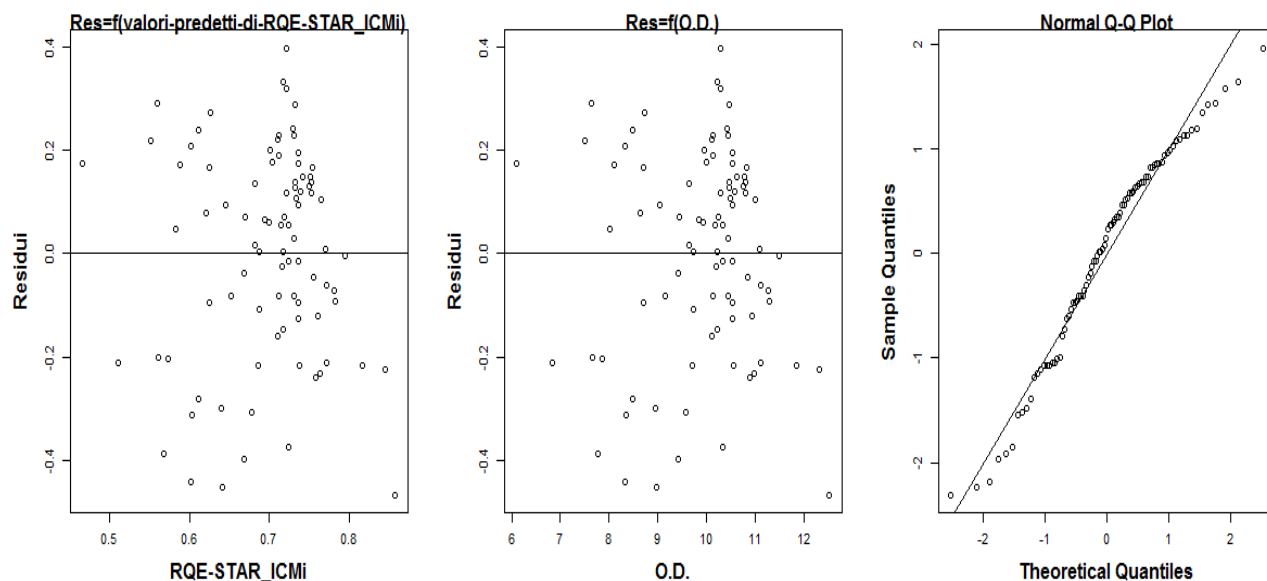


Figura 4.42. Grafici residui e Normal Q-Q Plot per regressione-lineare tra RQE-STAR_ICMi e OD.

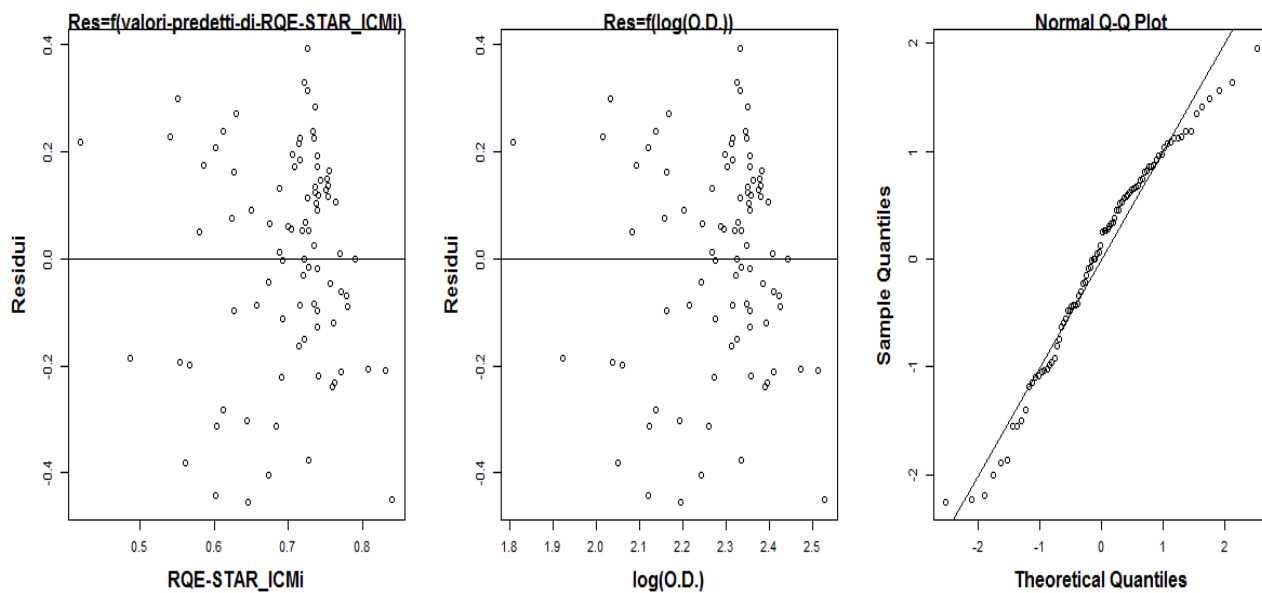


Figura 4.43. Grafici residui e Normal Q-Q Plot per regressione-lineare-logaritmica tra RQE-STAR_ICMi e OD

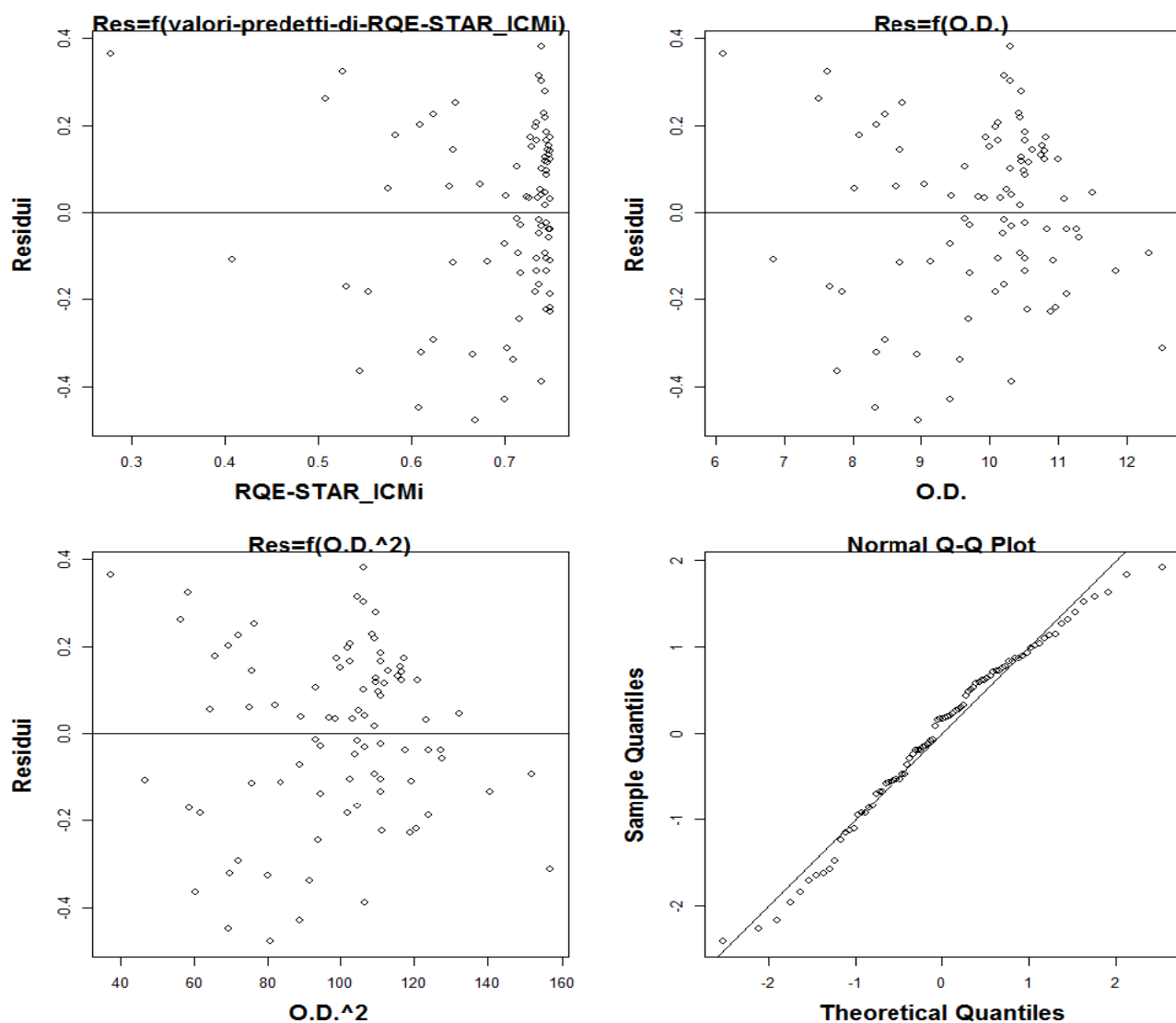


Figura 4.44. Grafici residui e Normal Q-Q Plot per regressione-polinomiale tra RQE-STAR_ICMi e OD

	lineare	logaritmica	polinomiale
R2 corretto	0,343	0,522	0,467
Statistica F	46,375	95,883	39,044
p_value_F	1,24E-06	1,13E-12	9,19E-10
p_value_Shapiro	0,366	0,002	0,009
p_value_Breush_Pagan	0,165	0,436	0,329

Tabella 4.17. Confronto dei risultati relativi alla regressione lineare, polinomiale e lineare-logaritmica con il fine di esaminare la relazione tra RQE-STAR_ICMi e N-NH₄. In giallo, la trasformazione che inizialmente è stata considerata approssimare bene i dati.

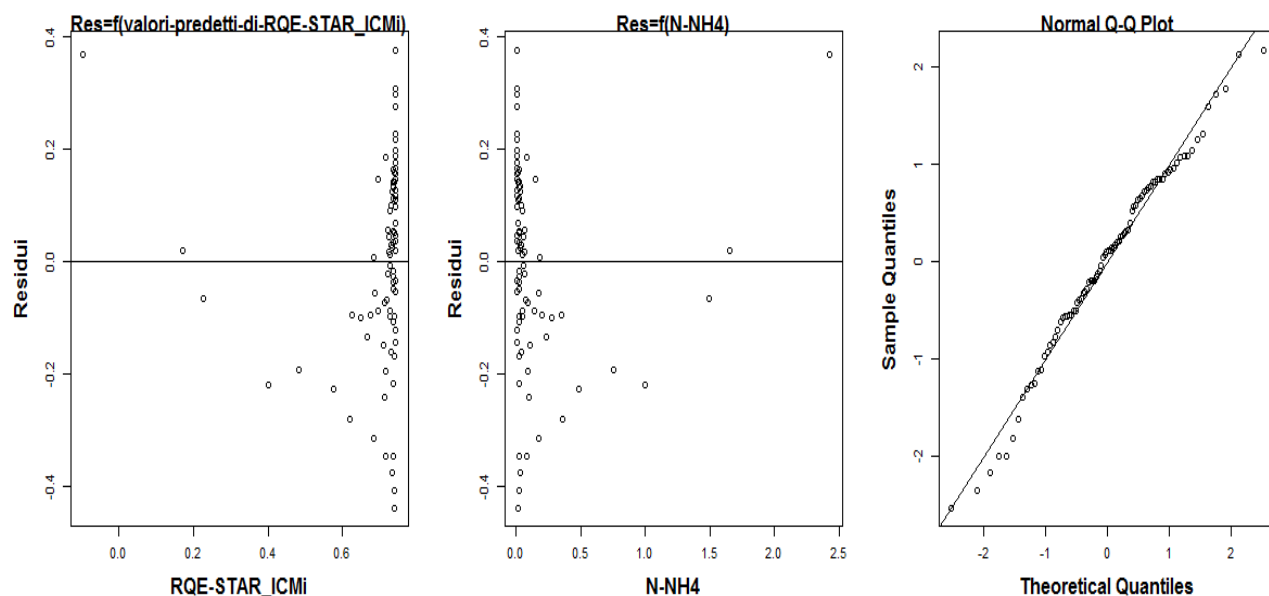


Figura 4.45. Grafici residui e Normal Q-Q Plot per regressione-lineare tra RQE-STAR_ICMi e N-NH₄.

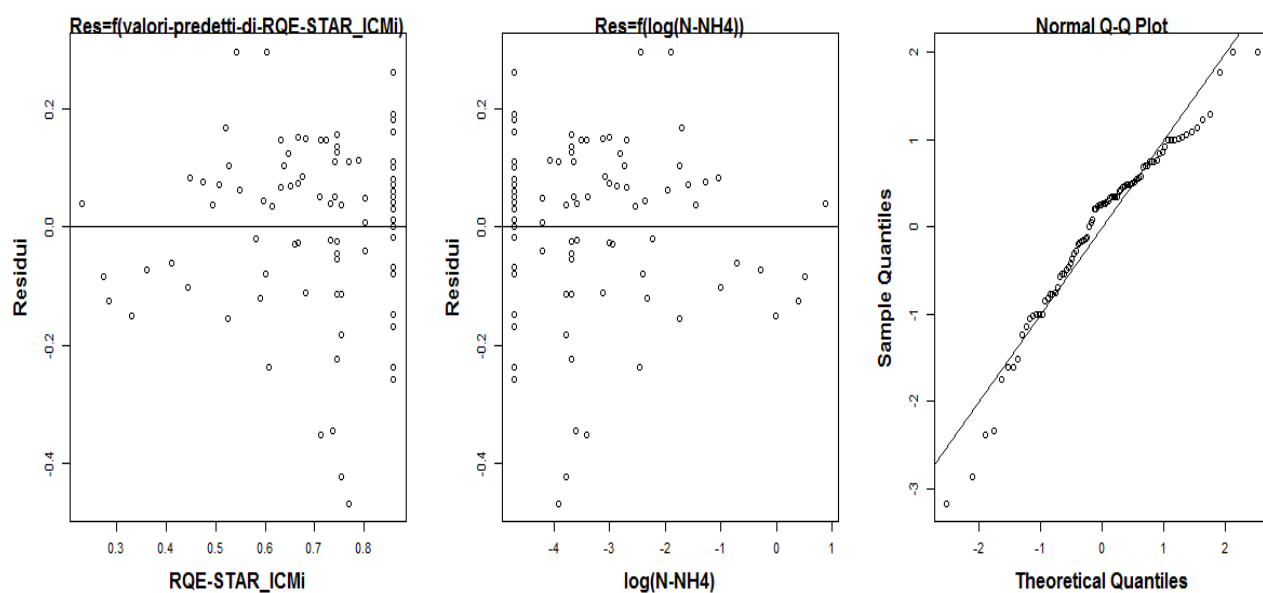


Figura 4.46. Grafici residui e Normal Q-Q Plot per regressione-lineare-logaritmica tra RQE-STAR_ICMi e N-NH₄.

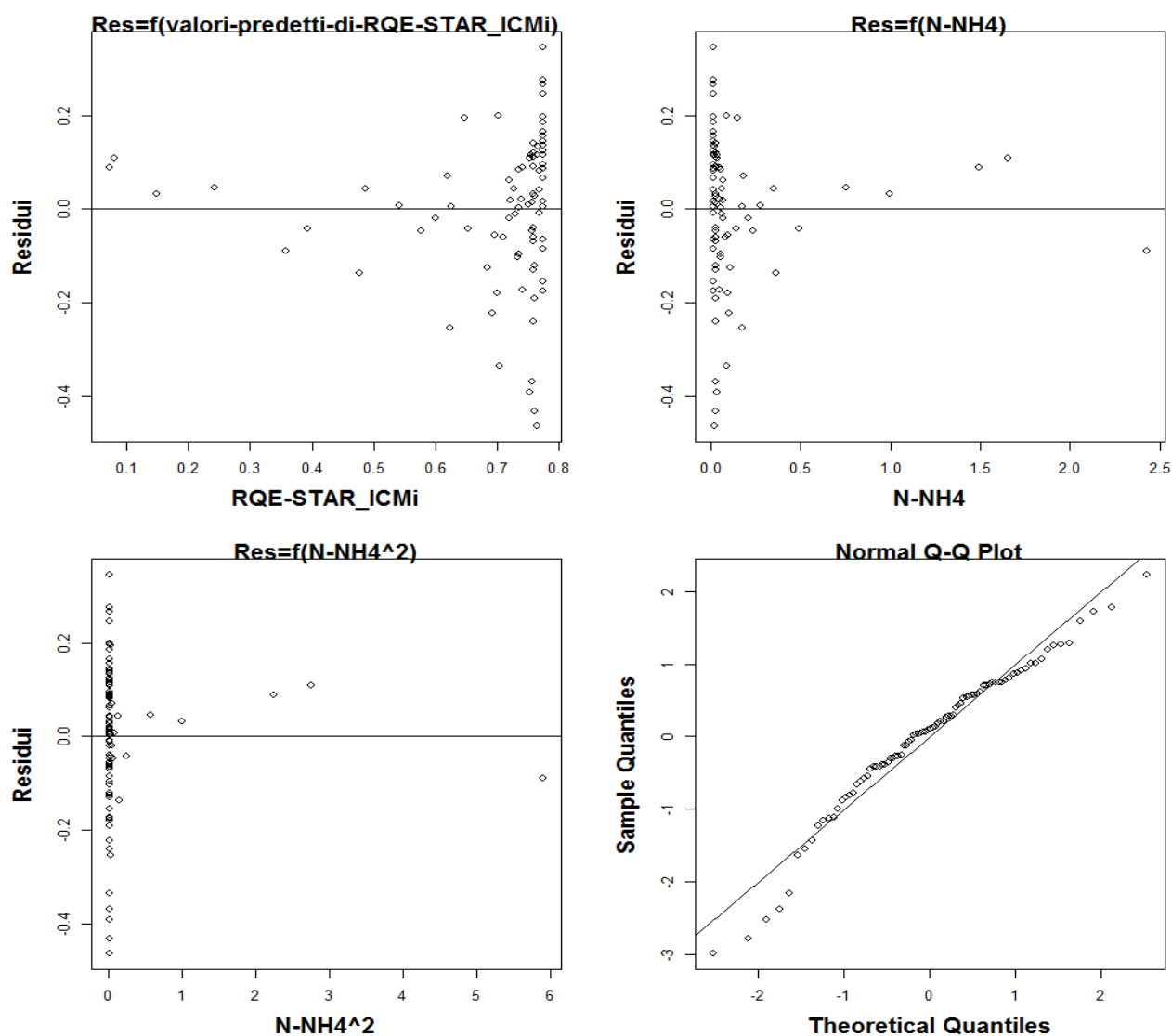


Figura 4.47. Grafici residui e Normal Q-Q Plot per regressione-polinomiale tra RQE-STAR_ICMi e N-NH₄

	lineare	logaritmica	polinomiale
R2 corretto	0,347	0,463	0,395
Statistica F	47,207	76,036	29,444
p_value_F	9,53E-07	1,83E-10	1,92E-07
p_value_Shapiro	0,0004	0,0002	0,0001
p_value_Breush_Pagan	0,701	0,781	0,857

Tabella 4.18. Confronto dei risultati relativi alla regressione lineare, polinomiale e lineare-logaritmica con il fine di esaminare la relazione tra RQE-STAR_ICMi e N-NO₃. In giallo, la trasformazione che inizialmente è stata considerata approssimare bene i dati.

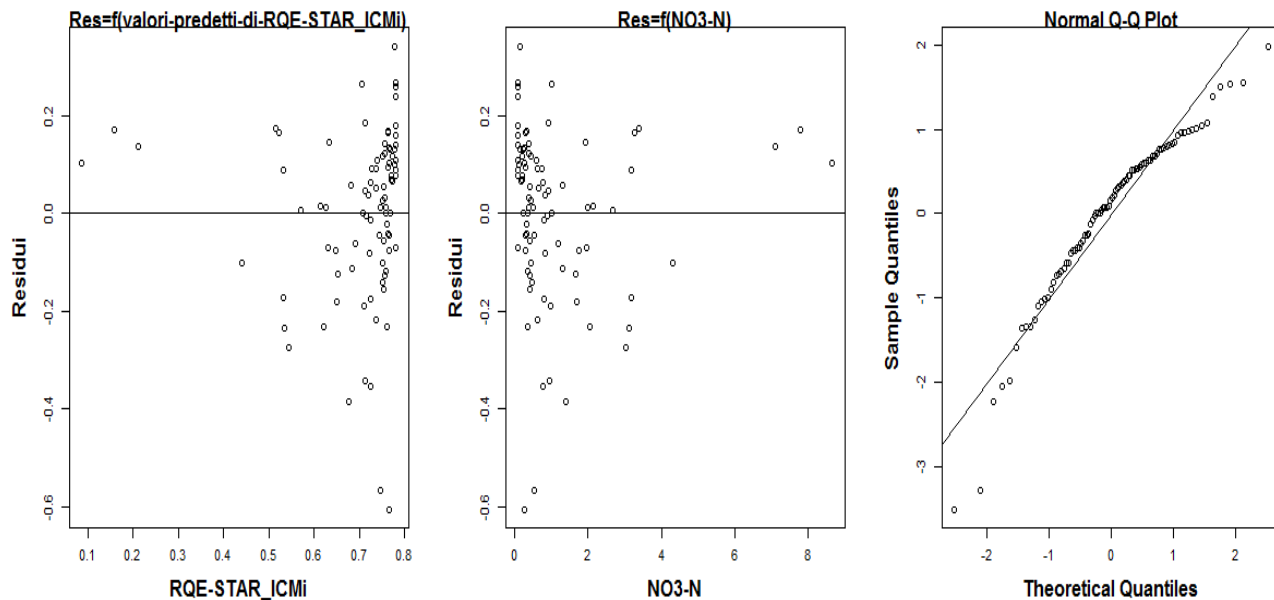


Figura 4.48. Grafici residui e Normal Q-Q Plot per regressione-lineare tra RQE-STAR_ICMi e N-NO_3 .

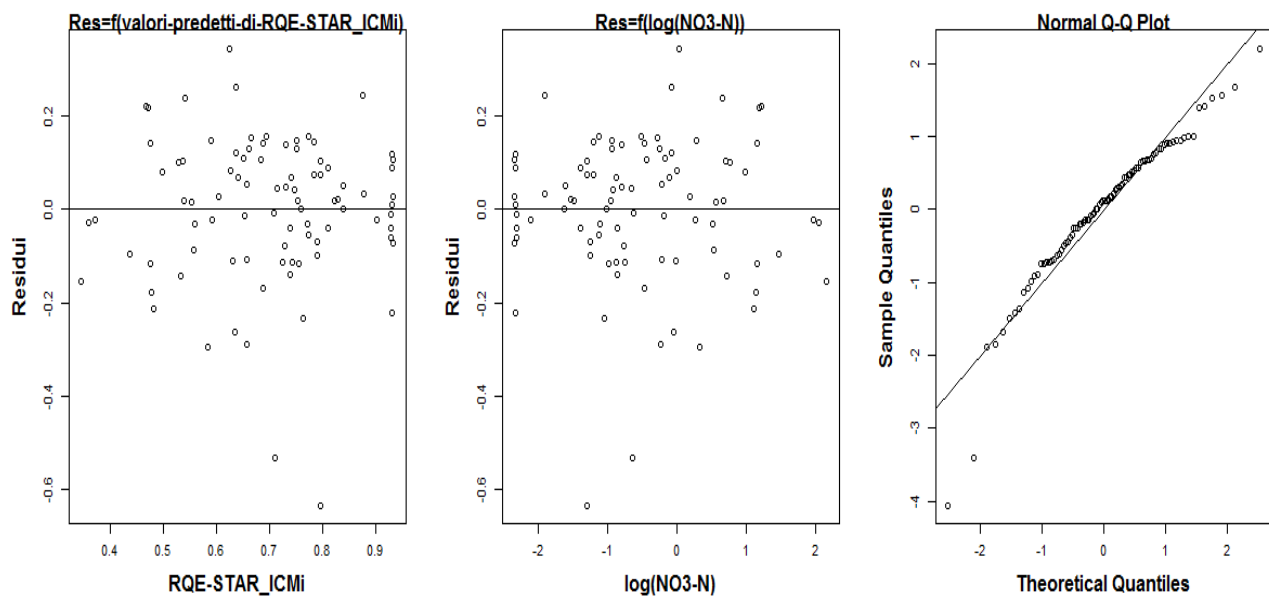


Figura 4.49. Grafici residui e Normal Q-Q Plot per regressione-lineare-logaritmica tra RQE-STAR_ICMi e N-NO_3 .

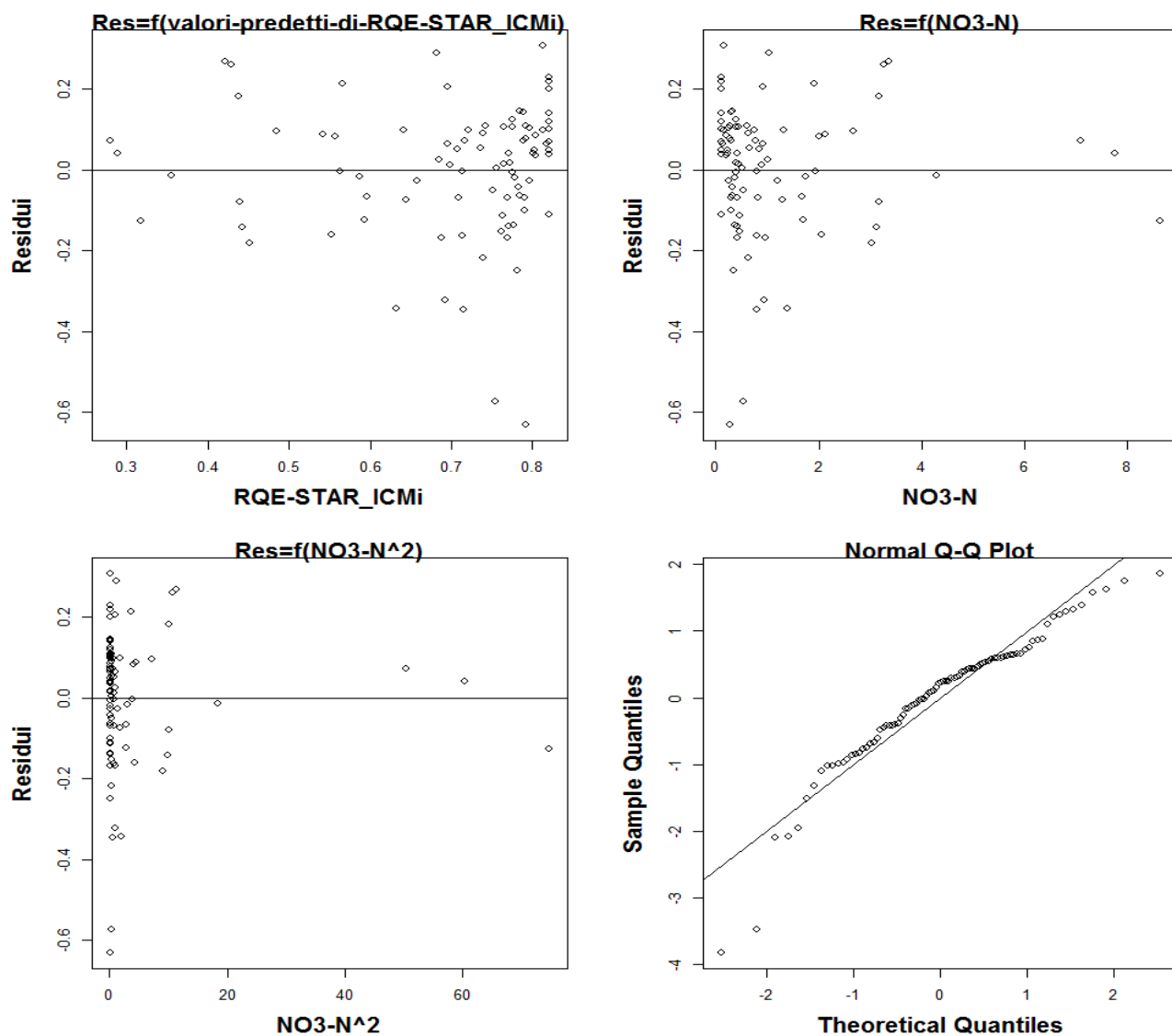


Figura 4.50. Grafici residui e Normal Q-Q Plot per regressione-polinomiale tra RQE-STAR_ICMi e N-NO₃.

	lineare	logaritmica	polinomiale
R2 corretto	0,250	0,331	0,313
Statistica F	30,002	44,103	20,818
p_value_F	4,17E-04	2,57E-06	4,30E-05
p_value_Shapiro	0,0565	0,5638	0,0340
p_value_Breush_Pagan	0,003	0,006	0,003

Tabella 4.19. Confronto dei risultati relativi alla regressione lineare, polinomiale e lineare-logaritmica con il fine di esaminare la relazione tra RQE-STAR_ICMi e PT. In giallo, la trasformazione che inizialmente è stata considerata approssimare bene i dati.

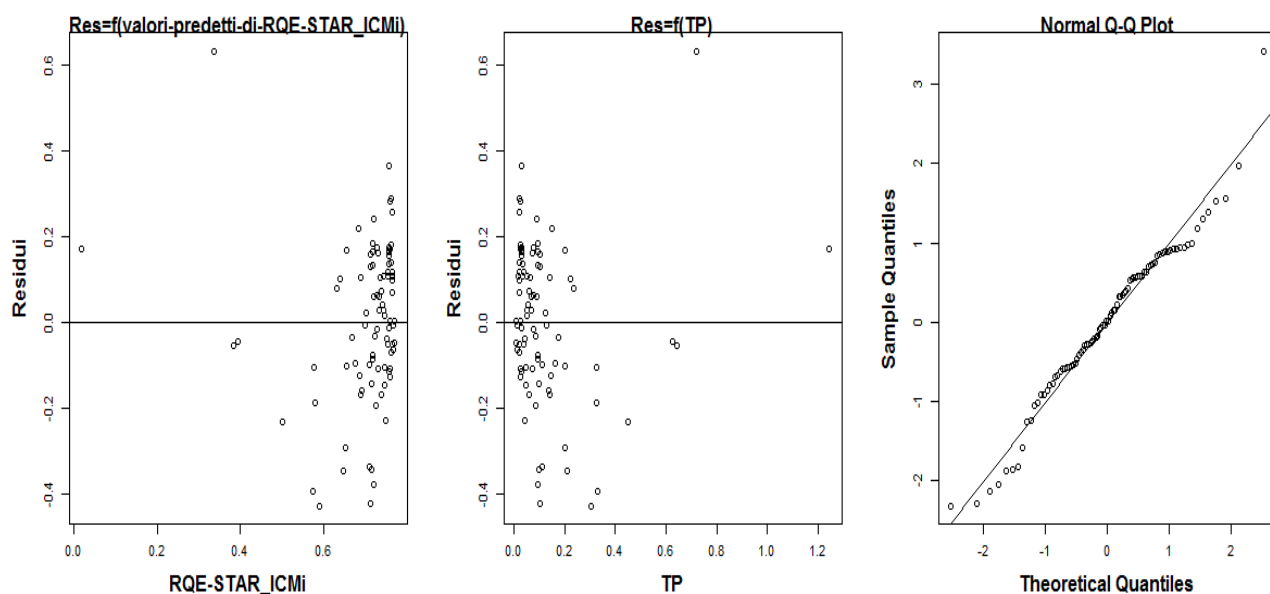


Figura 4.51. Grafici residui e Normal Q-Q Plot per regressione-lineare tra RQE-STAR_ICMi e PT.

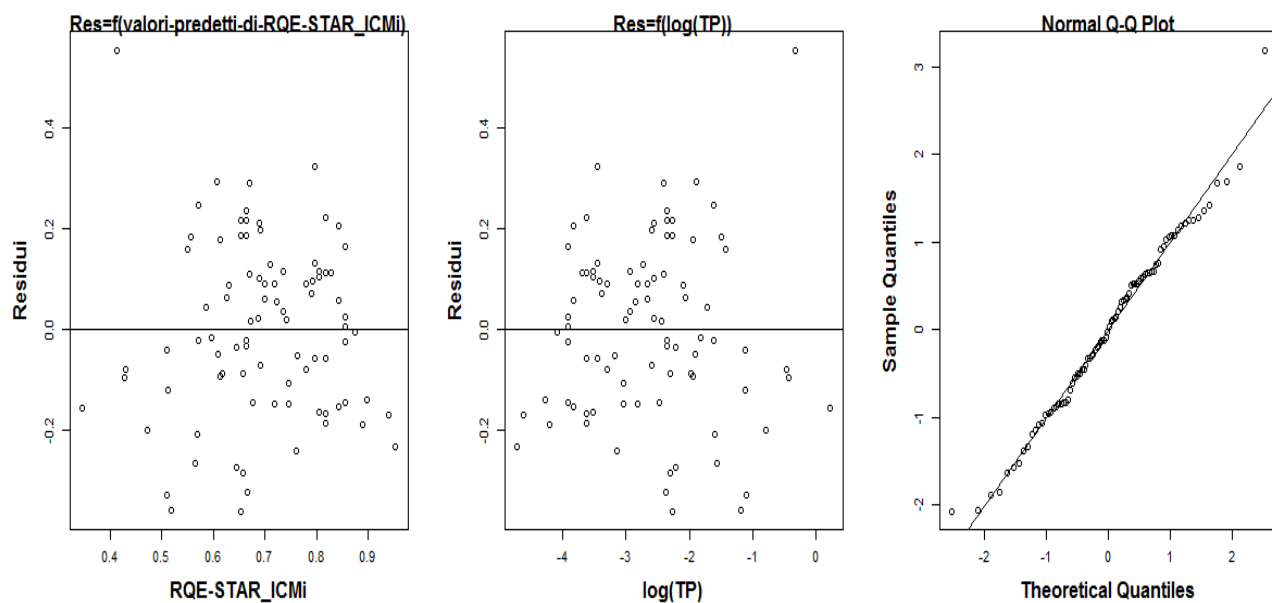


Figura 4.52. Grafici residui e Normal Q-Q Plot per regressione-lineare-logaritmica tra RQE-STAR_ICMi e PT.

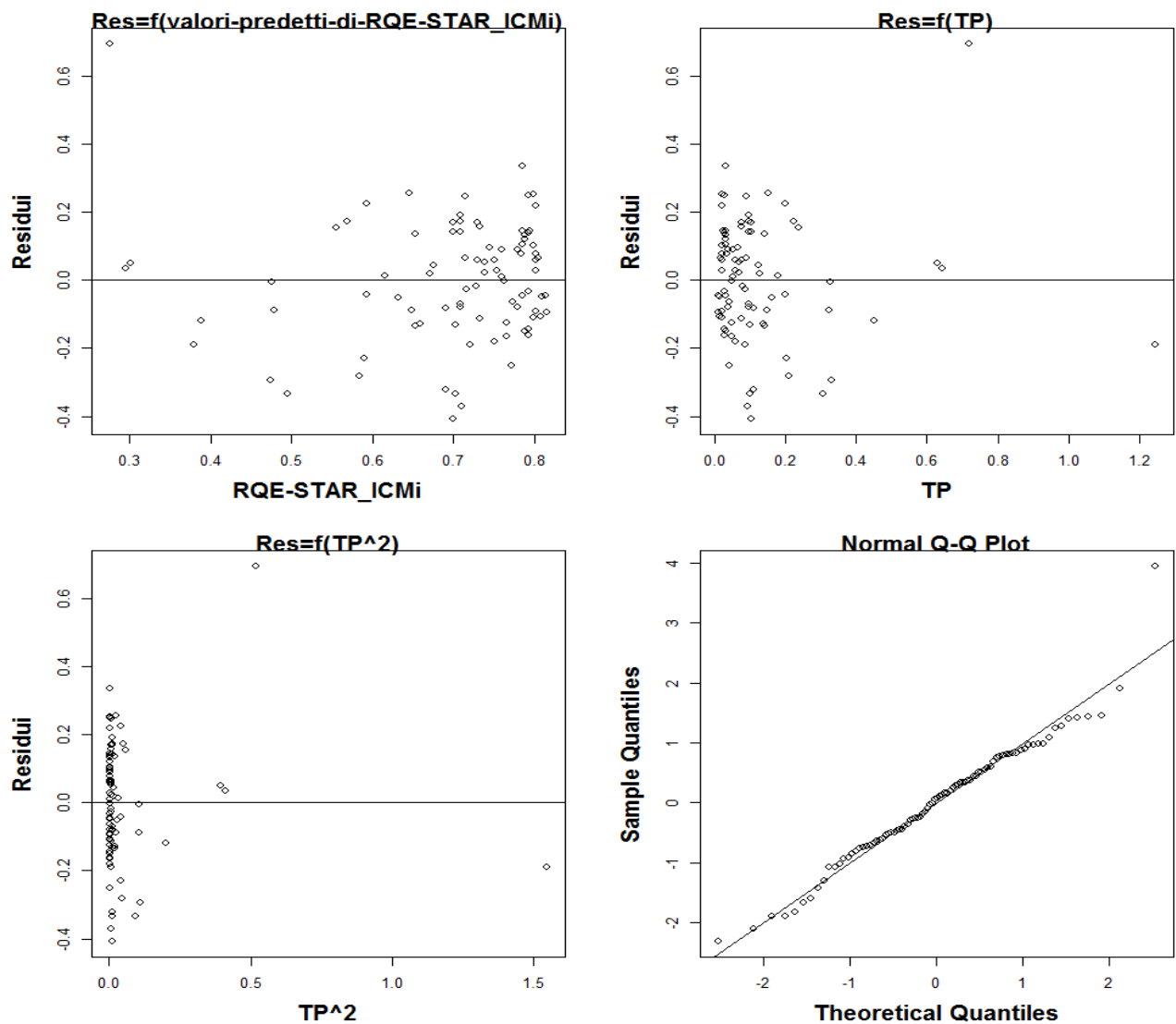


Figura 4.53. Grafici residui e Normal Q-Q Plot per regressione-polinomiale tra RQE-STAR_ICMi e PT.

	lineare	logaritmica	polinomiale
R2 corretto	0,358	0,315	0,365
Statistica F	49,450	41,060	26,000
p_value_F	4,58E-10	7,58E-06	1,55E-09
p_value_Shapiro	0,8810	0,7620	0,6682
p_value_Breush_Pagan	0,978	0,438	0,431

Tabella 4.20. Confronto dei risultati relativi alla regressione lineare, polinomiale e lineare-logaritmica con il fine di esaminare la relazione tra ISECI e RQE-STAR_ICMi. In giallo, la trasformazione che inizialmente è stata considerata approssimare bene i dati.

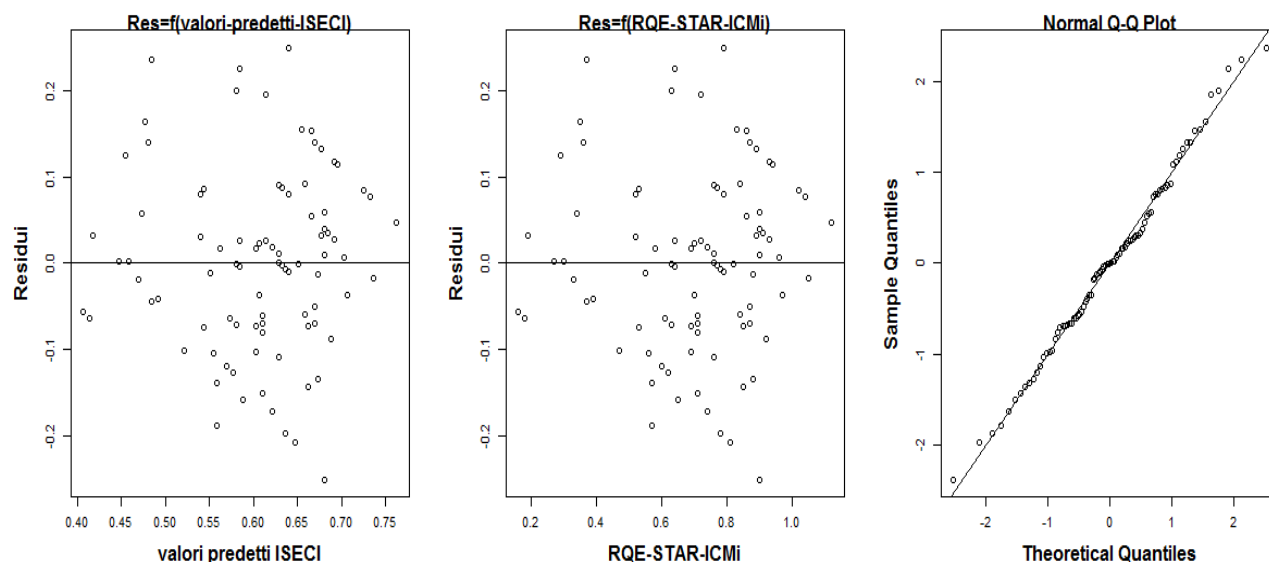


Figura 4.54. Grafici residui e Normal Q-Q Plot per regressione-lineare tra ISECI e RQE-STAR_ICMi.

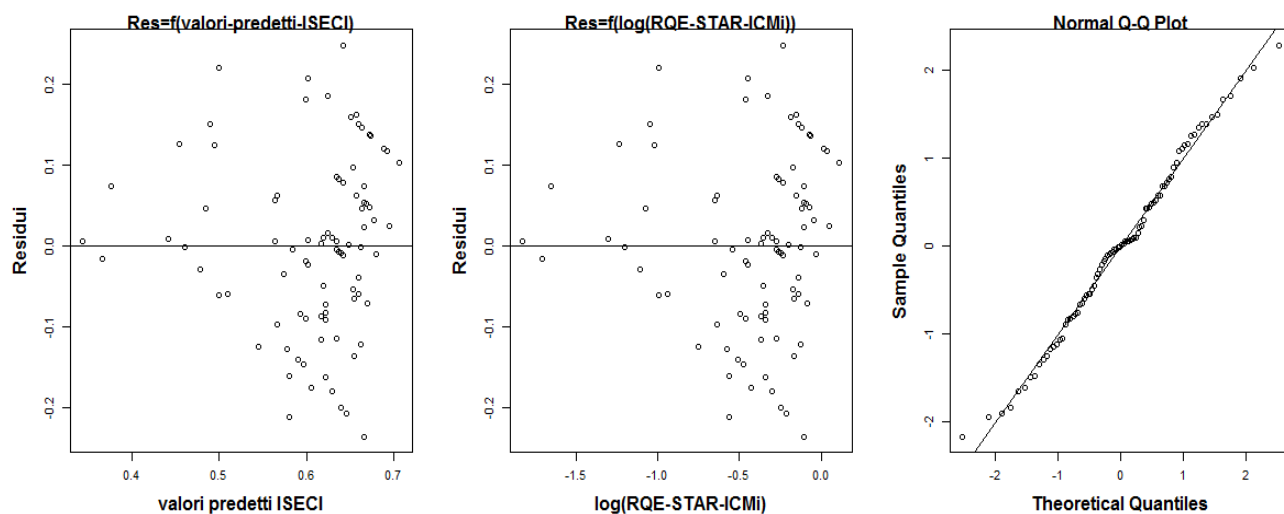


Figura 4.55. Grafici residui e Normal Q-Q Plot per regressione-lineare-logaritmica tra ISECI e RQE-STAR_ICMi.

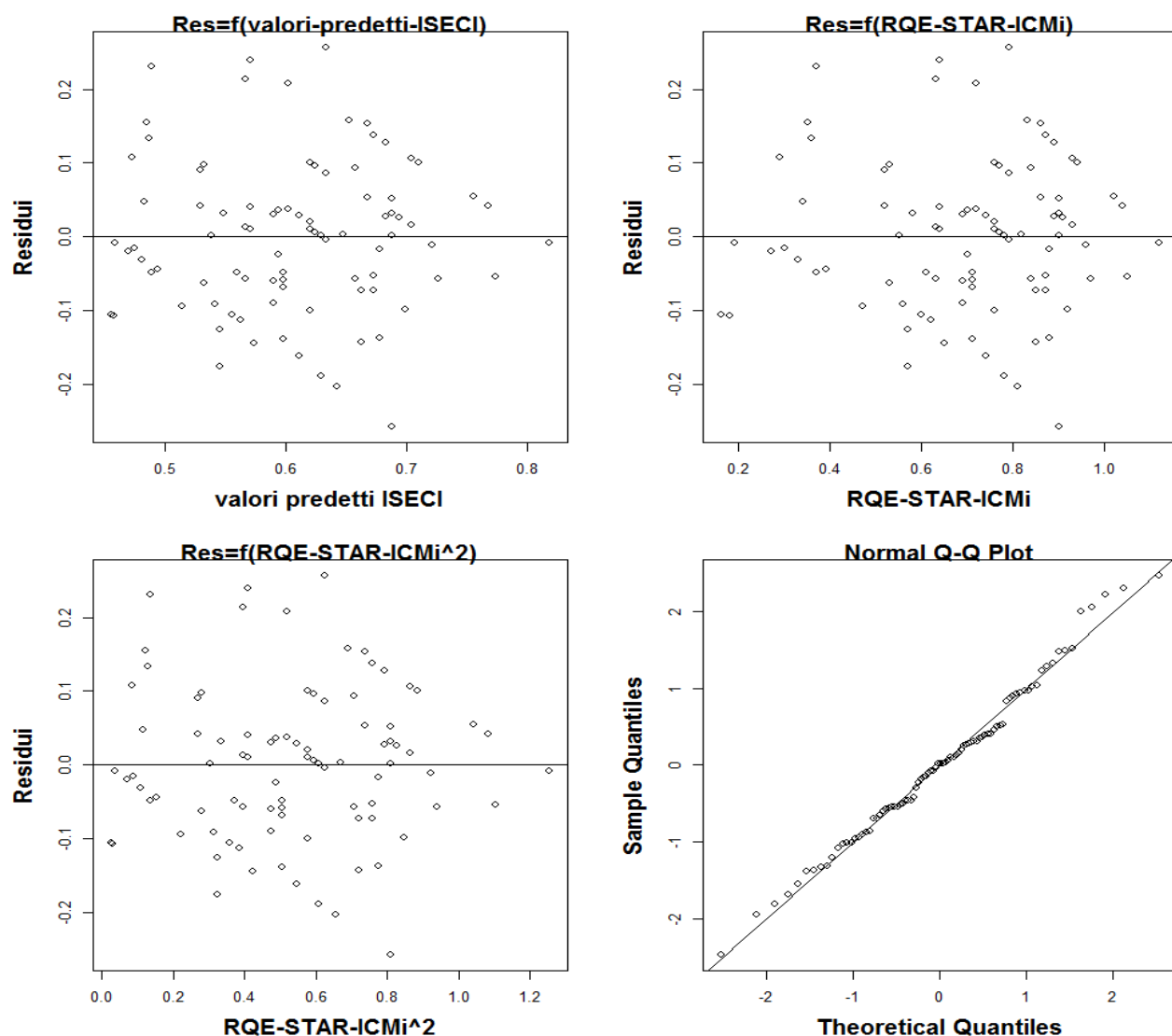


Figura 4.56. Grafici residui e Normal Q-Q Plot per regressione-polinomiale tra ISECI e RQE-STAR_ICMi.

	lineare	logaritmica	polinomiale
R2 corretto	0,174	0,185	0,182
Statistica F	19,280	20,750	10,700
p_value_F	3,20E-05	1,72E-05	7,18E-05
p_value_Shapiro	0,0398	0,0293	0,0276
p_value_Breush_Pagan	0,211	0,214	0,264

Tabella 4.21. Confronto dei risultati relativi alla regressione lineare, polinomiale e lineare-logaritmica con il fine di esaminare la relazione tra ISECI e RQE-IBMR. In giallo, la trasformazione che inizialmente è stata considerata approssimare bene i dati.

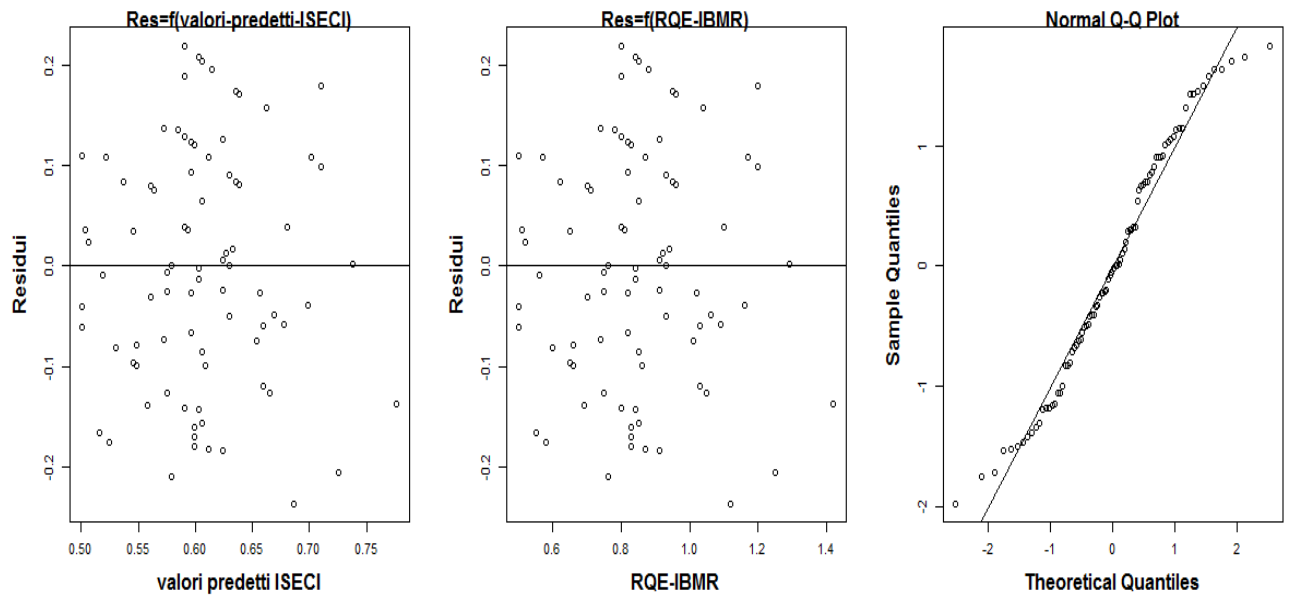


Figura 4.57. Grafici residui e Normal Q-Q Plot per regressione-lineare tra ISEC e RQE-IBMR.

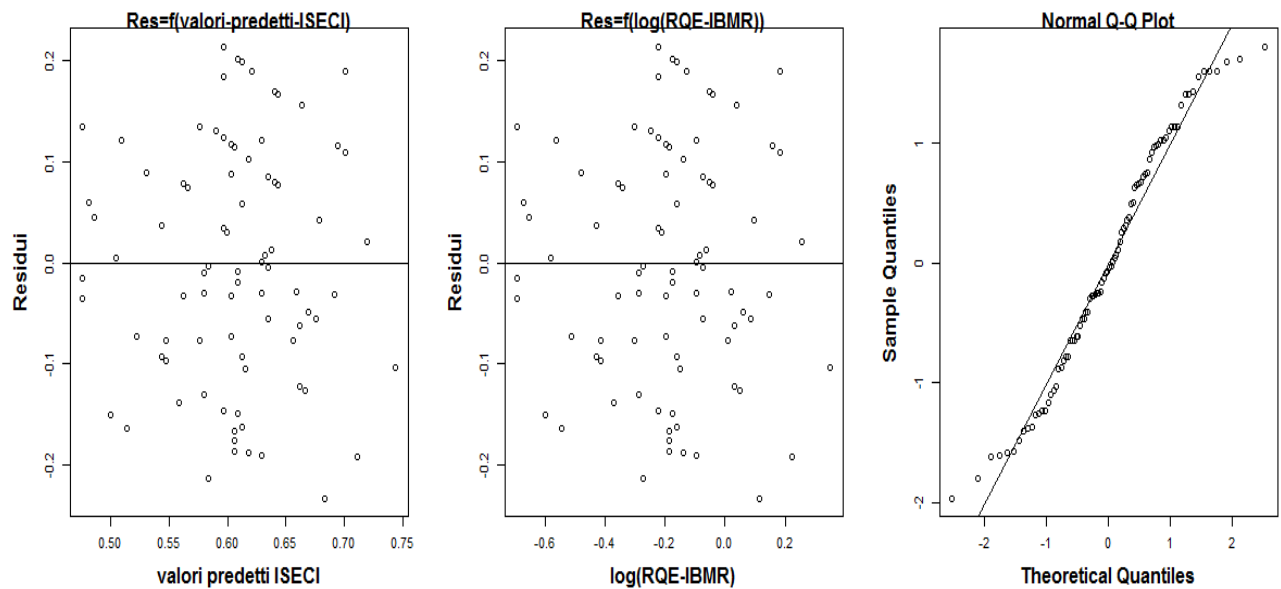


Figura 4.58. Grafici residui e Normal Q-Q Plot per regressione-lineare-logaritmica tra ISEC e RQE-IBMR.

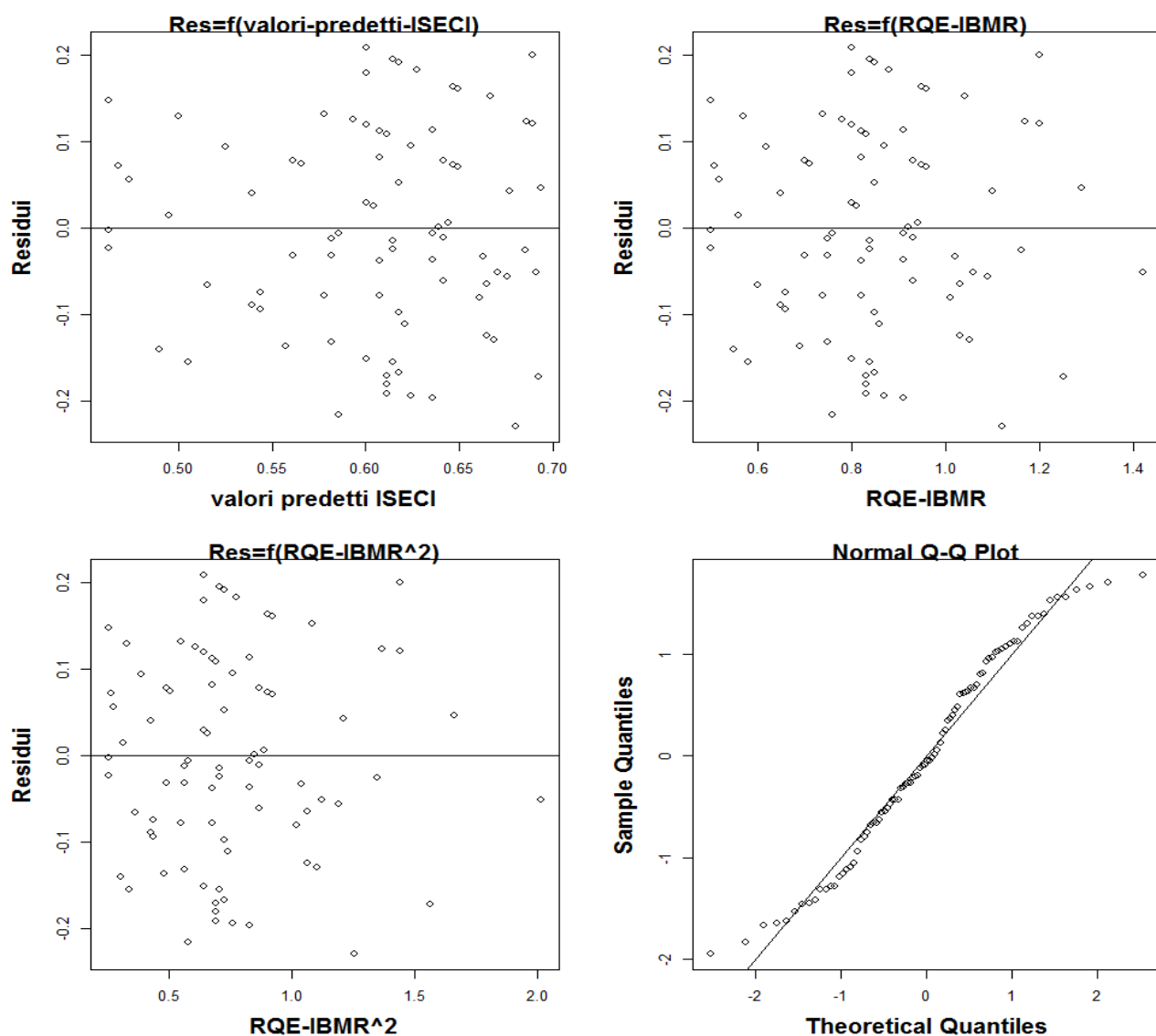


Figura 4.59. Grafici residui e Normal Q-Q Plot per regressione-polinomiale tra ISECI e RQE-IBMR.

	lineare	logaritmica	polinomiale
R2 corretto	0,016	0,044	0,114
Statistica F	2,398	4,960	6,603
p_value_F	0,1251	0,0286	0,0022
p_value_Shapiro	0,0190	0,0195	0,0229
p_value_Breush_Pagan	0,320	0,838	0,146

Tabella 4.22. Confronto dei risultati relativi alla regressione lineare, polinomiale e lineare-logaritmica con il fine di esaminare la relazione tra ISECI e RQE-ICMi. In giallo, la trasformazione che inizialmente è stata considerata approssimare bene i dati.

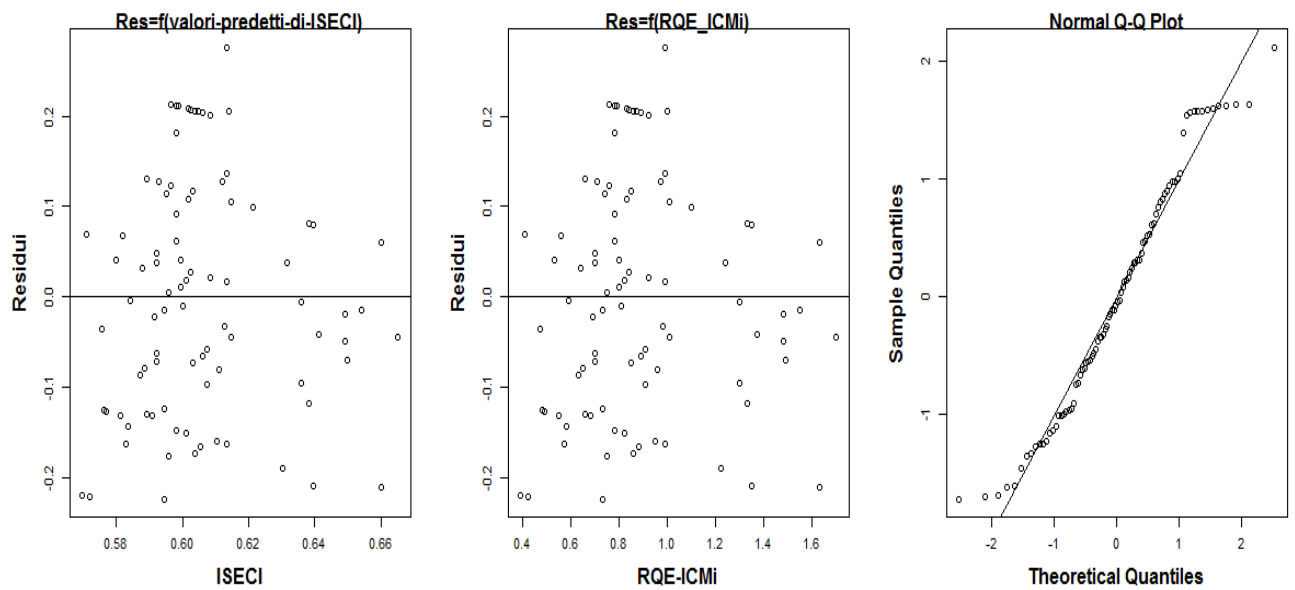


Figura 4.60. Grafici residui e Normal Q-Q Plot per regressione-lineare tra ISECI e RQE-ICMi.

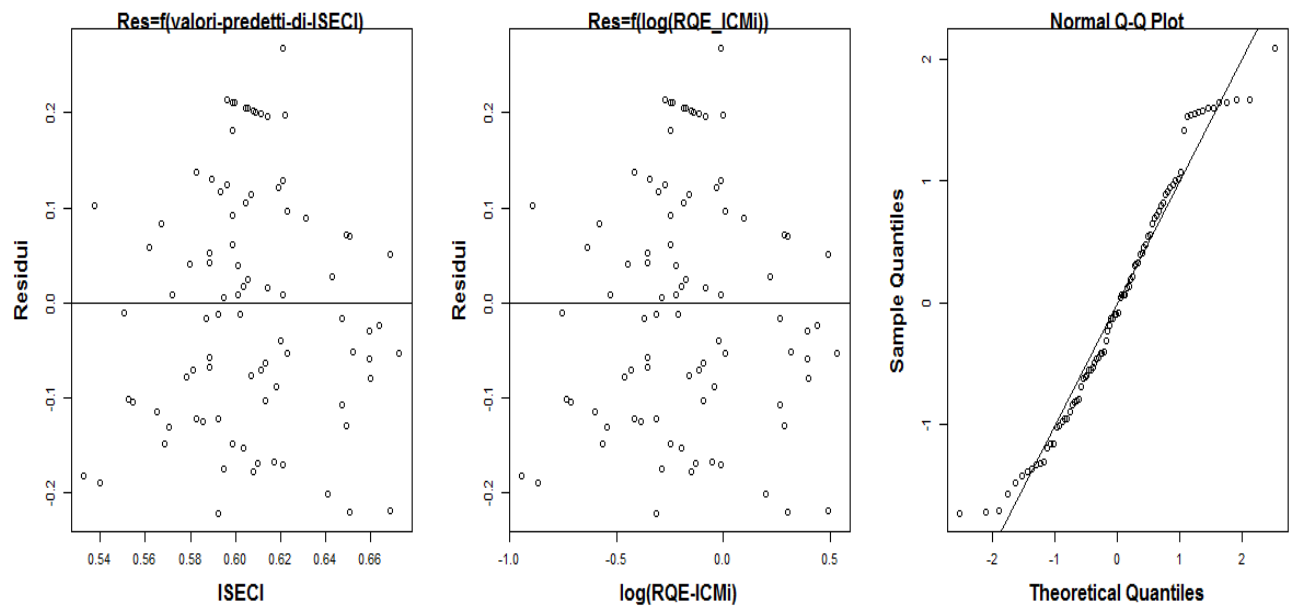


Figura 4.61. Grafici residui e Normal Q-Q Plot per regressione-lineare-logaritmica tra ISECI e RQE-ICMi.

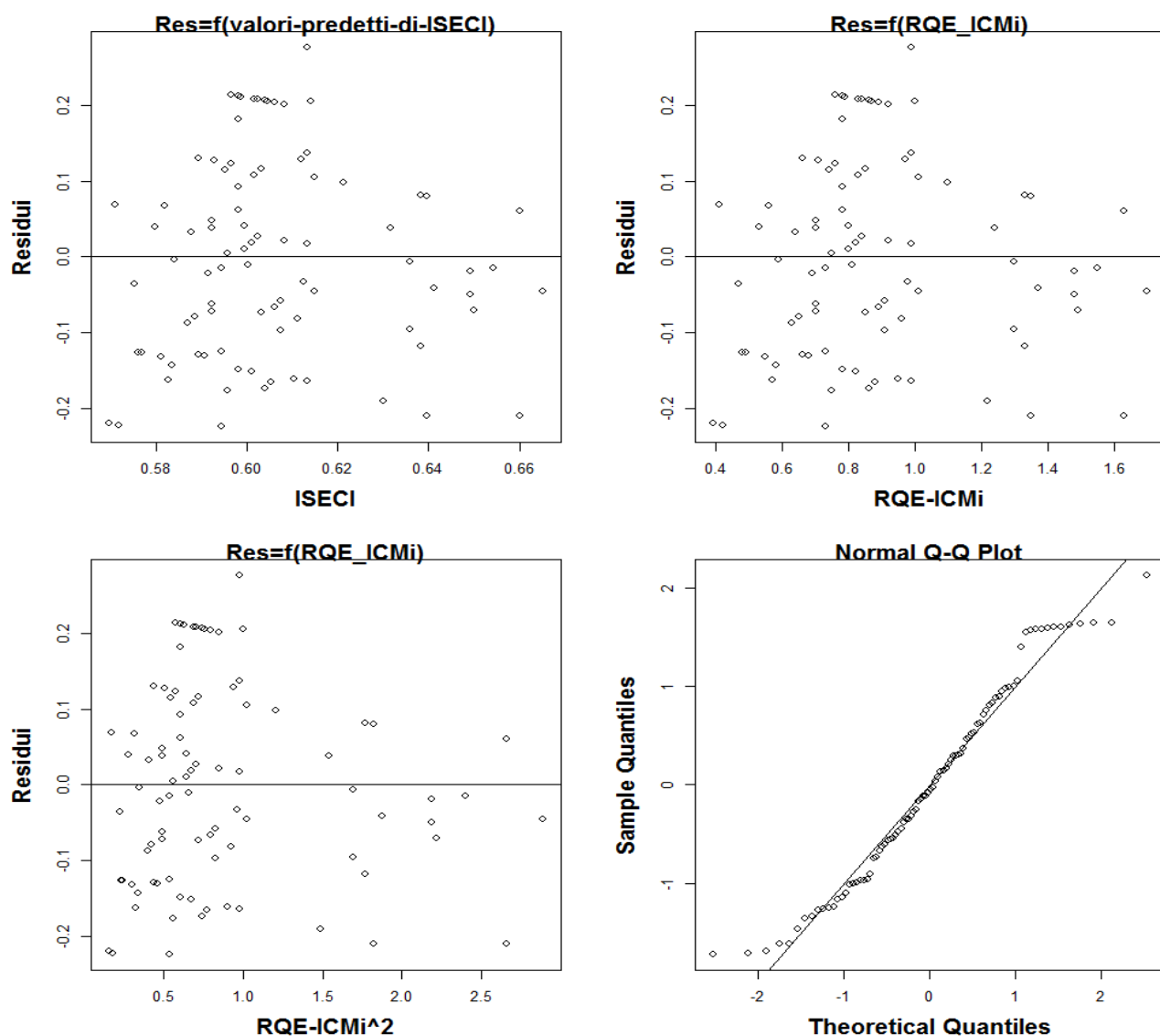


Figura 4.62. Grafici residui e Normal Q-Q Plot per regressione-polinomiale tra ISECI e RQE-ICMi.

	lineare	logaritmica	polinomiale
R2 corretto	0,098	0,105	0,111
Statistica F	10,490	11,210	6,413
p_value_F	0,0017	0,0012	0,0025
p_value_Shapiro	0,0179	0,2389	0,0358
p_value_Breush_Pagan	0,286	0,116	0,152

Tabella 4.23. Confronto dei risultati relativi alla regressione lineare, polinomiale e lineare-logaritmica con il fine di esaminare la relazione tra ISECI e N-NH₄. In giallo, la trasformazione che inizialmente è stata considerata approssimare bene i dati.

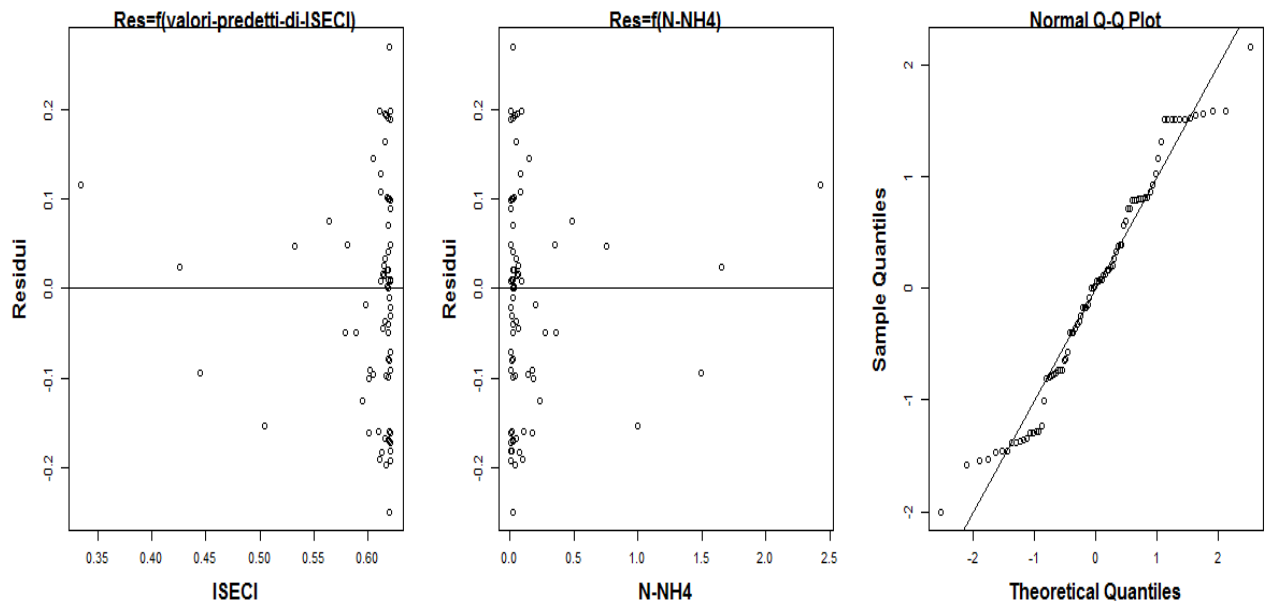


Figura 4.63. Grafici residui e Normal Q-Q Plot per regressione-lineare tra ISECI e N-NH₄.

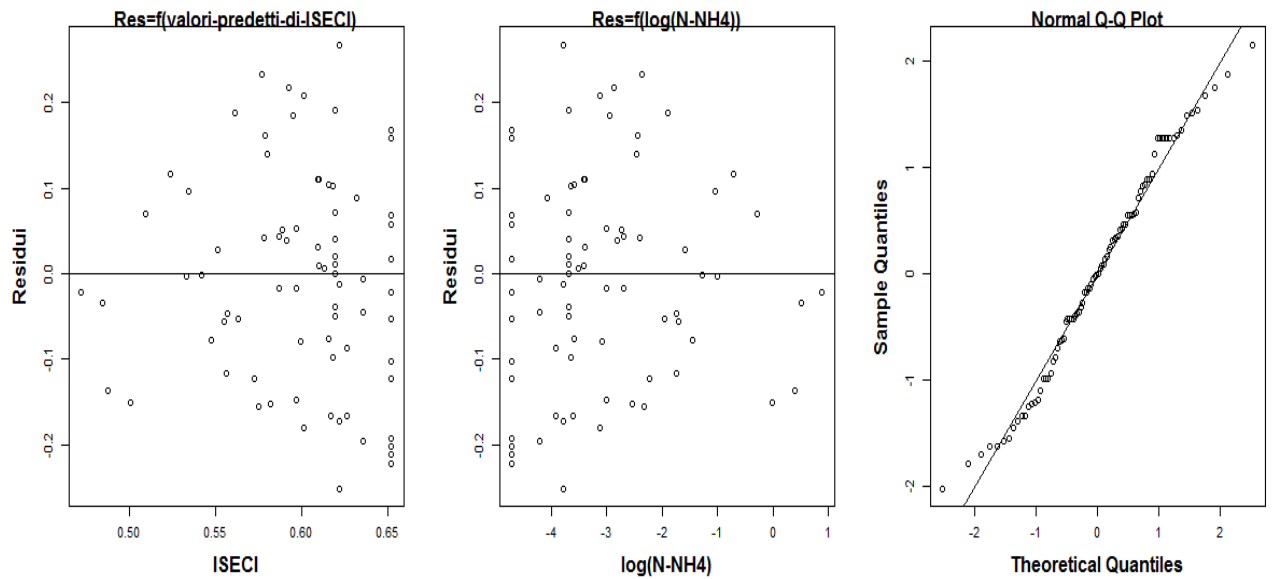


Figura 4.64. Grafici residui e Normal Q-Q Plot per regressione-lineare-logaritmica tra ISECI e N-NH₄.

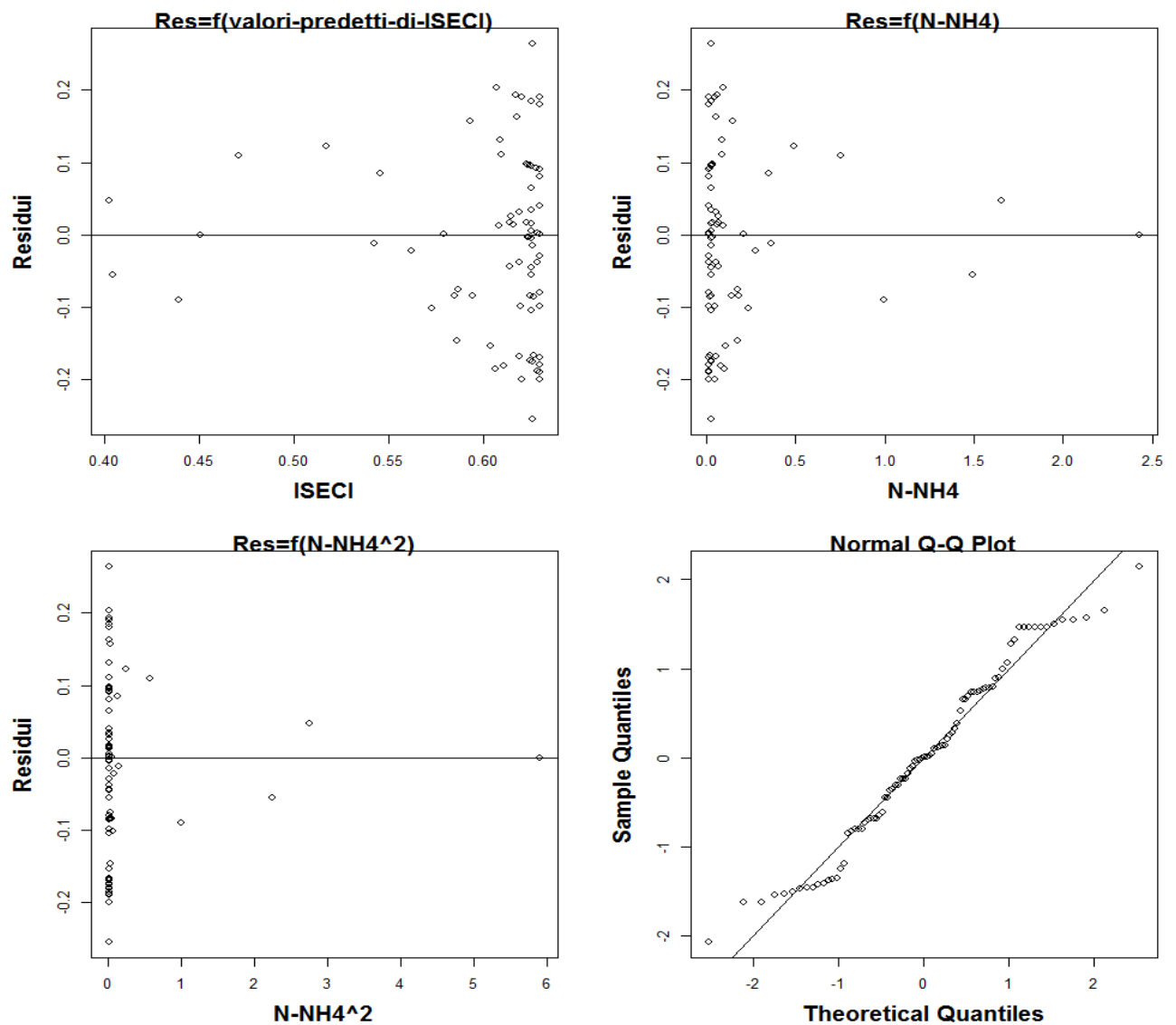


Figura 4.65. Grafici residui e Normal Q-Q Plot per regressione-polinomiale tra ISECI e N-NH₄.

	lineare	logaritmica	polinomiale
R2 corretto	0,1082	0,2244	0,166
Statistica F	11,56	26,18	9,657
p_value_F	0,001026	1,87E-06	0,0001662
p_value_Shapiro	0,2097	0,5575	0,1859
p_value_Breush_Pagan	0,2042	0,2379	0,1698

Tabella 4.24. Confronto dei risultati relativi alla regressione lineare, polinomiale e lineare-logaritmica con il fine di esaminare la relazione tra ISECI e N-NO₃. In giallo, la trasformazione che inizialmente è stata considerata approssimare bene i dati.

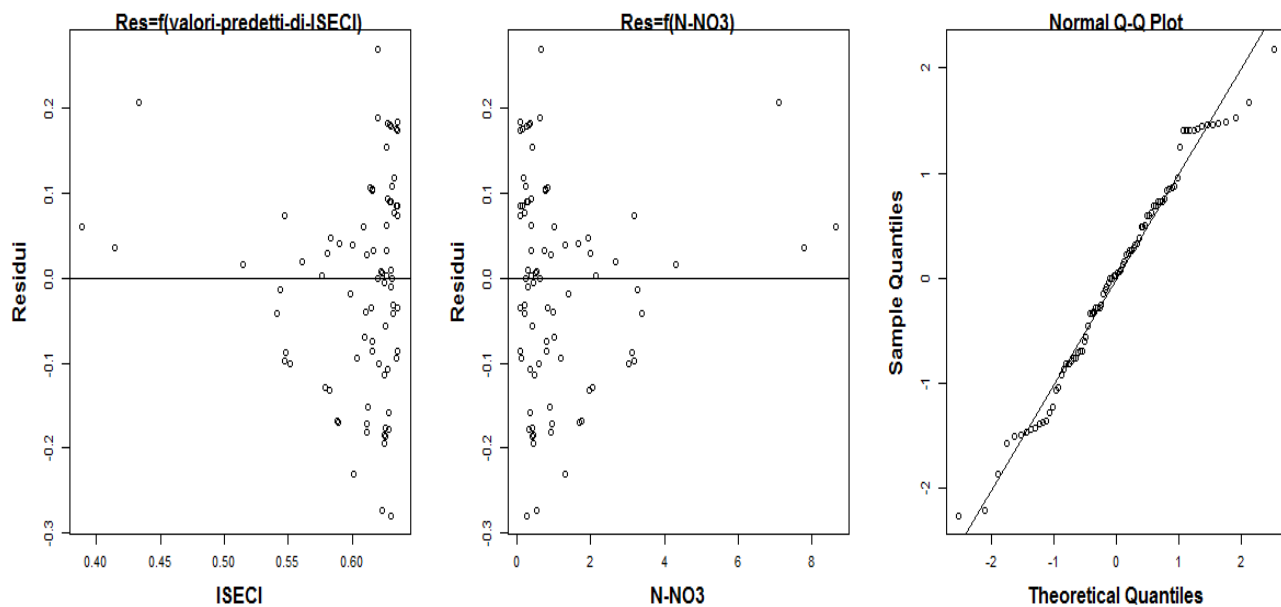


Figura 4.66. Grafici residui e Normal Q-Q Plot per regressione-lineare tra ISECI e $N-NO_3$.

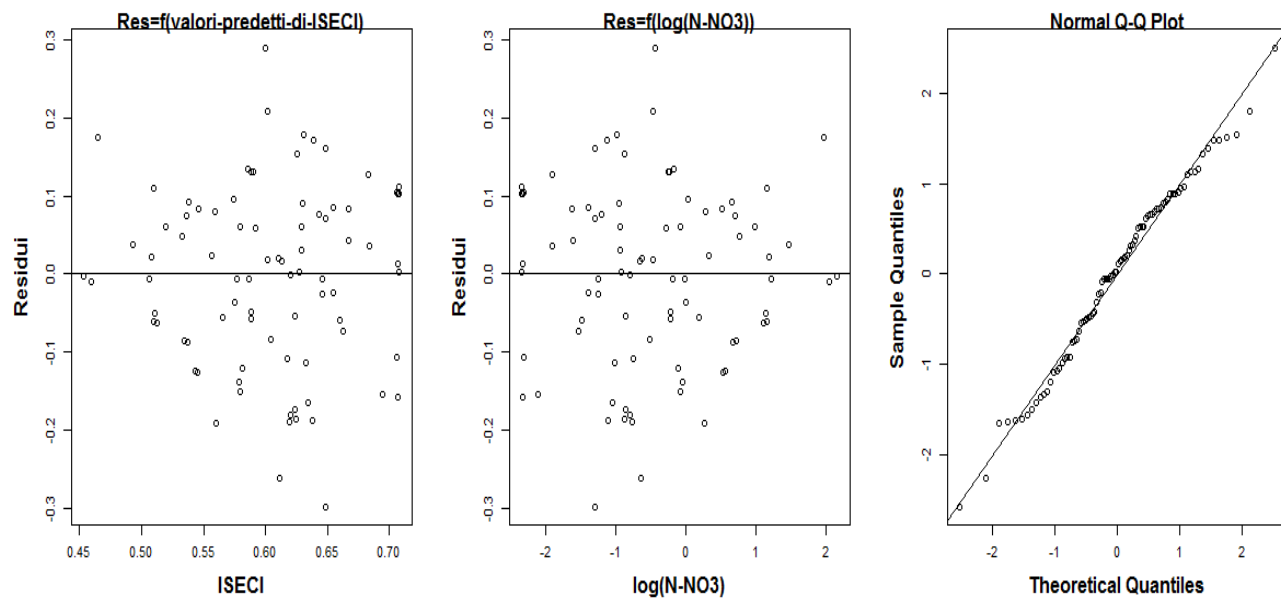


Figura 4.67. Grafici residui e Normal Q-Q Plot per regressione-lineare-logaritmica tra ISECI e $N-NO_3$.

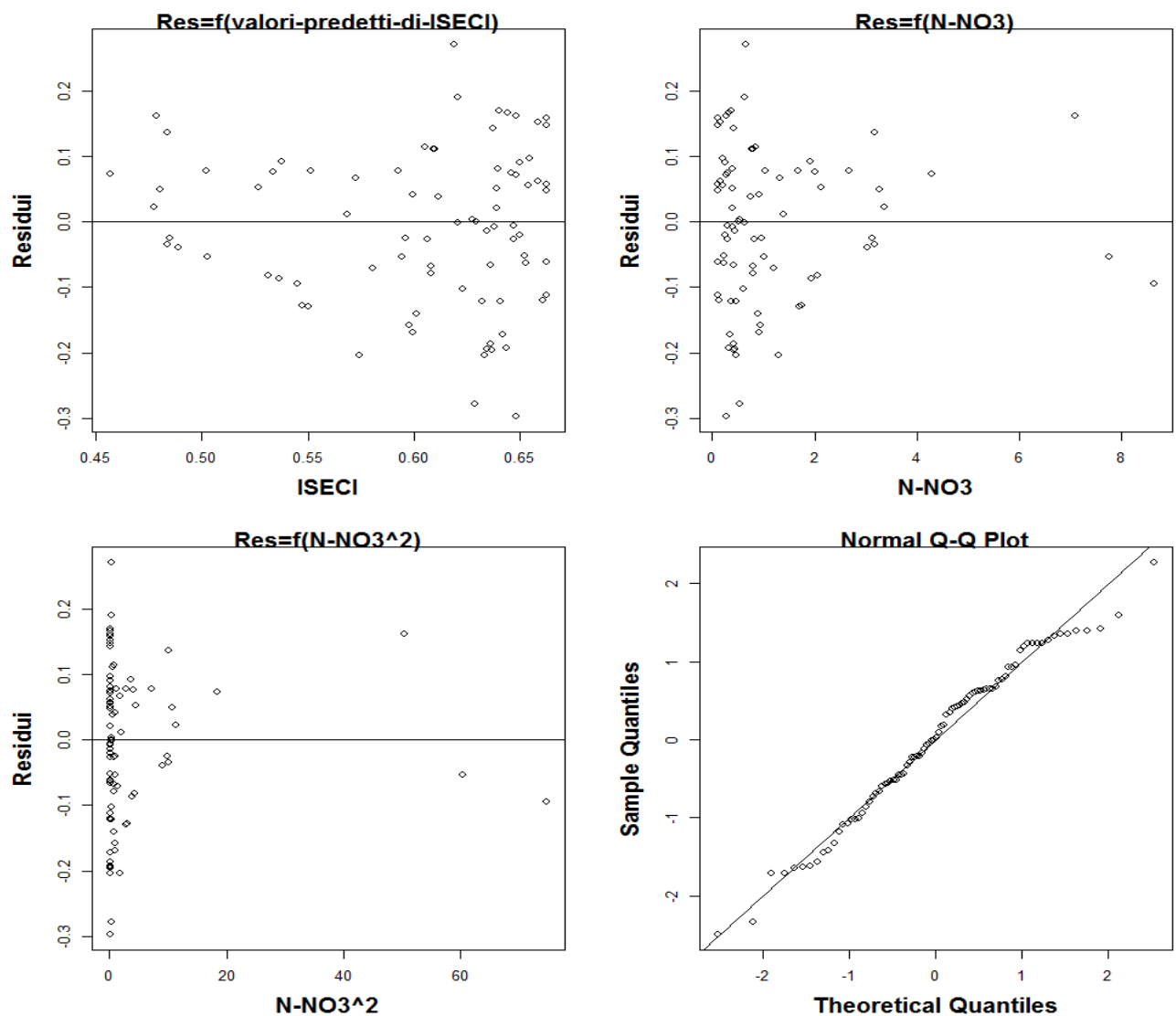


Figura 4.68. Grafici residui e Normal Q-Q Plot per regressione-polinomiale tra ISECI e N-NO₃.

	lineare	logaritmica	polinomiale
R2 corretto	0,011	0,015	0,045
Statistica F	1,939	2,354	3,058
p_value_F	0,167	0,129	0,052
p_value_Shapiro	0,022	0,02	0,043
p_value_Breush_Pagan	0,403	0,427	0,508

Tabella 4.25. Confronto dei risultati relativi alla regressione lineare, polinomiale e lineare-logaritmica con il fine di esaminare la relazione tra ISECI e OD. In giallo, la trasformazione che inizialmente è stata considerata approssimare bene i dati.

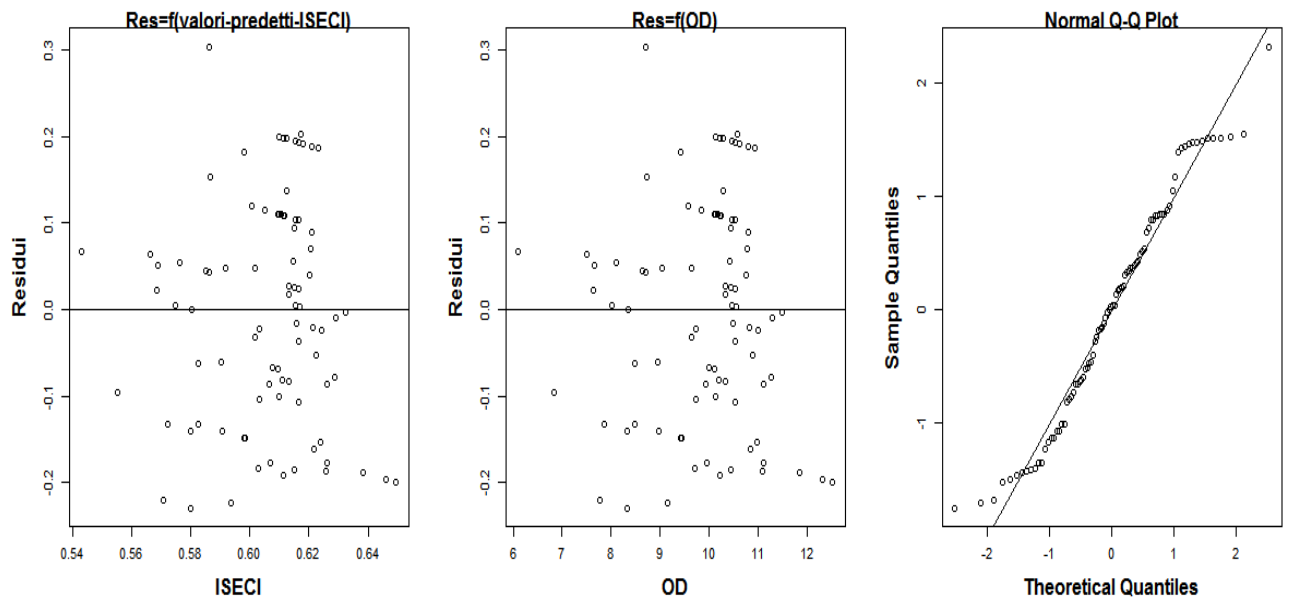


Figura 4.69. Grafici residui e Normal Q-Q Plot per regressione-lineare tra ISECI e OD.

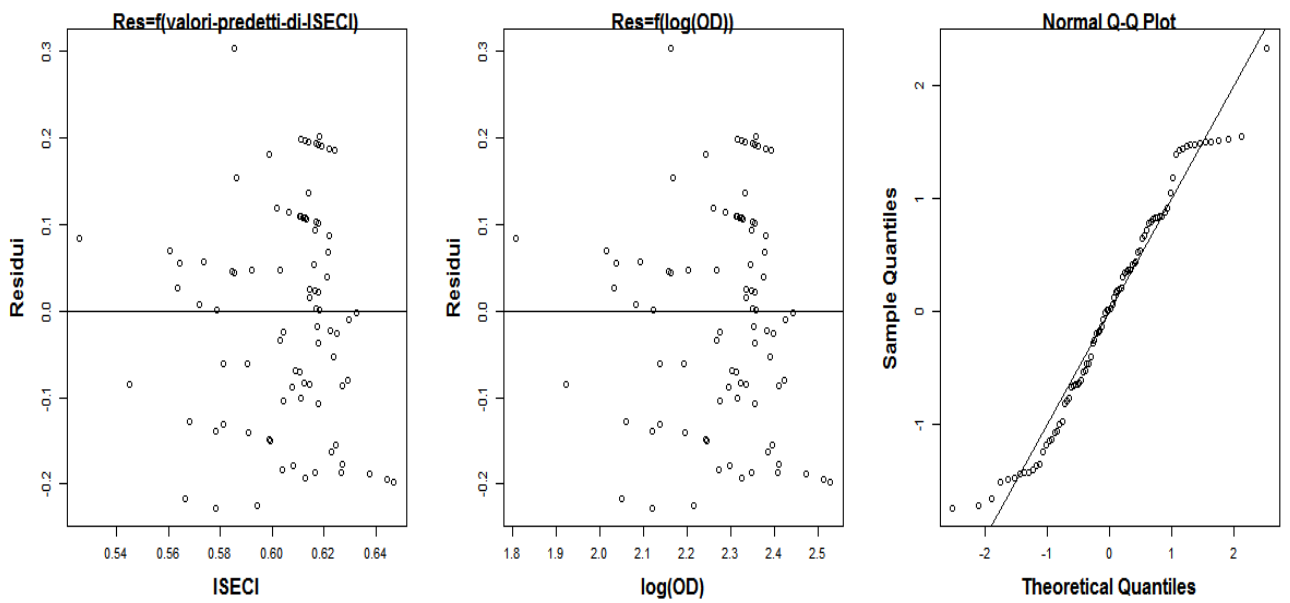


Figura 4.70. Grafici residui e Normal Q-Q Plot per regressione-lineare-logaritmica tra ISECI e OD.

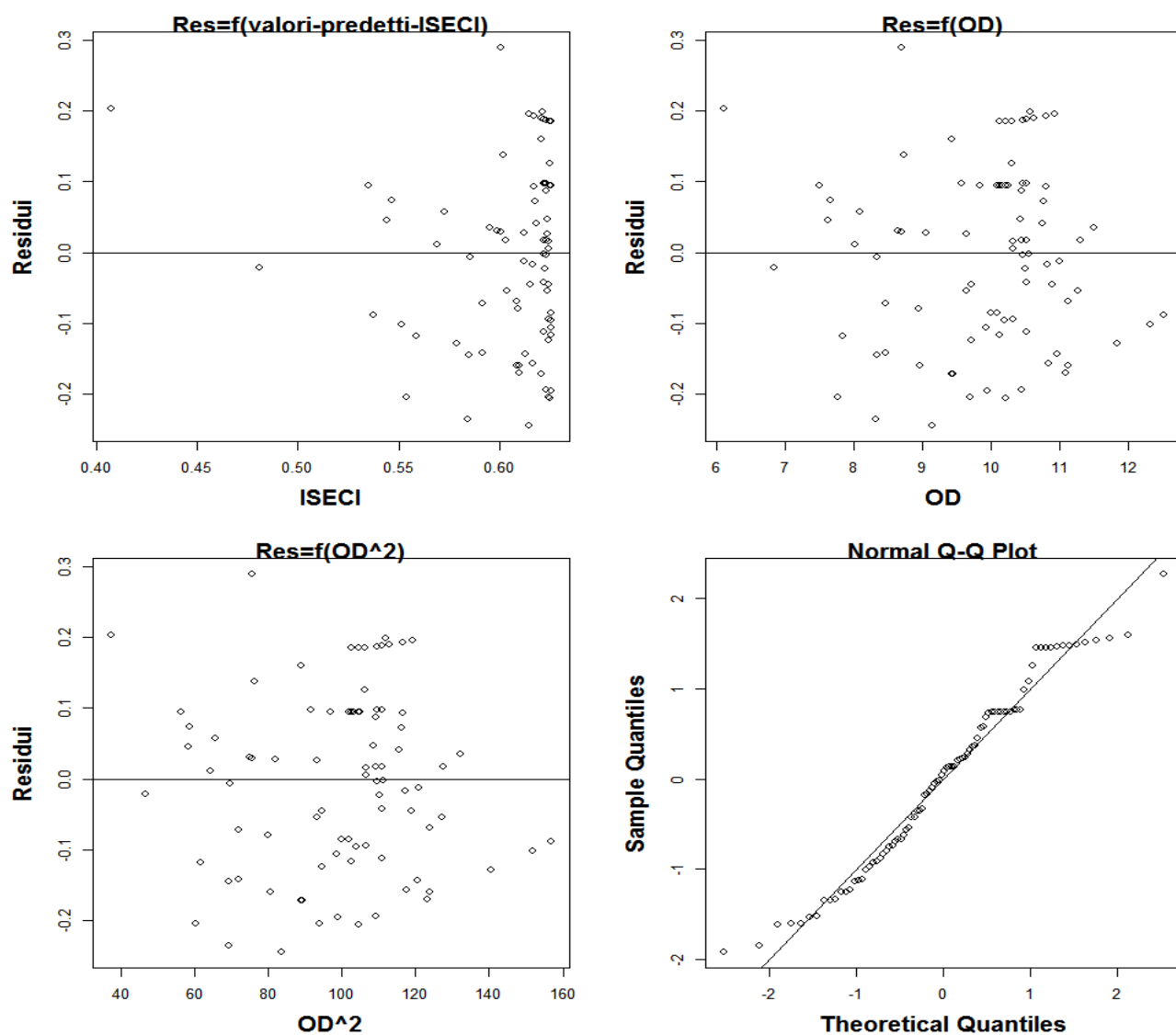


Figura 4.71. Grafici residui e Normal Q-Q Plot per regressione polinomiale tra ISEC1 e OD.

	lineare	logaritmica	polinomiale
R2 corretto	0,08138	0,179	0,1221
Statistica F	9	19,97	7,05
p_value_F	0,00408	2,380E-05	0,00147
p_value_Shapiro	0,05514	0,4257	0,2311
p_value_Breush_Pagan	0,8173	0,6492	0,8398

Tabella 4.26. Confronto dei risultati relativi alla regressione lineare, polinomiale e lineare-logaritmica con il fine di esaminare la relazione tra ISEC1 e PT. In giallo, la trasformazione che inizialmente è stata considerata approssimare bene i dati.

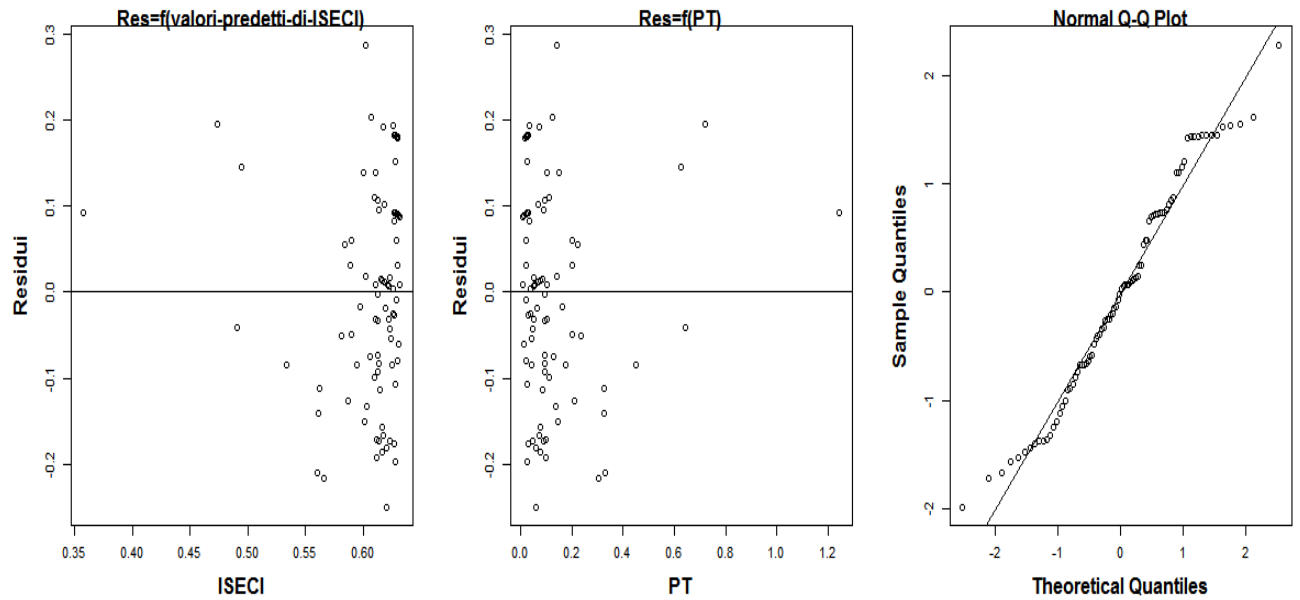


Figura 4.72. Grafici residui e Normal Q-Q Plot per regressione-lineare tra ISECI e PT.

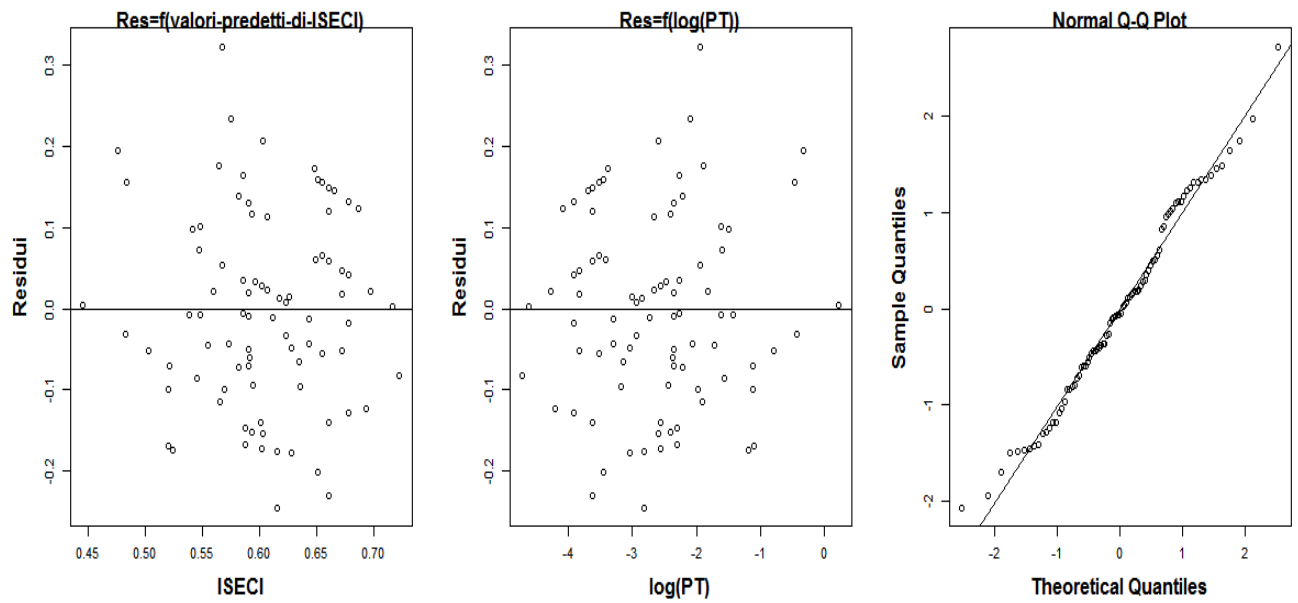


Figura 4.73. Grafici residui e Normal Q-Q Plot per regressione-lineare-logaritmica tra ISECI e PT.

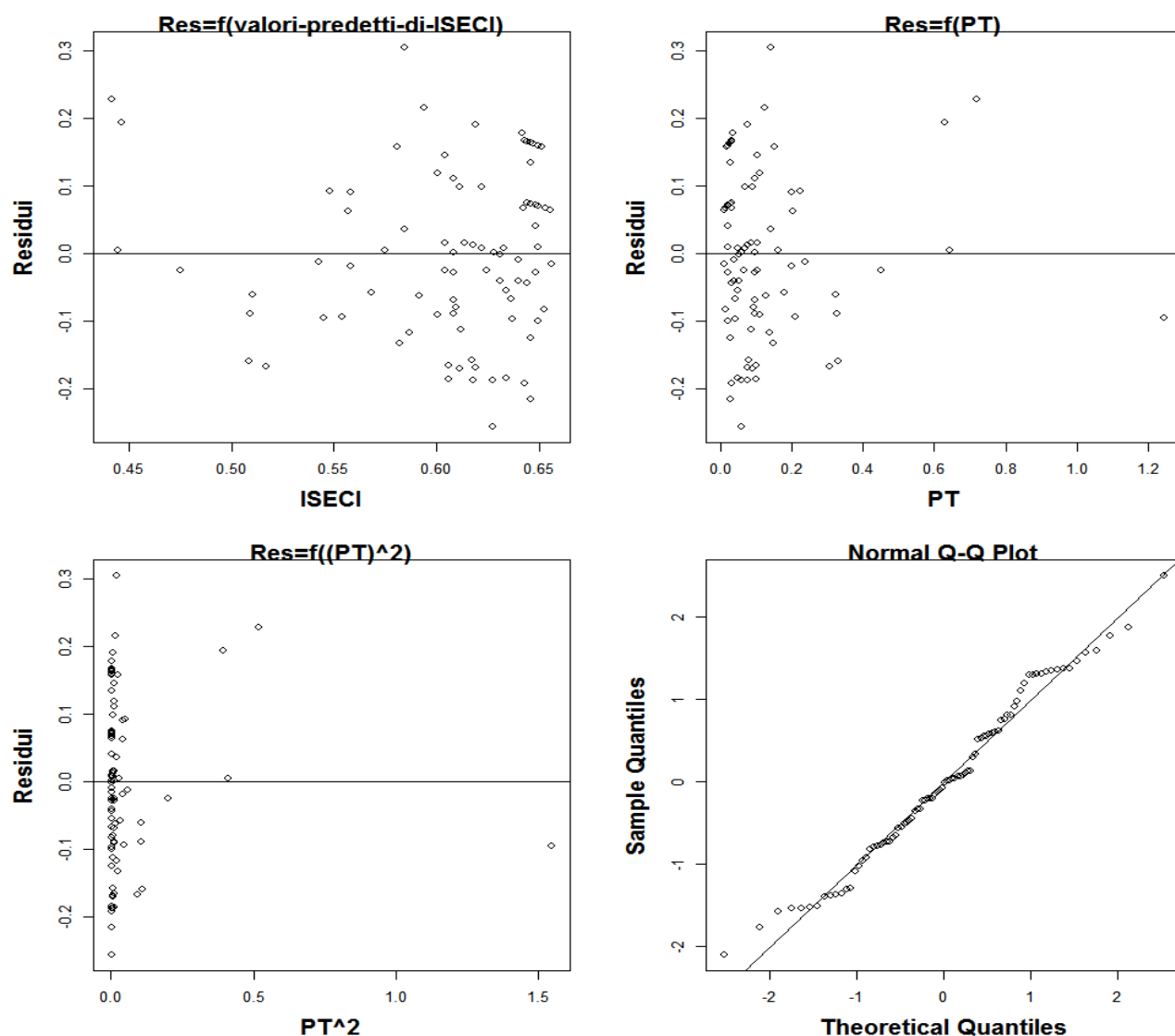


Figura 4.74. Grafici residui e Normal Q-Q Plot per regressione polinomiale tra ISECI e PT.

4.2.2. Regressione lineare multipla

Il secondo approccio, come già descritto nei metodi, era basato sul capire come ciascun RQE fosse simultaneamente relazionato alle variabili scelte come predittori usando la regressione multipla. I risultati sono riportati per ciascun RQE ed articolati nel seguente modo: analisi approfondita del primo modello di regressione multipla caratterizzato da forme funzionali lineari per ogni regressore e con inserimento intercetta, confronto con altri modelli e applicazione della procedura backward sul modello selezionato. Questa analisi approfondita racchiude i test di significatività, l'interpretazione dei grafici dei residui e la verifica

dell'assunzione di normalità sia graficamente che attraverso formali test statistici. Il confronto tra modelli invece è ottenuto variando le forme funzionali suggerite sia dai risultati relativi all'applicazione della regressione lineare semplice sia dai grafici dei residui della regressione lineare multipla. Esso è stato affiancato con un secondo approccio basato sui “partial residual plots” (detti anche “component+residual plots”). Quando il risultato del test non è statisticamente significativo è indicato in rosso.

4.2.2.1. RQE-IBMR

Il primo modello costruito è stato ottenuto applicando la regressione lineare multipla. I risultati sono sintetizzati nella tabella 4.27.

4.2.2.1.1. Test di significatività

Il risultato del test-F mostra che il modello spiega una quota significativa della varianza del RQE_IBMR: il p-value ad esso associato è minore di 0,05 (tabella 4.27). Ma nutrienti come N-NH₄, N-NO₃, PT, OD non contribuiscono in maniera significativa al modello. Poiché l'IBMR è definito come un metodo basato sulle macrofite per valutare lo stato trofico dei fiumi (Haury et al. 2006), si è proceduto a fare ulteriori trasformazioni per individuare un modello più rispondente alla realtà.

4.2.1.1.2. Grafici dei residui

I grafici dei residui contro i valori predetti del RQE-IBMR e contro il RQE-ICMi mostrano un pattern casuale (Figura 4.75); occorre evidenziare che il grafico dei residui contro il RQE-ICMi sembrerebbe anche essere caratterizzato dalla presenza di due clusters. Il grafico dei residui contro l'ossigeno mostra un pattern curvilineo, suggerendo un effetto quadratico che non è stato incluso nel modello.

I grafici dei residui relativi ai nutrienti mostrano sempre uno sbilanciamento rispetto all'asse delle X. Questo andamento dei residui, potrebbe dipendere dalle grandi oscillazioni del valore del RQE-IBMR in corrispondenza di bassi valori delle concentrazioni di N-NH₄, N-NO₃ e PT; questa forte incertezza nei dati è osservabile negli scatter-plot sviluppati durante l'analisi preliminare. Si evidenzia in tutti i grafici dei residui e nel Q-Q plot un'osservazione che non è congruente con le restanti e che risulta essere sempre la stazione numero 66 i cui valori sono riportati di seguito

Numero	Stazione	RQE IBMR	RQE ICMi	NH4(mg/L)	NO3(mg/L)	PTmg/L P)	OD(mg/L)
66	R1307AT15	1,42	0,8	0,065	1,325	0,225	9,05

Esso è stato considerato un outlier dal momento che presenta un valore inusuale (in questo caso molto elevato) della variabile dipendente, condizionatamente ai valori delle altre variabili indipendenti e presenta un residuo molto elevato. E' stato considerato un errore di misura e rimosso dai dati.

4.2.1.1.3. Verifica della normalità della distribuzione degli errori

Come già detto, il normal Q-Q plot è un grafico che permette di confrontare la funzione di distribuzione cumulativa empirica (ECDF) con la funzione di distribuzione teorica normale. Nella figura 4.76., affiancato dagli altri 2 grafici esso permette di avere una visione approfondita della normalità della distribuzione dei residui. Esso suggerirebbe una lieve dissimmetria a destra con una coda leggermente pesante. In effetti, il grafico relativo al confronto tra la curva di densità normale ed empirica mostra una distribuzione più allungata con una coda destra pesante rispetto ad una distribuzione normale, confermando quanto già dedotto dal normal Q-Q plot. Inoltre, quest'ultimo mostra sempre la presenza di un'osservazione anomala già identificata come un outlier: la stazione numero 66 (Figura 4.76.).

4.2.1.1.4. Confronto tra modelli

Come precedentemente indicato, è stato costruito un modello di partenza formato dalle seguenti forme funzionali: logaritmica per tutte le variabili indipendenti eccetto per l'OD rappresentato da una polinomiale. Sono stati costruiti più modelli variando di volta in volta la forma funzionale di una variabile indipendente selezionata. La tabella relativa al confronto di più modelli mostra che tutti i modelli hanno un livello di significatività molto prossimo (Figura 4.77.). Nella maggior parte dei modelli, i nutrienti costituiti da N-NH₄, N-NO₃ e PT non presentano significatività statistica. Invece, l'attribuzione della polinomiale all'ossigeno disciolto ne determina la sua significatività e determina la mancanza di significatività dell'intercetta. L'eliminazione della stazione 66 comporta un test di Shapiro-Wilk soddisfatto in tutti i modelli, sebbene in alcuni al limite. Nessun modello ha problemi di eteroschedasticità. L'aggiunta della variabile dicotomica relativa alle idroecoregioni non

implica grandi variazioni, poiché risulta essere sempre non significativa; l'unica osservazione che è possibile effettuare è che l'aggiunta della variabile dicotomica determina sempre una leggera riduzione della statistica F e quindi del livello di significatività totale di ciascun modello e un aumento del p-value associato al test di Shapiro ed una riduzione del p-value associato al test di Breush-Pagan.

Poiché nella maggior parte dei modelli variabili come N-NH₄, N-NO₃ e PT non risultano essere significative, si è individuata una nuova variabile definita NT (azoto totale) i cui valori sono ottenuti dalla somma di ciascun valore di N-NH₄ e N-NO₃; nuovamente è stata costruita una tabella basata sul confronto di più modelli ciascuno dei quali si differenzia dall'altro per una diversa forma funzionale attribuita a ciascuna variabile indipendente (Figura 4.78.). Si raggiungono analoghe considerazioni dedotte precedentemente: come per esempio quella relativa all'OD (l'attribuzione della polinomiale all'OD ne determina la sua significatività e determina la mancanza di significatività dell'intercetta); ma se ne aggiungono di nuove: come quella relativa alla variabile NT: l'attribuzione di una polinomiale o di una logaritmica alla variabile NT ne determina la sua significatività; il PT, invece, continua ad essere non significativo.

Le seguenti forme funzionali sono ritenute essere le più appropriate per le variabili considerate: logaritmica per il RQE-ICMi, polinomiale per NT ed OD, lineare per PT (viene considerata una relazione lineare non essendo significativa per nessuna forma funzionale considerata). Tali considerazioni sono anche supportate dai partial-residuals-plot ottenuti dal modello espresso con formule funzionali lineari per ogni regressore (Figura 4.79.).

La formulazione finale fu raggiunta attraverso la procedura di eliminazione backward, come mostrato in tabella 4.28. E' interessante osservare come l'annullamento dell'intercetta incida notevolmente sull'aumento di significatività dell'OD.

	Coefficienti di regressione	p-value
b0_ Intercetta	0,4638	5,35E-04
b1_ RQE_ ICMi	0,209	1,20E-03
b2_ N-NH4	0,0057	9,11E-01
b3_ N-NO3	-0,0074	5,53E-01
b4_ PT	-0,0847	3,84E-01
b5_ OD	0,0214	9,56E-02
	Valori delle principali statistiche	p-value
Adjusted R Squared	0,2574	
F-statistic	6,8810	1,43E-05
test-shapiro	0,9503	5,03E-04
test-Breush-pagan	2,1595	8,27E-01

Tabella 4.27. Risultati dei tests di significatività, di Shapiro-Wilk e di Breush Pagan inerenti al modello di regressione caratterizzato da forme lineari per ogni variabile indipendente.

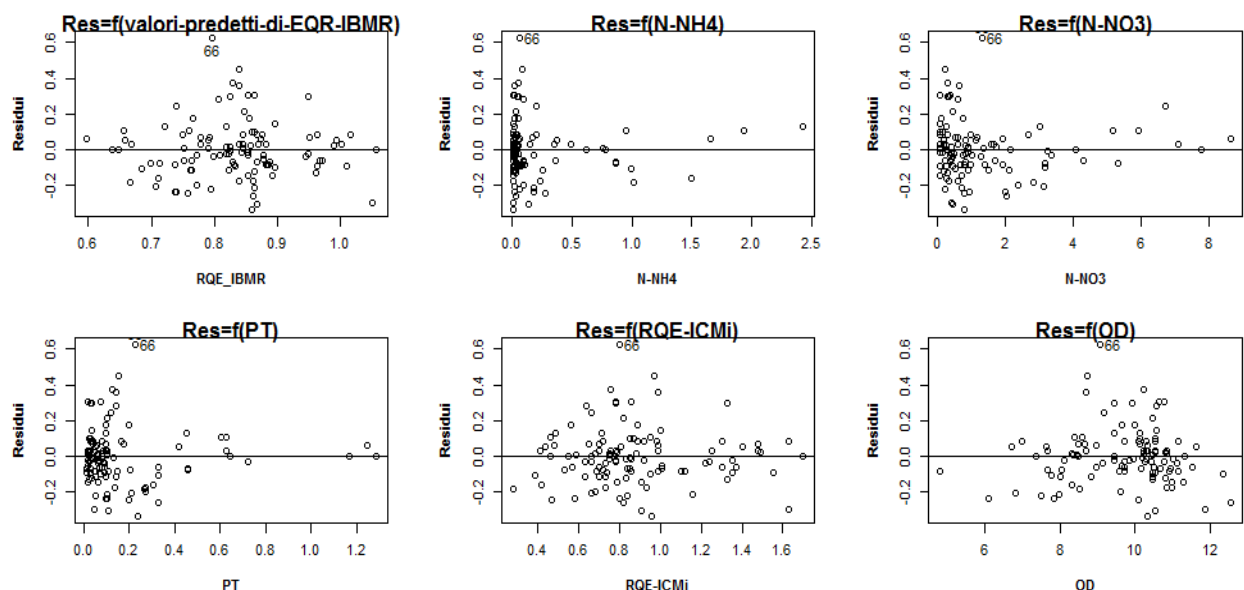


Figura 4.75. Grafici dei residui contro i valori predetti del RQE-IBMR e ciascuna variabile esplicativa inerenti al modello di regressione caratterizzato da forme lineari per ogni variabile indipendente.

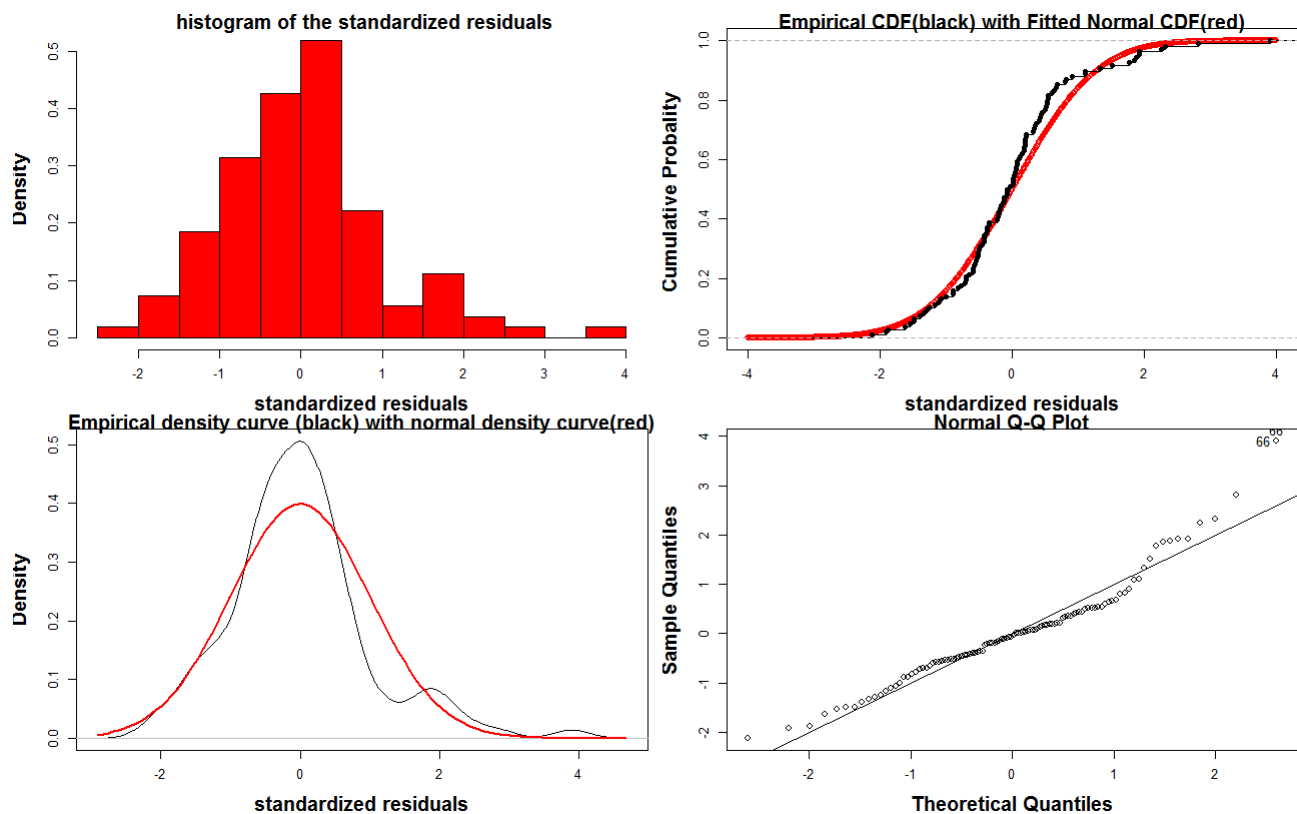


Figura 4.76. Verifica dell'assunzione di normalità della distribuzione degli errori attraverso l'istogramma dei residui standardizzati, confronto grafico tra la funzione di distribuzione cumulativa empirica (ecdf) e normale, tra la funzione di densità di probabilità empirica e normale e normal Q-Q plot dei residui standardizzati.

Forma funzionale	Lineare_ICMI		Log(ICMI)		Polinomiale_ICMI	
Presenza/assenza	senza variabile dummy	con variabile dummy	senza variabile dummy	con variabile dummy	senza variabile dummy	con variabile dummy
Sommario statistico						
b0_Intercept(p-value)	-0,9116(0,0517)	-0,8794(0,0597)	-0,6612(0,1493)	-0,615(0,1788)	-1,0481(0,0294)	-1,0174(0,0338)
b1_(RQE_ICMI)	0,2165(0,0003)	0,2288(0,0001)	0,2137(8,9E-05)	0,224(4,71E-05)	0,5670(0,0492)	0,5841(0,0422)
b1_(RQE_ICMI)^2_Polinomiale					-0,1690(0,2119)	-0,1712(0,2042)
b2_In(NH4-N)	-0,0017(0,8925)	-0,0004(0,9771)	0,0037(0,7673)	0,005(0,6767)	0,0037(0,7745)	0,0051(0,6935)
b3_In(NO3-N)	-0,0299(0,0756)	-0,0168(0,3840)	-0,0303(0,0694)	-0,017(0,3666)	-0,0296(0,0778)	-0,0163(0,3957)
b4_In(TP)	-0,0131(0,4889)	-0,0111(0,556)	-0,0105(0,5746)	-0,009(0,6457)	-0,0116(0,5389)	-0,0096(0,6105)
b5_DO	0,3186(0,0021)	0,3013(0,0036)	0,3193(0,0018)	0,302(0,0032)	0,3171(0,0021)	0,2995(0,0037)
b6_DO^2	-0,0166(0,0035)	-0,0155(0,0066)	-0,0166(0,0031)	-0,016(0,0060)	-0,0165(0,0036)	-0,0154(0,0069)
b7_Dummy		0,0510(0,1808)		0,051(0,1783)		0,0515(0,1747)
Adjusted R Squared	0,342640454	0,347962145	0,35585589	0,361204667	0,346421038	0,352109165
F-statistic	10,20852939	9,081035491	10,75990822	9,562477525	9,026274691	8,200976122
p-value-F	8,59792E-09	1,27221E-08	3,31152E-09	4,96097E-09	1,41746E-08	1,9362E-08
p-value-shapiro	0,03816	0,06425	0,065800	0,1135	0,069240	0,1418
p-value-Breush-pagan-test	0,2664	0,1809	0,231300	0,1226	0,1587	0,06342
Forma funzionale	Lineare_NH4-N		Log_NH4-N		Polinomiale_NH4-N	
Presenza/assenza	senza variabile dummy	con variabile dummy	senza variabile dummy	con variabile dummy	senza variabile dummy	con variabile dummy
Sommario statistico						
b0_Intercept(p-value)	-0,6724(0,1334)	-0,6247(0,1618)	-0,6612(0,1493)	-0,615(0,1788)	-0,6825(0,12578)	(-0,6375)(0,1520)
b1_In(RQE_ICMI)	0,2162(0,0001)	0,2297(0,00033)	0,2137(8,9E-05)	0,224(4,71E-05)	0,2060(0,0016)	0,2193(0,0001)
b2_(NH4-N)	0,0208(0,6316)	0,0332(0,4514)	0,0037(0,7673)	0,005(0,6767)	-0,1321(0,24904)	-0,1098(0,3406)
b2_(NH4-N)^2_polinomiale					0,0782(0,1502)	0,0727(0,1803)
b3_In(NO3-N)	-0,0295(0,0674)	-0,0151(0,4206)	-0,0303(0,0694)	-0,017(0,3666)	-0,0286(0,07498)	-0,0153(0,4139)
b4_In(TP)	-0,0121(0,5283)	-0,0111(0,5593)	-0,0105(0,5746)	-0,009(0,6457)	-0,0058(0,76480)	-0,0054(0,7813)
b5_DO	0,3180(0,0017)	0,2976(0,0034)	0,3193(0,0018)	0,302(0,0032)	0,3284(0,00120)	0,3087(0,0024)
b6_DO^2	-0,0166(0,0029)	-0,0153(0,0062)	-0,0166(0,0031)	-0,016(0,0060)	-0,0172(0,00193)	-0,0160(0,0043)
b7_Dummy		0,0548(0,1509)		0,051(0,1783)		0,0509(0,1811)
Adjusted R Squared	0,356776308	0,363748951	0,35585589	0,361204667	0,363791864	0,368981592
F-statistic	10,79915416	9,657272013	10,75990822	9,562477525	9,658877371	8,747802664
p-value-F	3,096E-09	4,12889E-09	3,31152E-09	4,96097E-09	4,1161E-09	5,92447E-09
p-value-shapiro	0,0422	0,07099	0,065800	0,1135	0,036320	0,05083
p-value-Breush-pagan-test	0,1847	0,1835	0,231300	0,1226	0,1832	0,2498
Forma funzionale	Lineare_NO3-N		Log_NO3-N		Polinomiale_NO3-N	
Presenza/assenza	senza variabile dummy	con variabile dummy	senza variabile dummy	con variabile dummy	senza variabile dummy	con variabile dummy
Sommario statistico						
b0_Intercept(p-value)	-0,7772(0,0924)	-0,6520(0,1540)	-0,6612(0,1493)	-0,615(0,1788)	-0,5960(0,2014)	-0,5448(0,2397)
b1_In(RQE_ICMI)	0,2075(0,0002)	0,2296(3,97E-05)	0,2137(8,9E-05)	0,224(4,71E-05)	0,2107(0,000121)	0,2278(4,36E-05)
b2_In(NH4-N)	-0,0013(0,9181)	0,0027(0,8280)	0,0037(0,7673)	0,005(0,6767)	0,0014(0,9110)	0,0039(0,7547)
b3_(NO3-N)	-0,0047(0,6574)	0,0055(0,6316)	-0,0303(0,0694)	-0,017(0,3666)	-0,0506(0,0686)	-0,0293(0,3333)
b3_(NO3-N)^2_polinomiale					0,0064(0,739)	0,0046(0,2140)
b4_In(TP)	-0,0228(0,2227)	-0,0175(0,3448)	-0,0105(0,5746)	-0,009(0,6457)	-0,0172(0,3602)	-0,0144(0,4392)
b5_DO	0,3361(0,0012)	0,2971(0,0040)	0,3193(0,0018)	0,302(0,0032)	0,3099(0,0027)	0,2855(0,0057)
b6_DO^2	-0,0174(0,0023)	-0,0151(0,0079)	-0,0166(0,0031)	-0,016(0,0060)	-0,0161(0,0045)	-0,0146(0,0102)
b7_Dummy		0,0750(0,0382)		0,051(0,1783)		0,0611(0,1045)
Adjusted R Squared	0,335462389	0,357396692	0,35585589	0,361204667	0,35016323	0,361037246
F-statistic	9,918234425	9,422003112	10,75990822	9,562477525	9,159697962	8,486732965
p-value-F	1,43024E-08	6,5193E-09	3,31152E-09	4,96097E-09	1,08961E-08	1,0396E-08
p-value-shapiro	0,01446	0,07312	0,065800	0,1135	0,067870	0,165
p-value-Breush-pagan-test	0,2264	0,06042	0,231300	0,1226	0,07506	0,06019
Forma funzionale	Linear_TP		Log_TP		Polynomial_TP	
Presenza/assenza	senza variabile dummy	con variabile dummy	senza variabile dummy	con variabile dummy	senza variabile dummy	con variabile dummy
Sommario statistico						
b0_Intercept(p-value)	-0,6389(0,1635)	-0,6003(0,1894)	-0,6612(0,1493)	-0,615(0,1788)	-0,6243(0,1757)	-0,5865(0,2024)
b1_In(RQE_ICMI)	0,2186(3,55E-05)	0,2293(1,8E-05)	0,2137(8,9E-05)	0,224(4,71E-05)	0,2127(8,90E-05)	0,2235(4,75E-05)
b2_In(NH4-N)	0,0039(0,7572)	0,0049(0,7000)	0,0037(0,7673)	0,005(0,6767)	0,0047(0,7130)	0,0056(0,6594)
b3_In(NO3-N)	-0,0319(0,0418)	-0,0195(0,2842)	-0,0303(0,0694)	-0,017(0,3666)	-0,0298(0,0683)	-0,0174(0,3518)
b4_(TP)	-0,0385(0,6113)	-0,0210(0,7844)	-0,0105(0,5746)	-0,009(0,6457)	-0,1368(0,5177)	-0,1160(0,5831)
b4_(TP)^2_polinomiale					0,083080,6181	0,0801(0,6292)
b5_DO	0,3214(0,0017)	0,3037(0,0031)	0,3193(0,0018)	0,302(0,0032)	0,3202(0,0018)	0,3027(0,0033)
b6_DO^2	-0,0167(0,0029)	-0,0156(0,0057)	-0,0166(0,0031)	-0,016(0,0060)	-0,0166(0,0032)	-0,0155(0,0062)
b7_Dummy		0,0501(0,1883)		0,051(0,1783)		0,0499(0,1922)
Adjusted R Squared	0,355488531	0,360318991	0,35585589	0,361204667	0,350618581	0,355335323
F-statistic	10,74427561	9,52965607	10,75990822	9,562477525	9,176037879	8,303320936
p-value-F	3,40156E-09	5,2873E-09	3,31152E-09	4,96097E-09	1,05515E-08	1,5484E-08
p-value-shapiro	0,085	0,1347	0,065800	0,1135	0,063820	0,1272
p-value-Breush-pagan-test	0,5021	0,2879	0,231300	0,1226	0,2749	0,1625
Forma funzionale	Lineare_DO		Log_DO		Polinomiale_DO	
Presenza/assenza	senza variabile dummy	con variabile dummy	senza variabile dummy	con variabile dummy	senza variabile dummy	con variabile dummy
Sommario statistico						
b0_Intercept(p-value)	0,7046(5,77E-08)	0,6362(1,83E-06)	0,4380(0,061)	0,3415(0,15)	-0,6389(0,1635)	-0,6003(0,1894)
b1_In(RQE_ICMI)	0,2038(0,000175)	0,2193(6,43E-05)	0,2018(0,000177)	0,2180(6,24E-05)	0,2186(3,55E-05)	0,2293(1,8E-05)
b2_In(NH4-N)	0,0115(0,3749)	0,0121(0,3464)	0,0114(0,3738)	-0,0184(0,3246)	0,0039(0,7572)	0,0049(0,7000)
b3_In(NO3-N)	-0,0363(0,0259)	-0,0194(0,3033)	-0,0354(0,0290)	0,0119(0,350)	-0,0319(0,0418)	-0,0195(0,2842)
b4_(TP)	-0,0393(0,6180)	-0,0160(0,8403)	-0,0404(0,6057)	-0,0169(0,8301)	-0,0385(0,6113)	-0,0210(0,7844)
b5_DO	0,0195(0,1013)	0,0229(0,0559)	0,2013(0,0502)	0,2281(0,027)	0,3214(0,0017)	0,3037(0,0031)
b6_DO^2_polinomiale					-0,0167(0,0029)	-0,0156(0,0057)
b7_Dummy		0,0665(0,0886)		0,0670(0,083)		0,0501(0,1883)
Adjusted R Squared	0,302442578	0,315697091	0,310371179	0,324080494	0,355488531	0,360318991
F-statistic	10,1917632	9,150360319	10,5411746	9,470567887	10,74427561	9,52965607
p-value-F	6,32909E-08	5,61965E-08	3,66716E-08	3,16346E-08	3,40156E-09	5,2873E-09
p-value-shapiro	0,0415	0,1428	0,043570	0,1405	0,085	0,1347
p-value-Breush-pagan-test	0,5406	0,2821	0,520200	0,2488	0,5021	0,2879

Figura 4.77. Confronto tra modelli ottenuti variando le forme funzionali di un modello base caratterizzato da logaritmica per tutte le variabili indipendenti eccetto per l'OD rappresentato da una polinomiale. Il modello selezionato in giallo rappresenta quello che, di volta in volta, è stato ritenuto il migliore fra quelli confrontati.

Forma funzionale	Lineare_ICMI	Log(ICMI)_Mod.Iniziale	Polinomiale_ICMI
Sommario statistico			
b0_Intercetta(p-value)	-0,7429(0,1058)	-0,5481(0,2195)	-0,914(0,0570)
b1_(RQE_ICMI)(p-value)	0,2122(0,0002)	0,2047(6,97E-05)	0,5420(0,089)
b1_(RQE_ICMI^2)(p-value)			-0,1603(0,2186)
b2_(NT)(p-value)	-0,0569(0,0136)	-0,0533(0,0202)	-0,0538(0,0198)
b3_(NT^2)(p-value)	0,0066(0,0156)	0,0064(0,0179)	0,0065(0,0169)
b4_(PT)(p-value)	-0,1111(0,1753)	-0,0971(0,2336)	-0,0979(0,2350)
b5_OD(p-value)	0,3067(0,0025)	0,3128(0,0019)	0,3103(0,0022)
b6_OD^2(p-value)	-0,0159(0,0041)	-0,0163(0,0031)	-0,0162(0,0035)
R ² aggiustato	0,353836533	0,365637759	0,357263272
Test-F	10,67419609	11,18282614	9,41711149
p-value-F	3,83688E-09	1,61043E-09	6,58177E-09
p-value-test-Shapiro	0,1098	0,120500	0,135600
p-value-test-Breush-pagar	0,4991	0,468400	0,4206
Forma funzionale	Lineare_NT	Log_NT	Polinomiale_NT_Mod.Iniziale
Sommario statistico			
b0_Intercetta(p-value)	-0,7333(0,1041)	-0,6761(0,1245)	-0,5481(0,2195)
b1_Ln(RQE_ICMI)(p-value)	0,2232(2E-05)	0,2040(0,0001)	0,2047(6,97E-05)
b2_(NT)(p-value)	-0,0047(0,6508)	-0,0356(0,0235)	-0,0533(0,0202)
b3_(NT^2)(p-value)			0,0064(0,0179)
b4_(PT)(p-value)	-0,0758(0,3599)	-0,0189(0,7956)	-0,0971(0,2336)
b5_OD(p-value)	0,3417(0,0009)	0,3283(0,0010)	0,3128(0,0019)
b6_OD^2(p-value)	-0,0177(0,0017)	-0,0172(0,0017)	-0,0163(0,0031)
R ² aggiustato	0,335535416	0,367322661	0,365637759
Test-F	11,70538746	13,30839152	11,18282614
p-value-F	6,17653E-09	5,81718E-10	1,61043E-09
p-value-test-Shapiro	0,04368	0,06386	0,120500
p-value-test-Breush-pagar	0,5415	0,827800	0,468400
Forma funzionale	Lineare_PT_Mod_iniziale	Log_PT	Polinomiale_PT
Sommario statistico			
b0_Intercetta(p-value)	-0,5481(0,2195)	-0,6035(0,1798)	-0,5481(0,2217)
b1_Ln(RQE_ICMI)(p-value)	0,2047(6,97E-05)	0,2019(0,0001)	0,2028(0,0001)
b2_(NT)(p-value)	-0,0533(0,0202)	-0,0493(0,0477)	-0,0514(0,0412)
b3_(NT^2)(p-value)	0,0064(0,0179)	0,0055(0,0458)	0,0062(0,0312)
b4_(PT)(p-value)	-0,0971(0,2336)	-0,0153(0,4092)	-0,1337(0,5344)
b4_(PT^2)(p-value)			0,0314(0,8543)
b5_OD(p-value)	0,3128(0,0019)	0,3130(0,0019)	0,3130(0,0020)
b6_OD^2(p-value)	-0,0163(0,0031)	-0,0163(0,0032)	-0,0163(0,0032)
R ² aggiustato	0,365637759	0,36091743	0,359449456
Test-F	11,18282614	10,97712696	9,497521082
p-value-F	1,61043E-09	2,28408E-09	5,62798E-09
p-value-test-Shapiro	0,120500	0,07299	0,1086
p-value-test-Breush-pagar	0,468400	0,056730	0,176800
Forma funzionale	Lineare_OD	Log_OD	Polinomiale_OD_Mod_iniziale
Sommario statistico			
b0_Intercetta(p-value)	0,7511(1,16E-08)	0,5011(0,0343)	-0,5481(0,2195)
b1_Ln(RQE_ICMI)	0,1819(0,00052)	0,1798(0,0005)	0,2047(6,97E-05)
b2_(NT)(p-value)	-0,0575(0,0159)	-0,0561(0,0179)	-0,0533(0,0202)
b3_(NT^2)(p-value)	0,0072(0,0098)	0,0071(0,0113)	0,0064(0,0179)
b4_(PT)(p-value)	-0,1023(0,2278)	-0,1010(0,2310)	-0,0971(0,2336)
b5_OD(p-value)	0,0175(0,1326)	0,1849(0,0674)	0,3128(0,0019)
b6_OD^2(p-value)			-0,0163(0,0031)
R ² aggiustato	0,314068499	0,321418556	0,365637759
Test-F	10,70687624	11,04164417	11,18282614
p-value-F	2,83609E-08	1,69345E-08	1,61043E-09
p-value-test-Shapiro	0,06456	0,057040	0,120500
p-value-test-Breush-pagar	0,5609	0,560800	0,468400

Figura 4.78. Confronto tra modelli ottenuti variando le forme funzionali di un modello base caratterizzato dalla logaritmica per RQE_ICMI, la polinomiale per NT (azoto totale), polinomiale per ossigeno disciolto (OD). Il modello selezionato in giallo rappresenta quello che, di volta in volta, è stato ritenuto il migliore fra quelli confrontati.

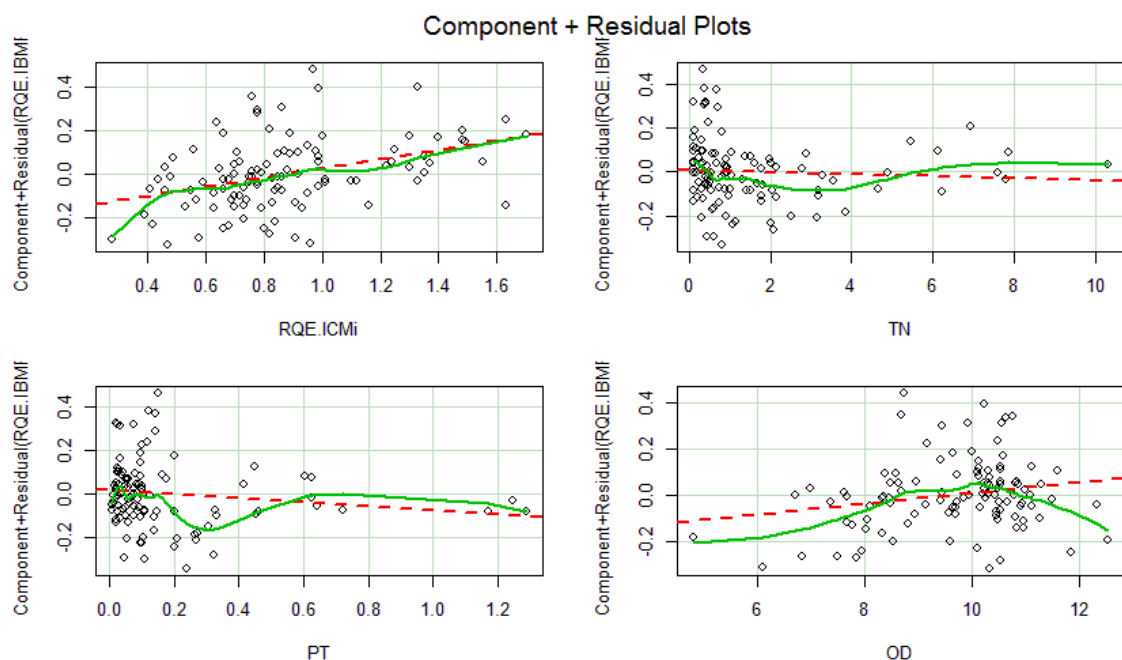


Figura 4.79. Partial residuals plot del modello caratterizzato da forme funzionali lineari per ogni regressore.

	STEP1	STEP2	STEP3
The removed variable		TP	Intercept
Statistic summary			
b0_ Intercept(p-value)	-0,548(0,2195)	-0,569(0,2027)	
b1_ ln(RQE_ICMi)(p-value)	0,205(6,97E-05)	0,214(2,85E-05)	0,209(4,19E-05)
b2_ (TN)(p-value)	-0,053(0,02020)	-0,058(0,0108)	-0,062(0,006)
b3_ (TN^2) (p-value)	0,0064(0,0179)	0,006(0,0244)	0,006(0,0129)
b5_ (TP)(p-value)	-0,097(0,2335)		
b6_ D0(p-value)	0,313(0,0018)	0,316(0,0017)	0,191(<2E-16)
b7_ DO^2(p-value)	-0,016(0,0030)	-0,016(0,0029)	-0,0098(3,95E-10)
Adjusted R Squared	0,3656	0,3629	0,9719
F-statistic	11,18	13,08	740
p-value-F	1,61E-09	8,15E-10	7,35E-78

Tabella 4.28. Applicazione del metodo backward al modello selezionato

4.2.2.2. RQE-ICMi

Anche in questo caso, il primo modello costruito è stato quello basato sull'analisi di regressione lineare multipla, con forme funzionali lineari per ciascun regressore.

4.2.2.2.1. Test di significatività

Il risultato del test-F mostra che il modello spiega una quota significativa della varianza del RQE_ICMi: il p-value ad esso associato è minore di 0,05 (tabella 4.29). I risultati dei test t per ciascuna delle variabili esplicative mostrano che solo RQE-IBMR e N-NH₄ risultano essere significative, sebbene N-NH₄ sia molto prossima alla soglia di significatività. Tali risultati determinano la necessità di individuare nuovi modelli più rappresentativi.

4.2.2.2.2. Grafici dei residui

L'analisi grafica dei residui (figura 4.80) suggerisce eteroschedasticità collegata soprattutto alle variabili RQE_IBMR, OD e N-NO₃: i punti si dispongono secondo un pattern a forma di imbuto. Anche, il test di Breush-Pagan conferma tale risultato, non essendo statisticamente significativo (tabella 4.29)

4.2.2.2.3. Verifica della normalità della distribuzione degli errori

Il Q-Q plot mostra una curva concava verso l'alto suggerendo una dissimmetria a destra. In effetti, l'istogramma non risulta essere simmetrico e la funzione di distribuzione cumulativa empirica (ecdf) e normale non si sovrappongono (figura 4.81).

4.2.2.2.4. Confronto tra modelli

Il modello di partenza scelto era basato sulle seguenti forme funzionali: logaritmica per tutte le variabili. Dal confronto di più modelli ottenuti attraverso la variazione di una forma funzionale si raggiungono le seguenti conclusioni (figura 4.83): a differenza del RQE-IBMR, in tutti i modelli relativi al RQE-ICMi, la variabile dicotomica è sempre significativa. Ciò contrasterebbe con la definizione stessa di RQE che dovrebbe già prendere in considerazione la specificità della stazione campionata in particolare la sua appartenenza ad una precisa eco-regione e ad un preciso livello tipologico. E' evidente, inoltre, che la sua introduzione incide soprattutto sulla significatività dell'N-NO₃. Infatti, poiché le più alte concentrazioni di N-NO₃ sono registrate nelle aree costiere e nei più bassi corsi dei fiumi, una forte correlazione tra le due variabili è ipotizzabile. In tutti i modelli relativi al RQE-ICMi è evidente la non

significatività del $N-NO_3$ e la significatività della variabile dummy. Analoghe considerazioni possono essere fatte per la correlazione tra $N-NO_3$ e PT che in forma logaritmica è significativo nella maggior parte dei modelli (figura 4.83.).

Nessun modello è caratterizzato da una distribuzione normale dei residui e molti non rispettano l'assunzione di omoschedasticità. L'introduzione della variabile dicotomica sembra ridurre l'eteroschedasticità, lasciando ipotizzare anche la possibilità di due campioni provenienti da due diverse popolazioni. Però la sua introduzione contrasterebbe, come detto prima, con la definizione di RQE.

In nessun modello l'OD è significativo (figura 4.83). Ma è molto probabile che l'annullamento dell'intercetta comporti la sua significatività

Questo è stato verificato per il modello selezionato in giallo caratterizzato dalla logaritmica per ogni regressore e per quello di regressione lineare multipla (Figura 4.84.). Il primo è molto interessante: i partial residual plots (Figura 4.82) evidenzerebbero un'assenza di correlazione lineare di tale indice sia con il PT che con $N-NH_4$ determinata dall'addensamento dei punti sul lato sinistro del diagramma e la trasformazione logaritmica potrebbe essere un rimedio; ma, anche questo modello continuerebbe a violare le assunzioni di base della regressione: normalità ed omoschedasticità.

Così, poiché si vuole porre l'attenzione soprattutto sulla violazione dell'assunzioni di base della regressione avente come variabile dipendente il RQE-ICMi, è stato preferito adottare il primo modello di regressione lineare multipla a cui è stata applicata la procedura di eliminazione backward. Anche in questo caso è possibile evidenziare come l'annullamento dell'intercetta comporti la significatività dell'OD (Figura 4.84.). La procedura backward è stata applicata comparando i test di significatività ottenuti attraverso gli errori standard OLS che quelli consistenti all'eteroschedasticità (Tabella 4.30). I risultati ottenuti dalla procedura backward non sono influenzati dalla stima corretta degli errori standard. Essi mostrano che le variabili che contribuiscono significativamente al modello sono: $N-NH_4$, OD e RQE_IBMR.

	Coefficienti di regressione	p-value
b0_ Intercetta	0,18539	0,362
b1_ RQE_IBMR	0,47017	0,0012
b2_ N-NH4	-0,15834	0,0369
b3_ N-NO3	-0,0217	0,244
b4_ PT	-0,04508	0,7576
b5_ OD	0,03734	0,0522
	Valori delle principali statistiche	p-value
R ² aggiustato	0,316	
test-F	10,88	2,08E-08
test-Shapiro-Wilk	0,94042	0,000111
test-Breush-Pagan	13,786000	0,01703

Tabella 4.29. Risultati dei tests di significatività, di Shapiro-Wilk e di Breush Pagan inerenti al modello di regressione caratterizzato da forme lineari per ogni variabile indipendente.

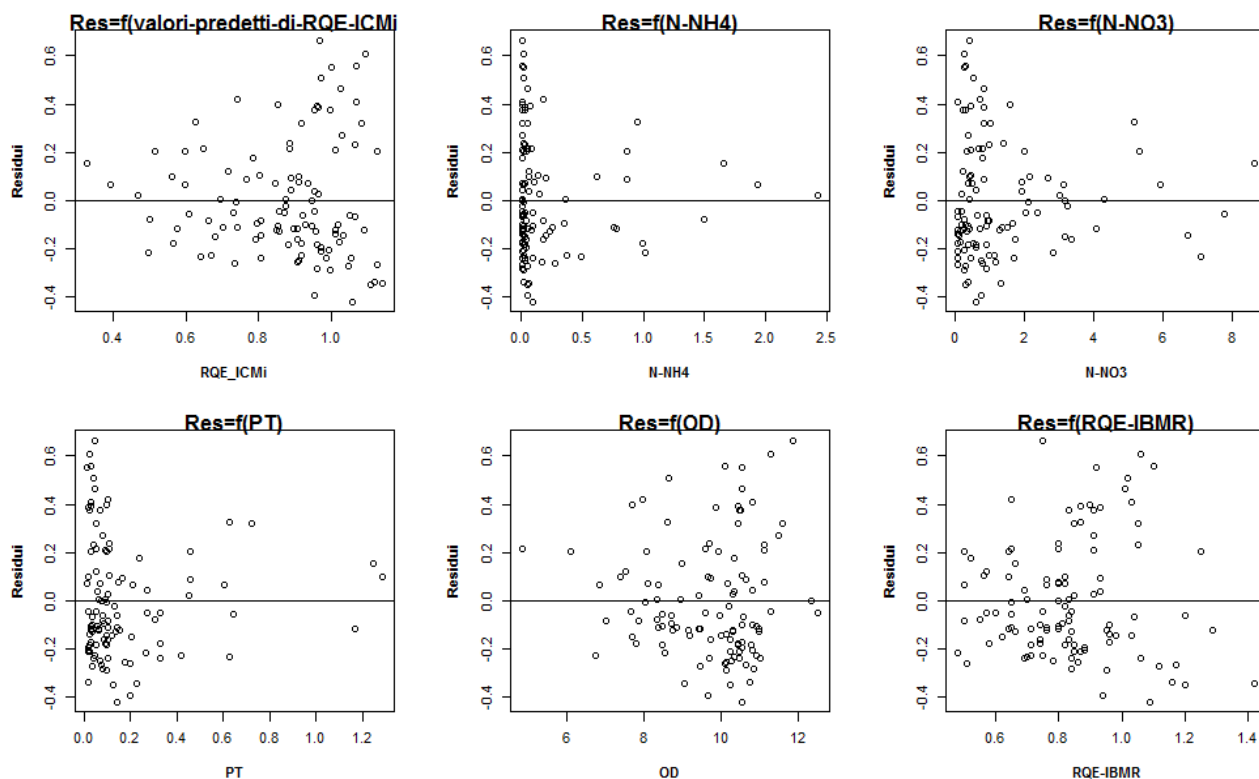


Figura 4.80. Grafici dei residui contro i valori predetti del RQE-ICMi e ciascuna variabile esplicativa inerenti al modello di regressione caratterizzato da forme lineari per ogni variabile indipendente.

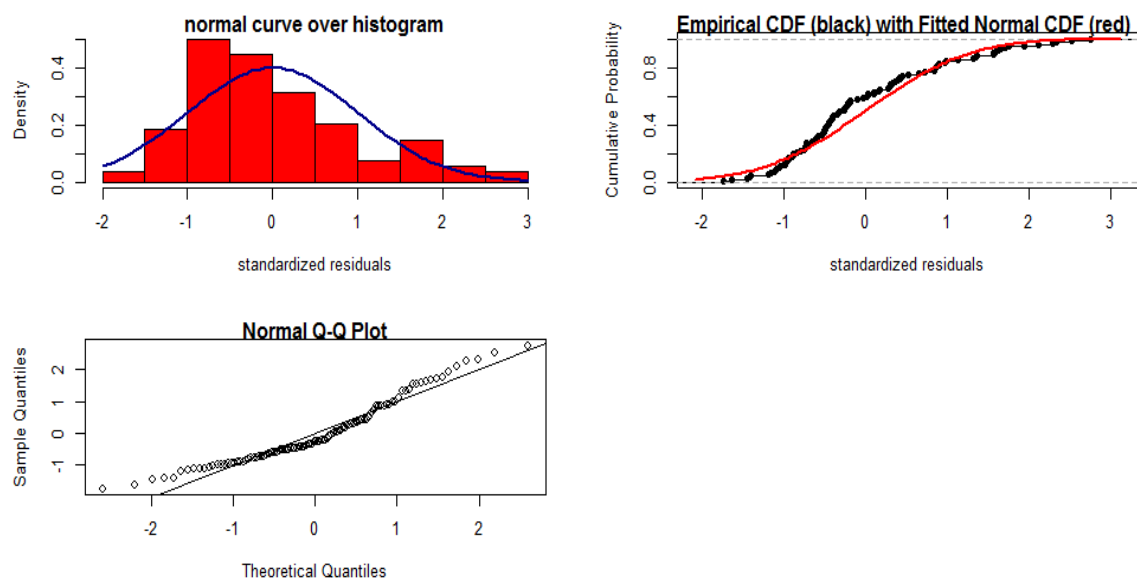


Figura 4.81. Verifica dell'assunzione di normalità della distribuzione degli errori attraverso l'istogramma dei residui standardizzati, confronto grafico tra la funzione di distribuzione cumulativa empirica (ecdf) e normale e normal Q-Q plot dei residui standardizzati.

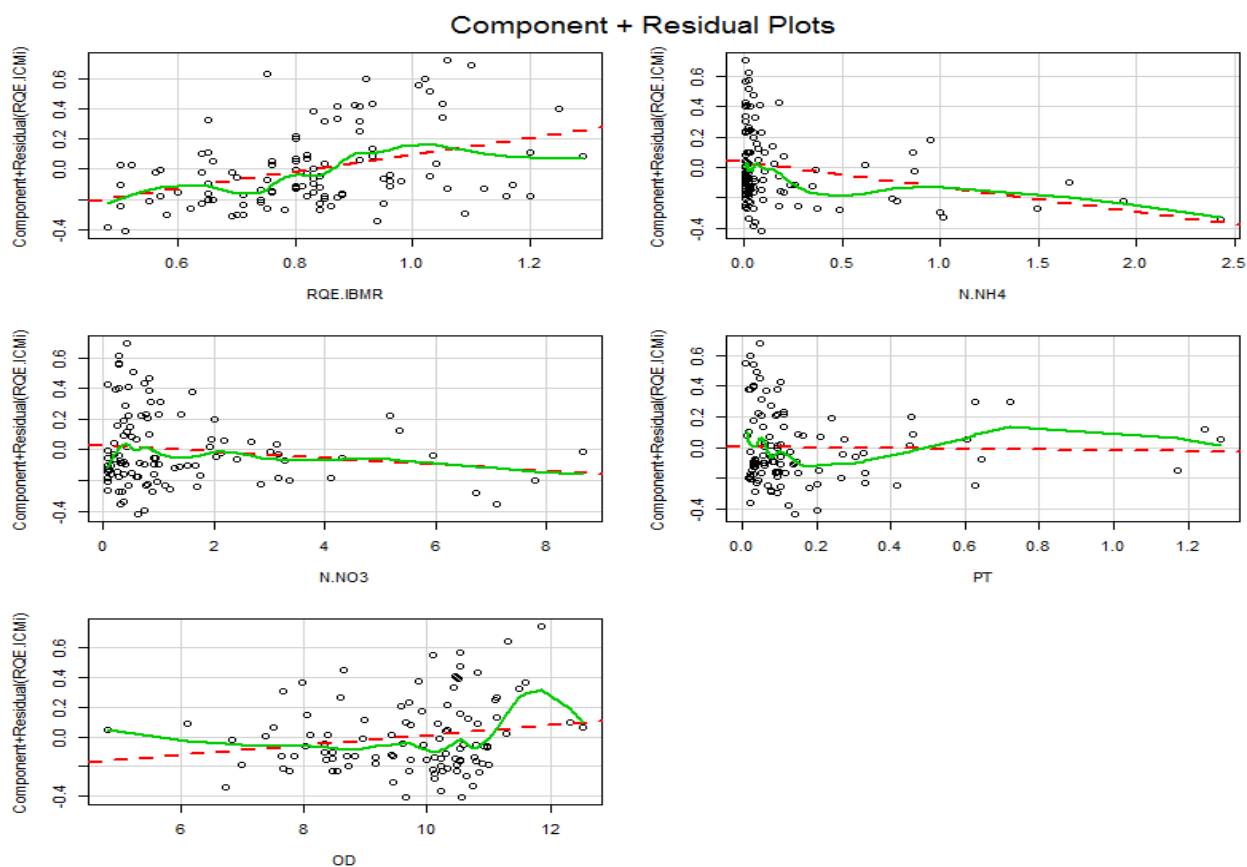


Figura 4.82. Partial residuals plot del modello caratterizzato da forme funzionali lineari per ogni regressore.

Forma funzionale	Lineare_IBMR		Log(IBMR)		Polinomiale_IBMR	
senza/con dummy	senza	con	senza	con	senza	con
Risultati						
b0_Intercetta(p-value)	-0,3004(0,4054)	-0,0770(0,833)	0,2497(0,5227)	0,5210(0,1939)	-0,284(0,5591)	-0,099(0,8358)
b1_(RQE_IBMR)	0,5095(0,00115)	0,545(0,0004)	0,4078(0,0015)	0,4411(0,0005)	0,456(0,6681)	0,621(0,5518)
b1_(RQE_IBMR^2)					0,0310(0,9592)	-0,044(0,9415)
b3_In(N-NO3)	0,0306(0,2763)	-0,0052(0,869)	0,0274(0,3287)	-0,009(0,772)	0,030(0,2832)	-0,006(0,8614)
b2_In(N-NH4)	-0,044(0,0257)	-0,044(0,023)	-0,0435(0,0281)	-0,043(0,025)	-0,0441(0,0266)	-0,044(0,0243)
b4_In(PT)	-0,0787(0,0110)	-0,079(0,009)	-0,0762(0,0144)	-0,077(0,012)	-0,0790(0,0132)	-0,079(0,0115)
b5_In(OD)	0,192(0,2531)	0,113(0,500)	0,1768(0,2984)	0,094(0,581)	0,1944(0,2639)	0,110(0,5273)
b7_Dummy		-0,139(0,023)		-0,141(0,022)		-0,139(0,0241)
R ² aggiustato	0,3605	0,3866	0,3574	0,3845	0,3541	0,3804
test-F (p-value)	12,95(9,78E-10)	12,13(3,289E-10)	12,79(1,234E-09)	12,04(3,872E-10)	10,69(3,763E-09)	10,3(1,212E-09)
p-value-test-Shapiro-Wi	3,685E-05	9,73E-05	2,61E-05	8,26E-05	3,69E-05	9,73E-05
p-value-test-Breush_Pag	0,0172	0,04163	0,014670	0,033	0,0329	0,06828
Forma funzionale	Lineare_N-NO3		Log_N-NO3		Polinomiale_N-NO3	
con/senza dummy	senza	con	senza	con	senza	con
Statistic summary						
b0_Intercept(p-value)	0,350(0,3749)	0,670(0,0905)	0,2497(0,5227)	0,5210(0,1939)	0,3008(0,452)	0,682(0,096)
b1_In(RQE_IBMR)	0,390(0,00221)	0,453(0,000315)	0,4078(0,0015)	0,4411(0,0005)	0,4063(0,002)	0,451(0,0004)
b2_(N-NO3)	-0,0097(0,5817)	-0,032(0,089)	0,0274(0,3287)	-0,009(0,772)	0,022(0,629)	-0,037(0,453)
b2_(N-NO3)^2					-0,004(0,4569)	0,0007(0,903)
b3_In(N-NH4)	-0,037(0,0626)	-0,039(0,040)	-0,0435(0,0281)	-0,043(0,025)	-0,038(0,056)	-0,039(0,042)
b4_In(PT)	-0,056(0,0731)	-0,060(0,044)	-0,0762(0,0144)	-0,077(0,012)	-0,059(0,062)	-0,060(0,048)
b5_In(OD)	0,164(0,335)	0,081(0,624)	0,1768(0,2984)	0,094(0,581)	0,173(0,313)	0,079(0,6399)
b6_Dummy		-0,171(0,003)		-0,141(0,022)		-0,173(0,004)
R ² aggiustato	0,3532	0,4016	0,3574	0,3845	0,3504	0,3957
test-F (p-value)	12,58(1,686E-09)	12,86(1,015E-10)	12,79(1,234E-09)	12,04(3,872E-10)	10,53(4,928E-09)	10,91(3,827E-10)
p-value-test-Shapiro-Wi	3,64E-05	0,0001419	2,61E-05	8,26E-05	2,81E-05	0,00015
p-value-test-Breush_Pag	0,006113	0,02745	0,014670	0,033	0,017	0,042
Forma funzionale	Lineare_N-NH4		Log_N-NH4		Polinomiale_N-NH4	
con/senza dummy	senza	con	senza	con	senza	con
Statistic summary						
b0_Intercept(p-value)	0,330(0,408)	0,651(0,111)	0,2497(0,5227)	0,5210(0,1939)	0,3533(0,3817)	0,676(0,103)
b1_In(RQE_IBMR)	0,415(0,0013)	0,452(0,000387)	0,4078(0,0015)	0,4411(0,0005)	0,4059(0,00199)	0,442(0,0006239)
b2_(N-NH4)	-0,136(0,053)	-0,156(0,024)	-0,0435(0,0281)	-0,043(0,025)	-0,2105(0,2693)	-0,233(0,210)
b2_(N-NH4)^2					0,0388(0,6748)	0,040(0,656)
b3_In(N-NO3)	0,016(0,563)	-0,024(0,432)	0,0274(0,3287)	-0,009(0,772)	0,016(0,5602)	-0,024(0,437)
b4_In(PT)	-0,074(0,023)	-0,070(0,025)	-0,0762(0,0144)	-0,077(0,012)	-0,070(0,0345)	-0,067(0,039)
b5_In(OD)	0,213(0,209)	0,118(0,485)	0,1768(0,2984)	0,094(0,581)	0,2082(0,2225)	0,113(0,5079)
b7_Dummy		-0,157(0,011)		-0,141(0,022)		-0,158(0,012)
R ² aggiustato	0,350	0,385	0,3574	0,3845	0,3452	0,38
test-F (p-value)	12,44(2,058E-09)	12,06(3,726E-10)	12,79(1,234E-09)	12,04(3,872E-10)	10,31(7,158E-09)	10,28(1,25E-09)
p-value-test-Shapiro-Wi	7,89E-05	2,76E-04	2,61E-05	8,26E-05	5,67E-05	2,07E-04
p-value-test-Breush_Pag	0,0262	0,0413	0,014670	0,033	0,040	0,06
Forma funzionale	Lineare_PT		Log_PT		Polinomiale_PT	
con/senza dummy	con	senza	con	senza	con	senza
Statistic summary						
b0_Intercept(p-value)	0,4068(0,3073)	0,7071(0,0850)	0,2497(0,5227)	0,5210(0,1939)	0,4351(0,2722)	0,7266(0,0753)
b1_In(RQE_IBMR)	0,4558(0,0005)	0,4877(0,0002)	0,4078(0,0015)	0,4411(0,0005)	0,4245(0,0012)	0,4575(0,0004)
b2_In(N-NH4)	-0,0527(0,0104)	-0,0504(0,0121)	-0,0435(0,0281)	-0,043(0,025)	-0,0465(0,0248)	-0,0447(0,0276)
b3_In(N-NO3)	0,0060(0,8272)	-0,0307(0,3173)	0,0274(0,3287)	-0,009(0,772)	0,0166(0,5532)	-0,0198(0,5267)
b4_(PT)	-0,1303(0,3170)	-0,1716(0,1822)	-0,0762(0,0144)	-0,077(0,012)	-0,6433(0,0717)	-0,6535(0,0614)
b4_(PT)^2					0,4364(0,1222)	0,4109(0,1367)
b5_In(OD)	0,1886(0,2798)	0,1018(0,5583)	0,1768(0,2984)	0,094(0,581)	0,2042(0,2398)	0,1185(0,4939)
b7_Dummy		-0,1517(0,0172)		-0,141(0,022)		-0,1480(0,0193)
R ² aggiustato	0,3247	0,3558	0,3574	0,3845	0,3342	0,3638
test-F (p-value)	11,2(1,338E-08)	10,76(3,317E-09)	12,79(1,234E-09)	12,04(3,872E-10)	9,867(1,566E-08)	9,659(4,116E-09)
p-value-test-Shapiro-Wi	1,93E-05	1,78E-05	2,61E-05	8,26E-05	1,069E-05	1,68E-05
p-value-test-Breush_Pag	0,0269	0,084	0,014670	0,033	0,0314	0,092
Forma funzionale	Lineare_OD		Log_OD		Polinomiale_OD	
con/senza dummy	con	senza	con	senza	con	senza
Statistic summary						
b0_Intercept(p-value)	0,4145(0,0408)	0,5905(0,0061)	0,2497(0,5227)	0,5210(0,1939)	2,0055(0,0142)	1,9429(0,0159)
b1_In(RQE_IBMR)	0,4040(0,0015)	0,4360(0,0005)	0,4078(0,0015)	0,4411(0,0005)	0,4674(0,0003)	0,4873(0,0002)
b2_In(N-NH4)	-0,0420(0,0343)	-0,0423(0,0295)	-0,0435(0,0281)	-0,043(0,025)	-0,0315(0,1171)	-0,0332(0,0935)
b3_In(N-NO3)	0,0278(0,3203)	-0,0078(0,8030)	0,0274(0,3287)	-0,009(0,772)	0,0253(0,3582)	-0,0061(0,8446)
b4_In(PT)	-0,0758(0,0148)	-0,0763(0,0122)	-0,0762(0,0144)	-0,077(0,012)	-0,0730(0,0171)	-0,0739(0,0143)
b5_OD	0,0247(0,2016)	0,0148(0,4467)	0,1768(0,2984)	0,094(0,581)	-0,3222(0,0623)	0,2832(0,0977)
b6_OD^2					0,0190(0,0437)	0,0164(0,0797)
b7_Dummy		-0,1380(0,0252)		-0,141(0,022)		-0,1230(0,0455)
R ² aggiustato	0,3609	0,3862	0,3574	0,3845	0,3804	0,399
test-F (p-value)	12,97(9,48E-10)	12,11(3,397E-10)	12,79(1,234E-09)	12,04(3,872E-10)	11,84(5,312E-10)	11,05(2,957E-10)
p-value-test-Shapiro-Wi	3,32E-05	8,41E-05	2,61E-05	8,26E-05	4,21E-05	4,52E-05
p-value-test-Breush_Pag	0,0126	0,0337	0,014670	0,033	0,0200	0,069

Figura 4.83. Confronto tra modelli ottenuti variando le forme funzionali di un modello base caratterizzato dalla logaritmica per tutte le variabili indipendenti. Il modello selezionato in giallo rappresenta quello che, di volta in volta, è stato ritenuto il migliore fra quelli confrontati.

	Modello caratterizzato dalla forma funzionale logaritmica per ogni regressore		
	Intercetta	Senza intercetta	Con dummy
Risultati			
b0_ Intercept(p-value)	0,2497(0,5227)		
b1_ ln(RQE_IBMR)	0,4078(0,0015)	0,377(0,00143)	0,377(0,0012)
b3_ ln(N-NO3)	0,0274(0,3287)	0,03(0,274)	0,002(0,9413)
b2_ ln(N-NH4)	-0,0435(0,0281)	-0,043(0,03)	-0,042(0,0307)
b4_ ln(PT)	-0,0762(0,0144)	-0,079(0,01)	-0,082(0,0074)
b5_ ln(OD)	0,1768(0,2984)	0,283(2,44e-11)	0,309(4,38e-12)
b6_ Dummy			-0,118(0,0446)
R ² aggiustato	0,3574	0,936	0,935
test-F (p-value)	12,79(1,234E-09)	298,2(< 2,2e-16)	256,8(<2.2e-16)
p-value-Shapiro-Wilk	2,61E-05	3.84e-05	1,15E-04
p-value-Breush-Pagan-	0,015	0,018	0,02579
	Modello caratterizzato dalla forma funzionale lineare per ogni regressore		
	Intercetta	Senza intercetta	Con dummy
Risultati			
b0_ Intercetta(p-value)	0,185(0,362)		
b1_ (RQE_IBMR)	0,470(0,0012)	0,520(0,00012)	0,695(3,13e-06)
b2_ (N-NO3)	-0,0217 (0,244)	-0,019(0,31)	-0,033(0,081)
b3_ (N-NH4)	-0,158 (0,0369)	-0,149(0,047)	-0,156(0,033)
b4_ (PT)	-0,045 (0,7576)	-0,031(0,8318)	-0,024(0,864)
b5_ (OD)	0,037 (0,0522)	0,051(2,89e-05)	0,047(0,00016)
b6_ Dummy			-0,133(0,0189)
R ² aggiustato	0,316	0,9287	0,9329
test-F (p-value)	10,88(2,08E-08)	282,2(<2,2e-16)	249,1(<2,2e-16)
p-value-Shapiro-Wilk	0,0001	0,00046	0,00048
p-value-Breush-Pagan-	0,02	0,0409	0,063

Figura 4.84. Confronto tra modelli caratterizzati dalla forma funzionale lineare per ogni regressore e dalla funzionale logaritmica per ogni regressore con intercetta, senza intercetta e con variabile dummy. Il modello in giallo rappresenta quello finale selezionato a cui è stata applicata procedura backward.

	Errori standard OLS			
	STEP1	STEP2	STEP3	STEP4
La variabile rimossa		PT	Intercetta	N-NO3
Risultati				
b0_Intercetta(p-value)	0,185(0,362)	0,179(0,375)		
b1_RQE_IBMR(p-value)	0,470(0,0012)	0,475(9,7E-04)	0,522(9,99E-05)	0,61(1,82E-05)
b2_(N-NH4)(p-value)	-0,158 (0,0369)	-0,168(0,0152)	-0,156(0,021)	-0,2(6,2E-04)
b3_(N-NO3) (p-value)	-0,0217 (0,244)	-0,024(0,138)	-0,021(0,194)	
b5_(PT)(p-value)	-0,045 (0,7576)			
b6_OD(p-value)	0,037 (0,0522)	0,037(0,050)	0,051(2,62E-05)	0,042(6,7E-04)
test-F(p-value)	10,88(2,08E-08)	13,7(5,41E-09)	356(<2,2E-16)	476,3(<2,2E-16)
	Errori standard robusti all'eteroschedasticità (HC3)			
	STEP1	STEP2	STEP3	STEP4
La variabile rimossa		PT	Intercetta	N-NO3
Risultati				
b0_Intercetta(p-value)	0,185(0,334)	0,179(0,346)		
b1_RQE_IBMR(p-value)	0,470(0,0034)	0,475(2,8E-03)	0,522(5,25E-04)	0,61(2,3E-05)
b2_(N-NH4)(p-value)	-0,158 (1,546E-05)	-0,168(2,266E-06)	-0,156(4,97E-07)	-0,2(2,4E-11)
b3_(N-NO3) (p-value)	-0,0217 (0,110)	-0,024(0,037)	-0,021(0,076)	
b5_(PT)(p-value)	-0,045 (0,715)			
b6_OD(p-value)	0,037 (0,050)	0,037(0,049)	0,051(7,64E-05)	0,042(4,7E-04)
test-F(p-value)	13,884(2,424E-10)	18,048(2,929E-11)	460(<2,2E-16)	467(<2,2E-16)

Tabella 4.30. Applicazione metodo backward con errori standard OLS ed errori standard robusti all'eteroschedasticità. Nella procedura backward è stata reintrodotta l'intercetta essendo non statisticamente significativa.

4.2.2.3. RQE-STAR-ICMi

Il primo modello costruito è stato quello relativo alla regressione lineare multipla, con forme funzionali lineari per ogni repressore.

4.2.2.3.1. Test di significatività

I risultati mostrano che il modello ha un buon R^2 corretto; circa il 65% della varianza in RQE-STAR-ICMi può essere spiegata dal modello di regressione (Tabella 4.31.). Poiché la statistica F test cade nella regione di rifiuto ($p\text{-value} < 0,05$), l'ipotesi nulla secondo la quale tutti i coefficienti di regressione siano nulli è rifiutata e si può asserire che il modello di regressione è statisticamente significativo (Tabella 4.31.). Osservando i p-values associati a ciascuna variabile, risulta evidente che le variabili statisticamente significative sono: RQE-IBMR, RQE-ISECI, N-NH₄, N-NO₃ (Tabella 4.31.). Per l'OD è possibile effettuare la considerazione

valida anche per gli altri RQE: esso non risulta essere significativo poiché l'intercetta è stata inclusa inizialmente nel modello di regressione (Figura 4.88.).

4.2.2.3.2. Grafici dei residui

Nel plot contro l'ISECI, i residui appaiono dispersi casualmente; nei grafici dei residui contro l'OD e il RQE-IBMR è possibile individuare pattern curvilinei che suggeriscono la presenza di un effetto quadratico non incluso nel modello. Il plot contro il RQE-ICMi mostra che i residui si distribuiscono in due clusters. I grafici dei residui relativi all' N-NH_4 , N-NO_3 e PT evidenziano sempre uno sbilanciamento rispetto all'asse delle x (Figura 4.85.).

4.2.2.3.3. Verifica della normalità della distribuzione degli errori

Il Q-Q Plots dei residui standardizzati mostra che i punti tendono ad allinearsi lungo la bisettrice del grafico suggerendo che i residui seguano una distribuzione normale (Figura 4.86.); l'istogramma è abbastanza simmetrico, sebbene evidenzi una coda leggermente più lunga a sinistra, lasciando ipotizzare la presenza di qualche outlier (Figura 4.86.). La funzione di distribuzione cumulativa empirica (ecdf) e normale si sovrappongono (Figura 4.86.) ed il risultato del test di Shapiro-Wilk conferma che la verifica di normalità è soddisfatta, essendo statisticamente significativo (Tabella 4.31.).

4.2.2.3.4. Confronto tra modelli

Poiché i partial residual plots evidenzierebbero un lieve effetto quadratico relativo all'OD e al RQE-IBMR, si è proceduto ad un approfondimento di ciò (Figura 4.87.). Essi mostrano anche che una relazione lineare per N-NO_3 e per l'ISECI approssima bene i dati e una minore correlazione con l' N-NH_4 ed il PT e confermerebbero la presenza di due clusters per il RQE-ICMi (Figura 4.87.). Sono stati confrontati il modello caratterizzato dalle forme funzionali polinomiali per il RQE-IBMR ed OD, logaritmica per N-NH_4 ed il PT, lineare per N-NO_3 , ISECI e RQE-ICMi con quello caratterizzato da forme funzionali lineari per ogni regressore (Figura 4.88.). I risultati mostrano che il primo modello presenta intercetta negativa; ciò è incompatibile con il significato stesso di RQE (essi sono espressi come un valore numerico tra zero ed uno, con valori tendenti ad uno o superanti l'unità, per rappresentare lo stato ecologico "elevato" e con valori che tendono a zero per rappresentare uno stato ecologico "cattivo"). L'intercetta corrisponde al valore medio della variabile dipendente quando le variabili indipendenti sono uguali a 0. In tal caso, poiché i regressori rappresentano fattori

indispensabili, è necessario porla pari a zero. Ma, il suo annullamento determina nel modello considerato non significativo l'ossigeno disciolto (Figura 4.88.). E' stato selezionato, allora, il modello con regressori lineari che è statisticamente significativo, rispetta le ipotesi di base delle regressioni ed inoltre è più aderente ad un modello teorico: in tale modello è evidente come l'annullamento dell'intercetta determina la significatività dell'OD (Figura 4.88.). L'introduzione della variabile dummy non apporta grandi modifiche, non essendo mai statisticamente significativa: la sua introduzione riduce la significatività globale del modello (Figura 4.88.). Al modello selezionato è stata applicata la procedura backward (Tabella 4.32.).

	Coefficiente di regressione	p-value
b0 Intercetta	0,103786	0,470423
b1_RQE_IBMR	0,216948	0,016223
b2_RQE_ICMi	-0,005382	0,92421
b3_RQE_ISECI	0,495686	0,000109
b4_N-NH4	-0,2099	4,34E-05
b5_N-NO3	-0,049173	0,000538
b6_OD	0,018155	0,155176
b7_PT	0,147217	0,254946
Valori delle principali statistiche		
R ² corretto	0,6474	
test F	23,82	< 2.2e-16
test di Shapiro-Wilk	0,98687	0,5224
test di Breusch-Pagan	4,1609	0,7611

Tabella 4.31. Risultati dei tests di significatività, di Shapiro-Wilk, di Breush Pagan inerenti al modello di regressione caratterizzato da forme lineari per ogni variabile indipendente.

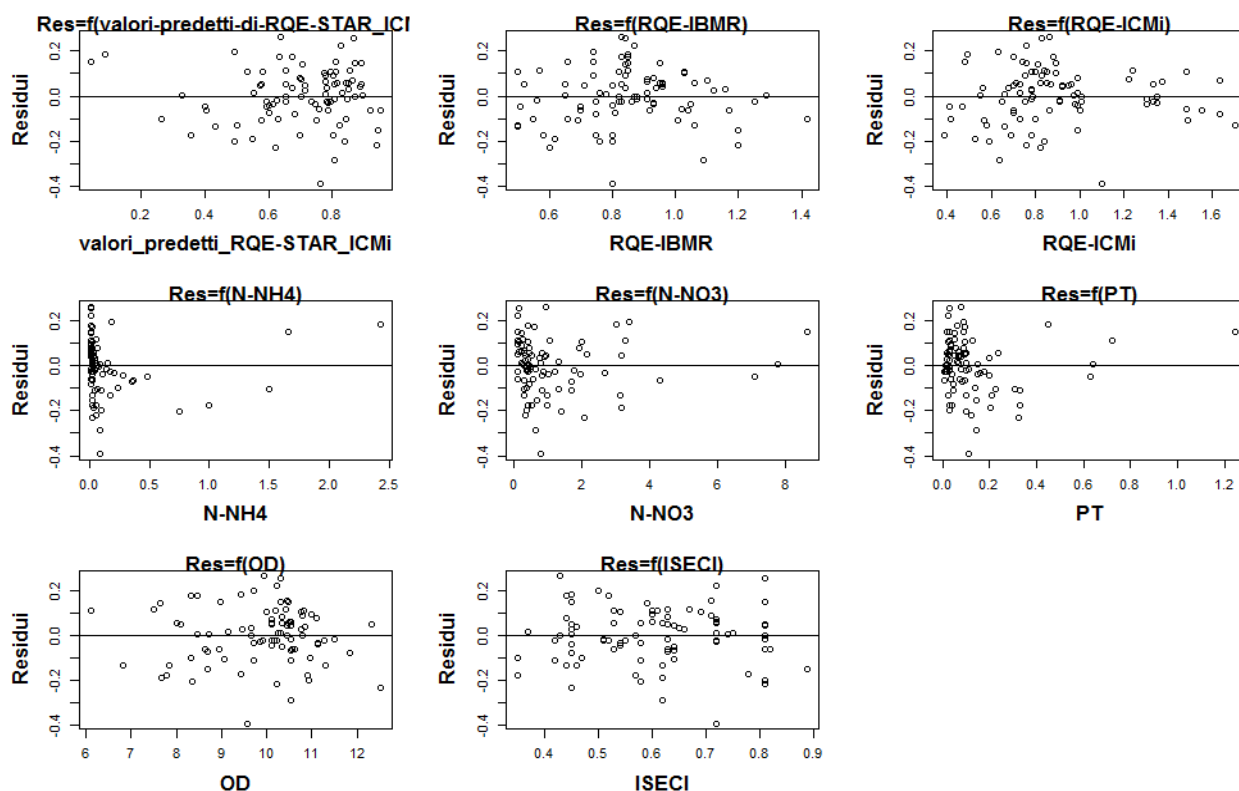


Figura 4.85. Grafici dei residui contro i valori predetti del RQE-STAR_ICMi e ciascuna variabile esplicativa inerenti al modello di regressione caratterizzato da forme lineari per ogni variabile indipendente.

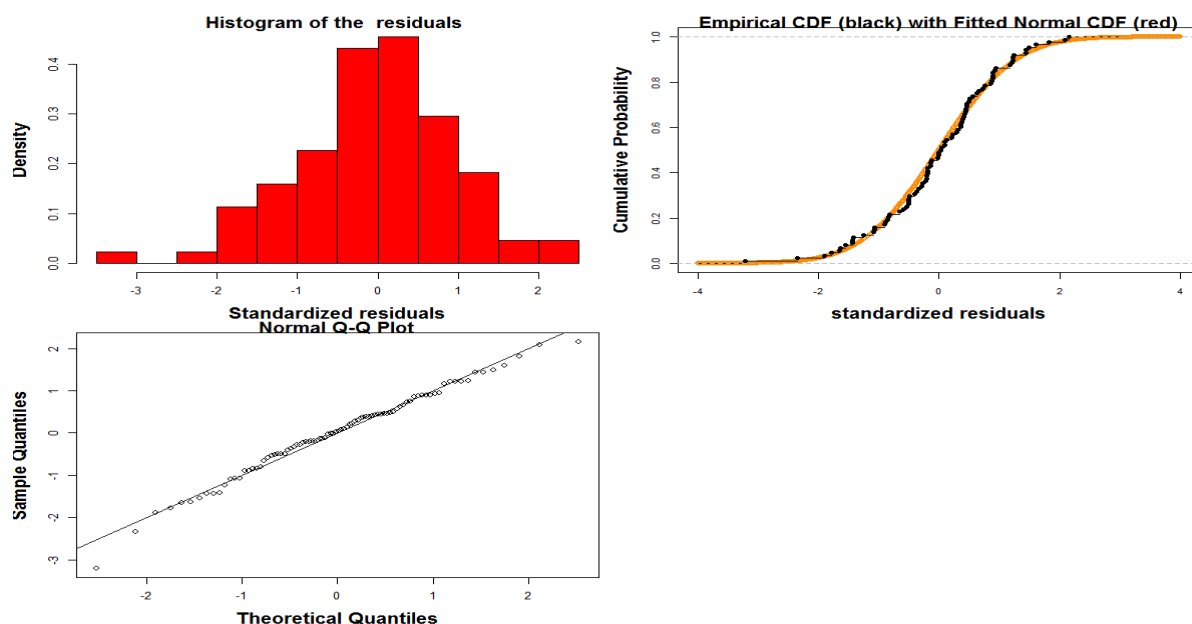


Figura 4.86. Verifica dell'assunzione di normalità della distribuzione degli errori attraverso l'istogramma dei residui standardizzati, confronto grafico tra la funzione di distribuzione cumulativa empirica (ecdf) e normale e normal Q-Q plot dei residui standardizzati.

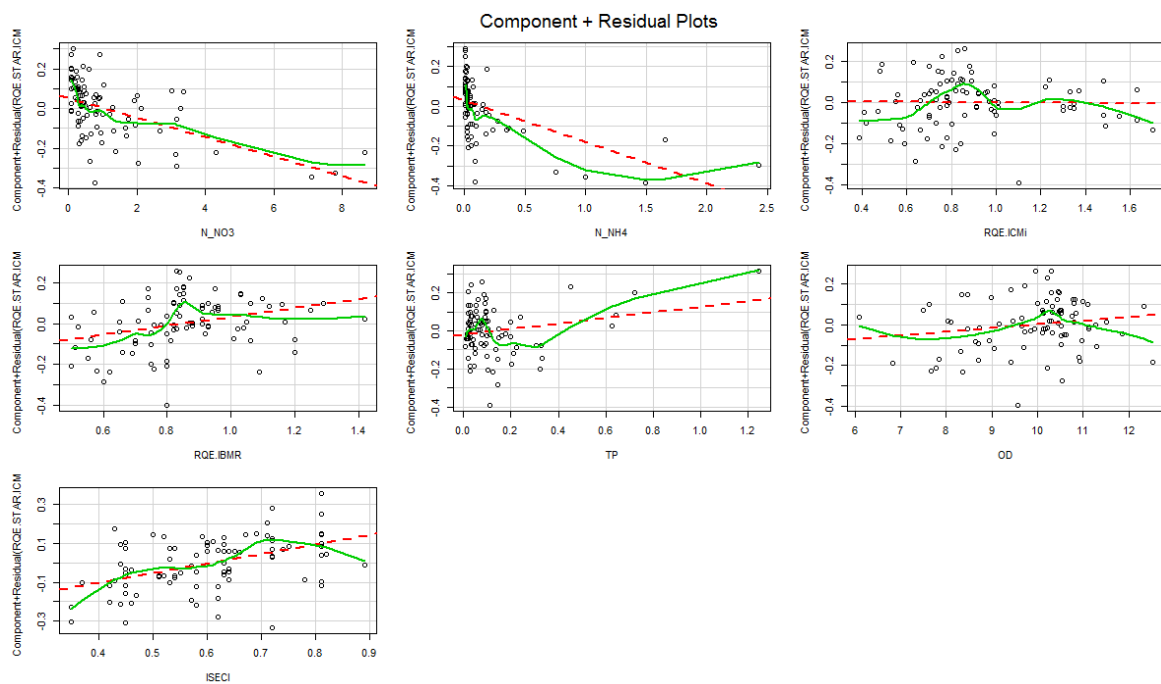


Figura 4.87. Partial residuals plot del modello caratterizzato da forme funzionali lineari per ogni regressore.

	Modello con forma funzionale lineare N-NO ₃ , ISECI, RQE-ICMi, polinomiale RQE-IBMR ed OD, logaritmica per N-NH ₄ e PT		
	Intercetta	Senza intercetta	Con dummy
Risultati			
b0_Intercetta(p-valu	-1,64(0,0045)		
b1_(RQE_IBMR)	0,63(0,209)	0,84(0,107)	-0,866(0,103)
b1_(RQE_IBMR^2)	-0,25(0,340)	-0,33(0,241)	-0,342(0,223)
b3_(N-NO ₃)	-0,03(0,0008)	-0,03(0,0005)	-0,031(0,004)
b2_ln(N-NH ₄)	-0,08(3,93E-12)	-0,07(1,72E-10)	-0,07(3,24E-10)
b4_ln(PT)	-0,007(0,664)	-0,009(0,585)	-0,008(0,622)
b5_(OD)	0,33(0,014)	-0,035(0,426)	-0,040(0,368)
b5_(OD^2)	-0,02(0,014)	0,002(0,431)	0,002(0,376)
b6_(ISECI)	0,36(0,0007)	0,376(0,0008)	0,373(0,0008)
b7_(RQE_ICMi)	-0,048(0,332)	-0,07(0,154)	-0,060(0,263)
b8_Dummy			0,024(0,443)
R ² aggiustato	0,77	0,98	0,9785
test-F (p-value)	34 (<2,2E-16)	447,3(<2,2E-16)	400,6 (<2,2E-16)
p-value-Shapiro-Will	0,9889(0,665)	0,9882(0,6122)	0,9871 (0,5356)
p-value-Breush-Paga	14,8(0,096)	11,49(0,175)	10,866(0,285)
	Modello caratterizzato dalla forma funzionale lineare per ogni regressore		
	Intercetta	Senza intercetta	Con dummy
Risultati			
b0_Intercetta(p-valu	0,103(0,470)		
b1_(RQE_IBMR)	0,217(0,016)	0,226(0,0113)	0,200(0,02868)
b3_(N-NO ₃)	-0,049(0,0005)	-0,047(0,0006)	-0,040(0,00558)
b2_(N-NH ₄)	-0,210(4,34E-05)	-0,200(4,81E-05)	-0,199(5,52E-05)
b4_(PT)	0,147(0,255)	0,143(0,265)	0,135(0,293)
b5_(OD)	0,018(0,155)	0,025(0,0041)	0,023(0,00893)
b6_(ISECI)	0,4496(0,0001)	0,530(8,56E-06)	0,514(1,70E-05)
b7_(RQE_ICMi)	-0,005(0,924)	-0,002(0,971)	0,016(0,778)
b8_Dummy			0,0397(0,2768)
R ² aggiustato	0,6474	0,9697	0,9697
test-F (p-value)	23,82(<2,2E-16)	402,8(<2,2E-16)	353,4(<2,2E-16)
p-value-Shapiro-Will	0,98687(0,5224)	0,98867(0,6482)	0,9860(0,4689)
p-value-Breush-Paga	4,1609(0,7611)	3,1146(0,7943)	3,411(0,8446)

Figura 4.88. Confronto tra modelli caratterizzati dalla forma funzionale lineare per N-NO₃, ISECI E RQE-ICMi, logaritmica per N-NH₄ e PT, polinomiale per OD e RQE-IBMR e il modello caratterizzato dalla forma funzionale lineare in ogni regressore con intercetta, senza intercetta e con variabile dummy. Il modello in giallo rappresenta quello finale selezionato a cui è stata applicata procedura backward

	STEP1	STEP2	STEP3	STEP4
Variabile rimossa		RQE-ICMi	Intercetta	PT
Risultati				
b0_Intercetta(p-value)	0,103(0,470)	0,103(0,471)		
b1_(RQE_IBMR)	0,217(0,016)	0,214(0,012)	0,225(0,0073)	0,224(0,008)
b3_(N-NO3)	-0,049(0,0005)	-0,049(0,0005)	-0,046(0,0006)	-0,036(0,0002)
b2_(N-NH4)	-0,210(4,34E-05)	-0,209(2,88E-05)	-0,200(3,2e-05)	-0,177(3,60E-05)
b4_(PT)	0,147(0,255)	0,148(0,250)	0,144(0,261)	
b5_(OD)	0,018(0,155)	0,018(0,148)	0,025(0,0015)	0,025(0,00137)
b6_(ISECI)	0,4496(0,0001)	0,497(8,11E-05)	0,530(4,46E-06)	0,533(6,17E-06)
b7_(RQE_ICMi)	-0,005(0,924)			
test-F (p-value)	23,82(<2,2E-16)	28,13(<2,2E-16)	475,7(<2,2E-16)	568,6(<2,2E-16)

Tabella 4.32. Applicazione metodo backward. Nella procedura backward è stata reintrodotta l'intercetta essendo non statisticamente significativa.

4.2.2.4. ISECI

Il primo modello costruito è stato quello relativo alla regressione lineare multipla, con forme funzionali lineari per ogni regressore.

4.2.2.4.1. Test di significatività

I risultati mostrano che nessuna variabile è significativa fatta eccezione per RQE-STAR-ICMi. Essi indicano, anche, che il modello ha un modesto R^2 corretto; circa il 35% della varianza in RQE-STAR-ICMi può essere spiegata dal modello di regressione (Tabella 4.33.). Poiché la statistica F test cade nella regione di rifiuto ($p\text{-value} < 0,05$), l'ipotesi nulla secondo la quale tutti i coefficienti di regressione siano nulli è rifiutata e si può asserire che il modello di regressione è statisticamente significativo (Tabella 4.33.).

4.2.2.4.2. Grafici dei residui

I grafici dei residui contro RQE-STAR-ICMi e RQE-IBMR mostrano un pattern casuale, un possibile lieve effetto quadratico per ossigeno, la presenza di due clusters per RQE-ICMi, uno sbilanciamento rispetto all'asse delle x per i nutrienti, accentuato per la variabile N-NH₄ (Figura 4.89.).

4.2.2.4.3. Verifica della normalità della distribuzione degli errori

Il Q-Q Plots dei residui standardizzati mostra che i punti tendono ad allinearsi lungo la bisettrice del grafico suggerendo che i residui seguano una distribuzione normale (Figura 4.90.); l'istogramma è abbastanza simmetrico, la funzione di distribuzione cumulativa

empirica (ecdf) e normale si sovrappongono (Figura 4.90.) ed il risultato del test di Shapiro-Wilk conferma che la verifica di normalità è soddisfatta, essendo statisticamente significativo (Tabella 4.33.).

4.2.2.4.4. Confronto tra modelli

Per l'ISECI, il confronto tra modelli è stato effettuato secondo la tabella indicata in Figura 4.92. E' stato scelto il modello di regressione lineare multipla, il modello con le forme funzionali suggerite sia dai partial-residual.plots e sia dall'analisi preliminare (logaritmica per i nutrienti, polinomiale per l'OD e lineare per RQE-STAR-ICMi, RQE-IBMR, RQE-ICMi) ed il modello di regressione multipla con forma funzionale lineare per tutte le variabili eccetto per l'OD caratterizzato da una polinomiale. Per ciascun tipo modello sono state valutate tre varianti: con intercetta, senza intercetta e con variabile dummy. I risultati evidenziano che per tutti i modelli con intercetta nessuna variabile è significativa fatta eccezione per RQE-STAR-ICMi. L'annullamento dell'intercetta determina, invece, anche la significatività dell'ossigeno disciolto sia nella forma funzionale lineare che polinomiale; l'aggiunta della variabile dummy non sembra apportare grandi modifiche, non essendo significativa ed influisce sulla significatività globale del modello facendola diminuire. Tutti i modelli sono caratterizzati dal test di Breush-Pagan e Shapiro-Wilk soddisfatti. Il modello selezionato è stato quello caratterizzato dalla forma funzionale lineare per tutte le variabili eccetto per l'OD in forma polinomiale, essendo considerato più corrispondente al modello teorico. A questo è stata applicata la procedura backward (Tabella 4.34.).

REGRESSIONE MULTIPLA TRA ISECI e RQE-STAR-ICMi, N-NH4, N-NO3, PT, OD, RQE IBMR, RQE-ICMi		
	Coefficienti di regressione	p-value
b0_Intercetta	0,34741	0,00301
b1_RQE_ICMi	-0,05259	0,26414
b2_RQE_STAR_ICMi	0,34620	0,00011
b3_RQE_IBMR	0,14244	0,06076
b2_N-NH4	0,00138	0,97564
b3_N-NO3	0,00279	0,82093
b4_PT	-0,03056	0,77820
b5_OD	-0,00576	0,59170
	Valori delle principali statistiche	p-value
R ² aggiustato	0,3477	
test-F	7,626	4,96E-04
test-Shapiro	0,992	0,889
test-Breush-Pagan	7,446	0,384

Tabella 4.33. Risultati dei tests di significatività, di Shapiro-Wilk, di Breus Pagan inerenti al modello di regressione caratterizzato da forme lineari per ogni variabile indipendente.

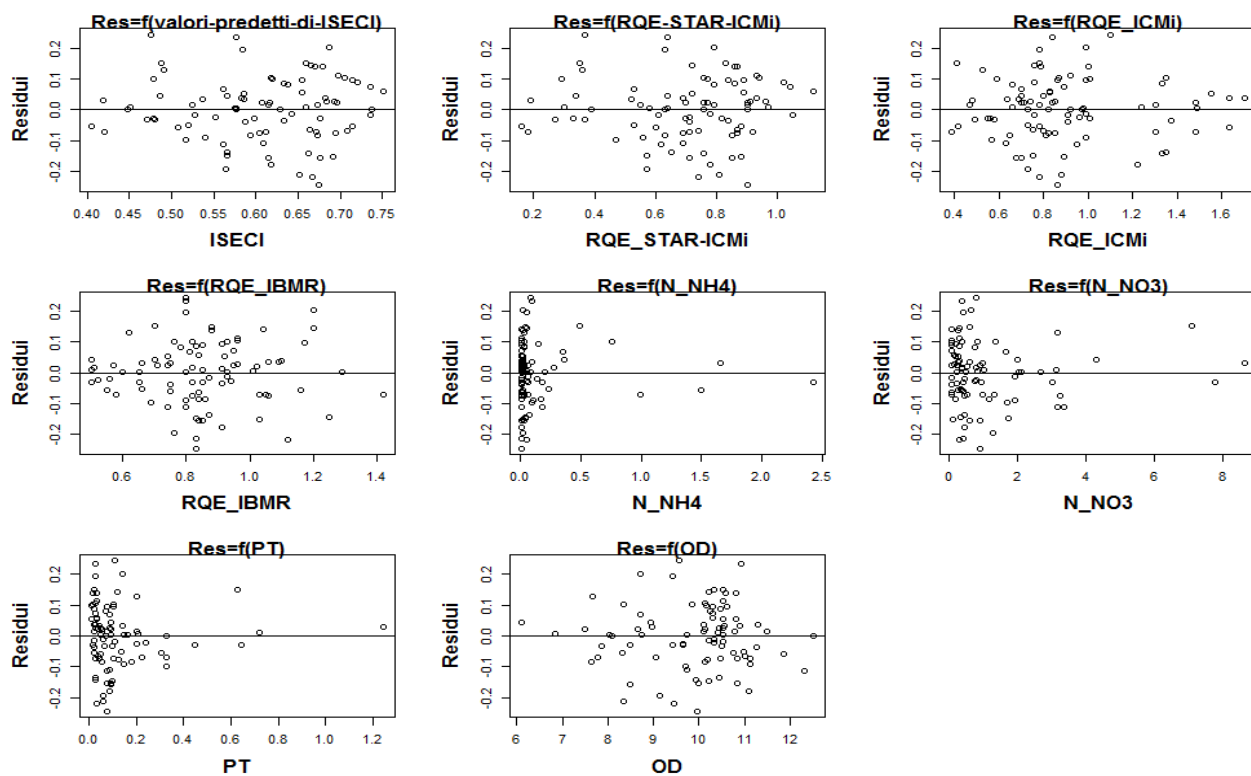


Figura 4.89. Grafici dei residui contro i valori predetti dell'ISECI e ciascuna variabile esplicativa inerenti al modello di regressione caratterizzato da forme lineari per ogni variabile indipendente.

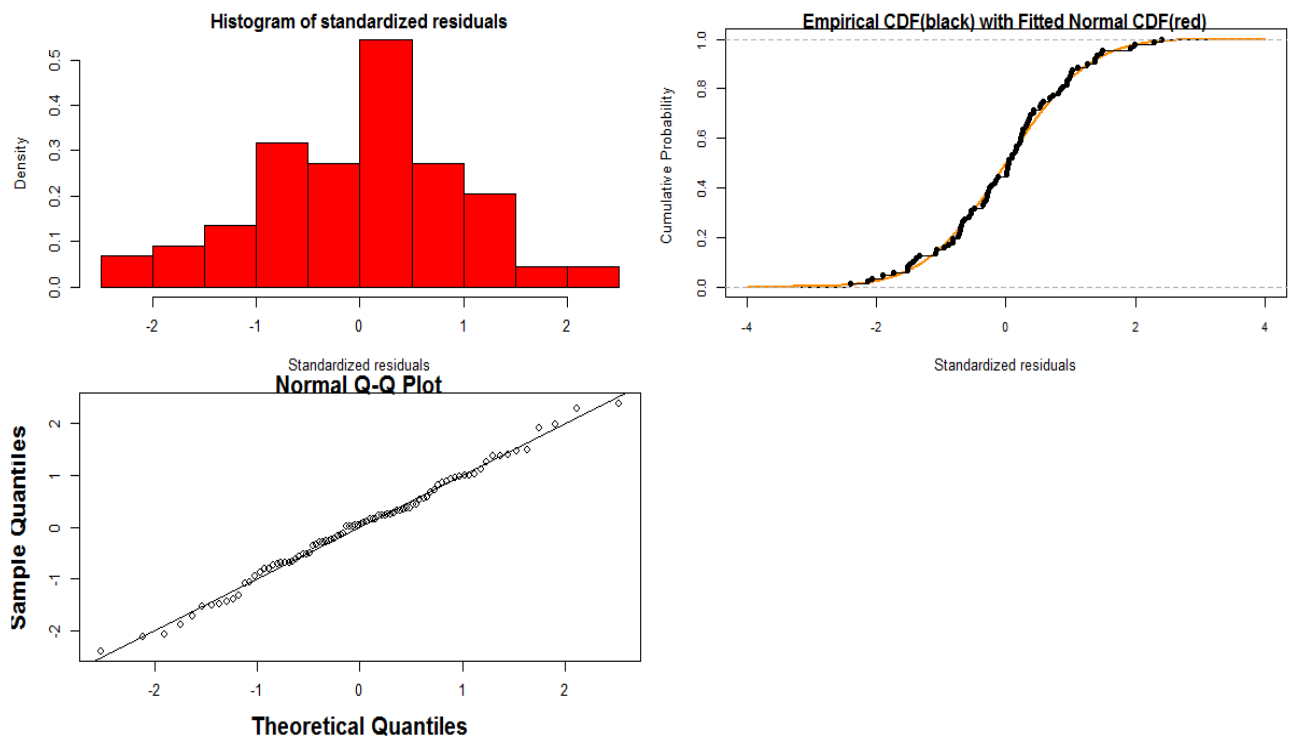


Figura 4.90. Verifica dell'assunzione di normalità della distribuzione degli errori attraverso l'istogramma dei residui standardizzati, confronto grafico tra la funzione di distribuzione cumulativa empirica (ecdf) e normale e normal Q-Q plot dei residui standardizzati.

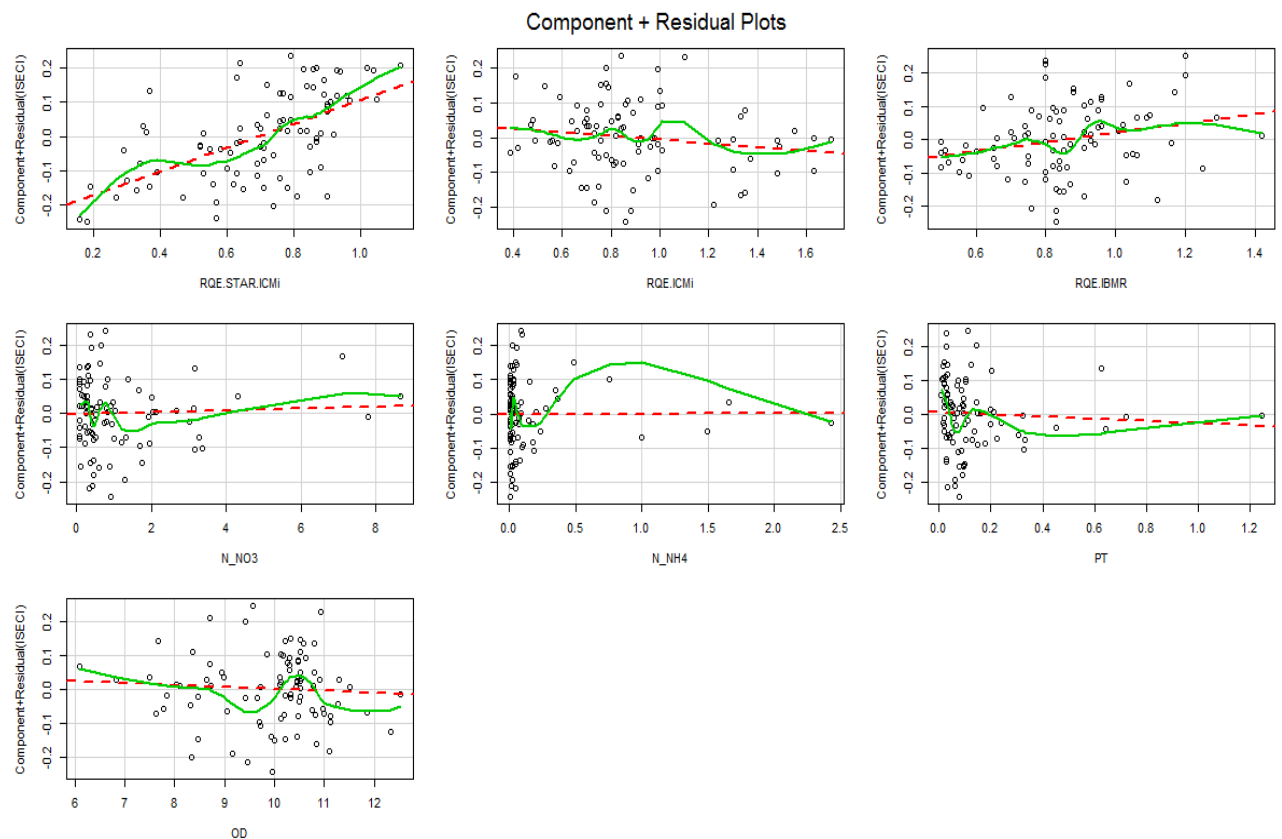


Figura 4.91. Partial residuals plot del modello caratterizzato da funzioni lineari per ogni regressore

Modelli di regressione lineare multipla			
	Con Intercetta	Senza Intercetta	Con variabile dummy
b0_Intercetta(p-value)	0,347(0,003)		
b1_RQE_ICMi(p-value)	-0,053(0,264)	-0,046(0,349)	-0,038(0,463)
b2_RQE_STAR_ICMi(p-value)	0,346(0,0001)	0,411(8,56e-06)	0,403(1,7e-05)
b3_RQE_IBMR(p-value)	0,142(0,061)	0,187(0,0174)	0,177(0,029)
b4_N-NH4(p-value)	0,001(0,976)	0,040(0,3819)	0,039(0,392)
b5_N-NO3(p-value)	0,003(0,821)	0,014(0,2521)	0,016(0,212)
b6_PT(p-value)	-0,031(0,778)	-0,051(0,653)	-0,054(0,639)
b7_OD(p-value)	-0,006(0,592)	0,018(0,0173)	0,018(0,024)
b8_Dummy(p-value)			0,017(0,600)
R ² aggiustato	0,3477	0,9674	0,967
test-F (p-value)	7,626 (4,96e-04)	374,1(< 2.2e-16)	324,5(< 2,2e-16)
test-Shapiro (p-value)	0,992(0,889)	0,9917(0,858)	0,9923(0,8897)
test-Breush-Pagan(p-value)	7,446(0,384)	7,744(0,258)	8,756 (0,271)
Regressione lineare con forme funzionali lineari per RQE_ICMi, RQE_STAR_ICMi, RQE_IBMR, logaritmiche per N_NH4, N_NO3, PT e polinomiale per OD			
	Con intercetta	Senza intercetta	Variabile dummy
b0_Intercetta(p-value)	0,392(0,517)		
b1_RQE_ICMi(p-value)	-0,068(0,173)	-0,064(0,193)	-0,074(0,150)
b2_RQE_STAR_ICMi(p-value)	0,318(0,005)	0,298(0,0057)	0,297(0,006)
b3_RQE_IBMR(p-value)	0,135(0,091)	0,124(0,111)	0,135(0,090)
b4_log(N-NH4)(p-value)	0,011(0,417)	0,008(0,548)	0,008(0,557)
b5_log(N-NO3)(p-value)	-0,008(0,577)	-0,008(0,588)	-0,014(0,410)
b6_log(PT)(p-value)	-0,023(0,139)	-0,016(0,138)	-0,023(0,140)
b7_OD(p-value)	-0,013(0,924)	0,072(1,94e-05)	0,075(1,75e-05)
b7_OD^2(p-value)	0,0003(0,960)	-0,004(0,00107)	-0,004(0,0009)
b8_Dummy(p-value)			-0,023(0,476)
R ² aggiustato	0,3746	0,972	0,972
test-F(p-value)	7,513(2,089e-07)	381,1(< 2.2e-16)	336,8(< 2.2e-16)
test-Shapiro(p-value)	0,988(0,6598)	0,988(0,634)	0,986(0,462)
test-Breush-Pagan(p-value)	11,167(0,1924)	8,126(0,3216)	7,487(0,4851)
Regressione multipla con forma lineare per tutte le variabili eccetto per l'OD caratterizzato da una polinomiale			
	Con Intercetta	Senza Intercetta	Con variabile dummy
b0_Intercetta (p-value)	0,112(0,845)		
b1_RQE_ICMi(p-value)	-0,049(0,312)	-0,047(0,318)	-0,046(0,354)
b2_RQE_STAR_ICMi(p-value)	0,34(0,0002)	0,338(0,0002)	0,337(0,0002)
b3_RQE_IBMR(p-value)	0,131(0,107)	0,126(0,101)	0,125(0,113)
b4_N-NH4(p-value)	-0,002(0,958)	-0,004(0,936)	-0,004(0,937)
b5_N-NO3(p-value)	0,003(0,824)	0,003(0,815)	0,003(0,810)
b6_PT(p-value)	-0,032(0,771)	-0,033(0,764)	-0,033(0,763)
b7_OD(p-value)	0,047(0,710)	0,072(0,0003)	0,072(0,0004)
b7_OD^2(p-value)	-0,003(0,675)	-0,004(0,003)	-0,004(0,003)
b8_Dummy(p-value)			0,002(0,947)
R ² aggiustato	0,341	0,971	0,97
test-F(p-value)	6,626(1,343e-06)	363(< 2.2e-16)	318,6(< 2,2e-16)
test-Shapiro(p-value)	0,991(0,828)	0,9906(0,786)	0,99059(0,7853)
test-Breush-Pagan(p-value)	10,19(0,252)	8,012(0,332)	8,275(0,4071)

Figura 4.92. Confronto tra il modello di regressione lineare multipla, il modello con forma funzionali logaritmica per i nutrienti, polinomiale per l'OD e lineare per RQE-STAR-ICMi, RQE-IBMR, RQE-ICMi ed il modello di regressione multipla con forma funzionale lineare per tutte le variabili eccetto per l'OD caratterizzato da una polinomiale. Il modello in giallo rappresenta quello finale selezionato a cui è stata applicata procedura backward

	STEP1	STEP2	STEP3	STEP4	STEP5	STEP6	STEP7
La variabile rimossa		N-NH4	INTERCETTA	N-NO3	PT	RQE-ICMi	RQE-IBMR
Risultati							
b0_Intercetta (p-value)	0,112(0,845)	0,116(0,836)					
b1_RQE_IBMR (p-value)	0,131(0,107)	0,130(0,105)	0,125(0,098)	0,124(0,098)	0,124(0,097)	0,103(0,149)	
b2_N-NH4 (p-value)	-0,002(0,958)						
b3_N-NO3 (p-value)	0,003(0,824)	0,003(0,804)	0,003(0,786)				
b4_RQE_STAR_ICMi (p-value)	0,34(0,0002)	0,342(1,8E-05)	0,341(1,63E-05)	0,335(8,00E-06)	0,343(6,08E-07)	0,333(8,69E-07)	0,375(2,17E-09)
b5_PT (p-value)	-0,032(0,771)	-0,034(0,726)	-0,036(0,705)	-0,02(0,788)			
b6_OD (p-value)	0,047(0,710)	0,046(0,711)	0,071(7,98E-05)	0,073(2,19E-05)	0,071(3,54E-06)	0,071(3,36E-06)	0,082(1,07E-09)
b6_OD^2 (p-value)	-0,003(0,675)	-0,003(0,675)	-0,004(0,002)	-0,004(0,001)	-0,004(0,007)	-0,004(0,0038)	-0,005(3,22E-05)
b7_RQE_ICMi (p-value)	-0,049(0,312)	-0,048(0,307)	-0,046(0,314)	-0,047(0,305)	-0,044(0,318)		
test-F(p-value)	6,626(1,343E-06)	7,668(4,576E-07)	420(<2,2E-16)	495,6(<2,2E-16)	601,4(<2,2E-16)	751,4(<2,2E-16)	988,1(<2,2E-16)

Tabella 4.34. Applicazione metodo blackward. Nella procedura blackward è stata reintrodotta l'intercetta essendo non statisticamente significativa.

4.2.3. Stima dei parametri con verifica assunzioni del sotto-modello finale

La tabella presenta i risultati finali delle quattro regressioni, dopo l'applicazione della procedura blackward (tabella 4.35.). Nella successiva descrizione delle quattro regressioni, è sottinteso che i coefficienti della regressione multipla non sono interpretati marginalmente poiché ciascuno di essi rappresenta le stime delle variazioni attese della variabile indipendente, mantenendo costanti gli altri (tabella 4.36.). Per tutte le quattro regressioni, è stato utilizzato lo stimatore OLS. Però, per la prima regressione le assunzioni di normalità ed omoschedasticità erano violate; poiché in grandi campioni, la normalità non costituisce un grande limite, sono stati soprattutto analizzati i metodi per ridurre gli effetti dell'eteroschedasticità. Come già indicato precedentemente, nella pratica è più comune usare gli errori standard consistenti all'eteroschedasticità rispetto allo stimatore WLS poiché essi non richiedono assunzioni sulla struttura della varianza. In particolare, è stato usato lo stimatore HC3. Attraverso questo approccio, il modello di regressione è sempre stimato usando gli OLS (i coefficienti regressione restano invariati); mentre gli errori standard stimati sono quelli consistenti all'eteroschedasticità. Come mostra la tabella 4.37., i valori degli errori standards OLS e HC3 sono molto vicini, ma N-NH₄ ha un' errore standard incorretto più grande di quello corretto. Essendo la variabile N-NH₄ maggiormente influenzata dalle correzioni per l'eteroschedasticità, sembrerebbe esserne la causa. Per tale motivo, solo per tale regressione sono stati rappresentati anche i plot dei residui del modello ottenuto dopo l'applicazione del metodo backward (figura 4.93). I test di significatività per tutti i coefficienti di regressione relazionati a RQE-ICMi non sono influenzati dalla stima corretta degli errori standard (tabella 4.37.). La formulazione finale mostra che il RQE-ICMi è positivamente

associato al RQE-IBMR e OD e negativamente al N-NH₄ (tabella 4.35., tabella 4.36.). Durante la procedura backward, due parametri erano stati eliminati: N-NO₃ e PT.

Per la seconda regressione mostrata in tabella 4.35., una nuova variabile definita NT era stata creata sommando ciascun valore di N-NH₄ con il corrispondente valore di N-NO₃. La formulazione finale mostra che il RQE-IBMR è positivamente correlato al RQE-ICMi (espresso in forma logaritmica), al termine lineare dell'OD, e al termine quadratico della variabile TN e negativamente correlato al termine lineare di TN e al termine quadratico dell'OD (tabella 4.35., tabella 4.36.). La variabile costituita dal PT non contribuiva significativamente al modello ed era stata eliminata. Le variabili più significative sono l'OD in forma quadratica e il RQE-ICMi (tabella 4.36.).

Per la terza regressione (tabella 4.35.), è stato preferito adottare una funzione di tipo lineare in ciascun termine. La formulazione finale mostra che il RQE-STAR-ICMi è positivamente associato all'ISECI, OD e al RQE-IBMR e negativamente relazionato al N-NH₄ ed N-NO₃ (tabella 4.35., tabella 4.36.); N-NH₄ ed ISECI sono le variabili più significative associate al RQE-STAR-ICMi (tabella 4.36.).

Per l'ultima regressione (tabella 4.35.), molte variabili non contribuivano significativamente al modello e furono eliminate. Esse erano: N-NO₃, N-NH₄, RQE-IBMR, RQE-ICMi e PT. La finale formulazione mostra che l'ISECI è correlato positivamente al RQE-STAR-ICMi, al termine lineare dell'OD e negativamente al termine quadratico dell'OD (tabella 4.35., tabella 4.36.).

E' interessante osservare come in tutti i modelli selezionati, rimuovendo l'intercetta, l'OD o aumentava di significatività (RQE_IBMR) o diventava significativo (RQE-ICMi, RQE-STAR-ICMi, ISECI). La dummy non è mai statisticamente significativa, tranne in alcuni modelli non presi in considerazione relativi al RQE-ICMi. Del resto, la sua introduzione avrebbe contrastato con la definizione di RQE.

4.2.3.1. I test di significatività

Per la prima regressione, l'R² aggiustato mostra che il 92% della varianza in RQE-ICMi può essere spiegata dal modello di regressione. L'usuale statistica F non può essere affidabile se c'è eteroschedasticità; così è stato calcolato una statistica robusta all'eteroschedasticità ed il

suo valore è pari a 467. Il corrispondente p-value è minore di 0,05, indicando che il modello spiega una quota significativa della varianza del RQE_ICMi (tabella 4.36.).

Per la seconda regressione, l' R^2 aggiustato mostra che il 97% della varianza in RQE-IBMR può essere spiegata dal modello di regressione. Il risultato del test-F è pari a 740 ed il corrispondente p-value è minore di 0,05, indicando che il modello spiega una quota significativa della varianza del RQE_IBMR (tabella 4.36.).

La terza regressione ha un R^2 che spiega il 97% della varianza in RQE-STAR-ICMi. Il valore della statistica F è 569 ed il corrispondente p-value che è minore di 0,05 indica significatività statistica (tabella 4.36.).

Per la quarta regressione, l' R^2 aggiustato di 0,96 implica che le variabili predittrici spiegano il 96% della varianza nell'ISECI. Il valore della statistica F è 988 ed il corrispondente p-value che è minore di 0,05 indica significatività statistica (tabella 4.36.).

4.2.3.2. Verifica dell'assunzione di normalità e omoschedasticità

La tabella 4.36. mostra che tutti i test diagnostici (per normalità ed omoschedasticità) applicati alle quattro regressioni sono soddisfatti ad eccezione della regressione associata al RQE-ICMi. Il normal Q-Q plot dei residui del RQE-ICMi mostra deviazioni dalla normalità (figura 4.94.). Per la corrispondente regressione, il test di Breush-Pagan indica la presenza di eteroschedasticità (tabella 4.36.). Allora, gli errori standard consistenti all'eteroschedasticità sono stati usati per dare risultati più accurati ai test e agli intervalli di confidenza. Questo metodo non ha richiesto l'uso di una trasformazione arbitraria della variabile dipendente. Per la regressione associata al RQE-IBMR, il Q-Q plot suggerisce una curva pesante a destra ma non da una forte indicazione di non-normalità (figura 4.94). In effetti, il test di Shapiro-Wilk è soddisfatto con un p-value più grande di 0,05 (tabella 4.36.). La stazione numero 66, essendo stata considerata un'outlier, è stata rimossa dai dati. Per la regressione relativa al RQE-ICMi e RQE-STAR-ICMi, i valori dei VIFs indicano che la collinearità non è alta per le variabili esplicative considerate (tabella 4.36.). Invece le regressioni relative all'ISECI e RQE-STAR-ICMi sono affette da collinearità poiché sono state usate regressioni polinomiali (tabella 4.36.). Ma, come suggerito da Berry (1987), nel caso di polinomiali, poiché le variabili non sono linearmente relazionate, l'analista non dovrebbe preoccuparsi della perfetta collinearità.

Var. risposta	Regressioni	osservazioni
RQE_ICMi	1. $RQE_ICMi = 0,61RQE_IBMR - 0,2N-NH_4 + 0,042OD$	107
RQE_IBMR	2. $RQE_IBMR = 0,21\ln(RQE_ICMi) - 0,06NT + 0,006NT^2 + 0,19 OD - 0,010 OD^2$	107
RQE_STAR_ICMi	3. $RQE_STAR_ICMi = -0,18N-NH_4 - 0,036N-NO_3 + 0,025OD + 0,22RQE_IBMR + 0,51ISECI$	88
ISECI	4. $ISECI = 0,08OD - 0,005OD^2 + 0,37RQE_STAR_ICMi$	88

Tabella 4.35. Regressioni che descrivono le relazioni tra ciascun RQE e le relative variabili indipendenti (altri RQE ed i 4 componenti del LIMeco)

Variabile risposta	variabili predittrici	β (p-value)	VIF	R^2	test-F (p-value)	test-Shapiro-Wilk (p-value)	test-Breush-Pagan (p-value)
RQE-ICMi	RQE-IBMR	0,61(2,3e-05)	1,186	0,92	467 (<2,2e-16)	0,953 (8,5e-04)	7,82 (0,02)
	N-NH4	-0,2(2,4e-11)	1,117				
	OD	0,042(4,7e-4)	1,175				
RQE-IBMR	$\ln(RQE_ICMi)$	0,21(4,19e-05)	1,48	0,97	740 (7,35e-78)	0,977 (0,06)	4,25(0,38)
	NT	-0,06(6e-3)	10,2				
	NT^2	0,0066(1,3e-02)	9,27				
	OD	0,19(<2e-16)	93,89				
	OD^2	-0,0098(3,95e-10)	95,11				
RQE-STAR_ICMi	N-NH4	-0,18(3,60e-05)	1,27	0,97	569(4,48e-62)	0,989(0,72)	2,93(0,57)
	N-NO3	-0,036(2,1e-04)	1,32				
	OD	0,025(1,3e-03)	1,13				
	ISECI	0,53(6,17e-06)	1,35				
	RQE-IBMR	0,22(7,6e-03)	1,34				
ISECI	RQE-STAR_ICMi	0,37(2,17e-09)	1,17	0,96	988(2,24e-65)	0,990 (0,80)	0,37(0,83)
	OD	0,08(1,07e-09)	134,80				
	OD^2	-0,005(3,22E-05)	133,10				

Tabella 4.36. Risultati del sottomodello finale.

		Stimatori degli errori standard			
		OLSE		Errori standard robusti (HC3)	
	β	SE	p-value	SE	p-value
RQE_IBMR	0,61	0,136	1,82e-05	0,145	2,3e-05
N-NH4	-0,2	0,057	6,2e-04	0,030	2,4e-11
OD	0,042	0,012	6,7e-04	0,012	4,7e-04

Tabella 4.37. Confronto tra gli errori standard OLS e HC3 e p-value associati per ciascun coefficiente β nell'analisi del RQE-ICMi

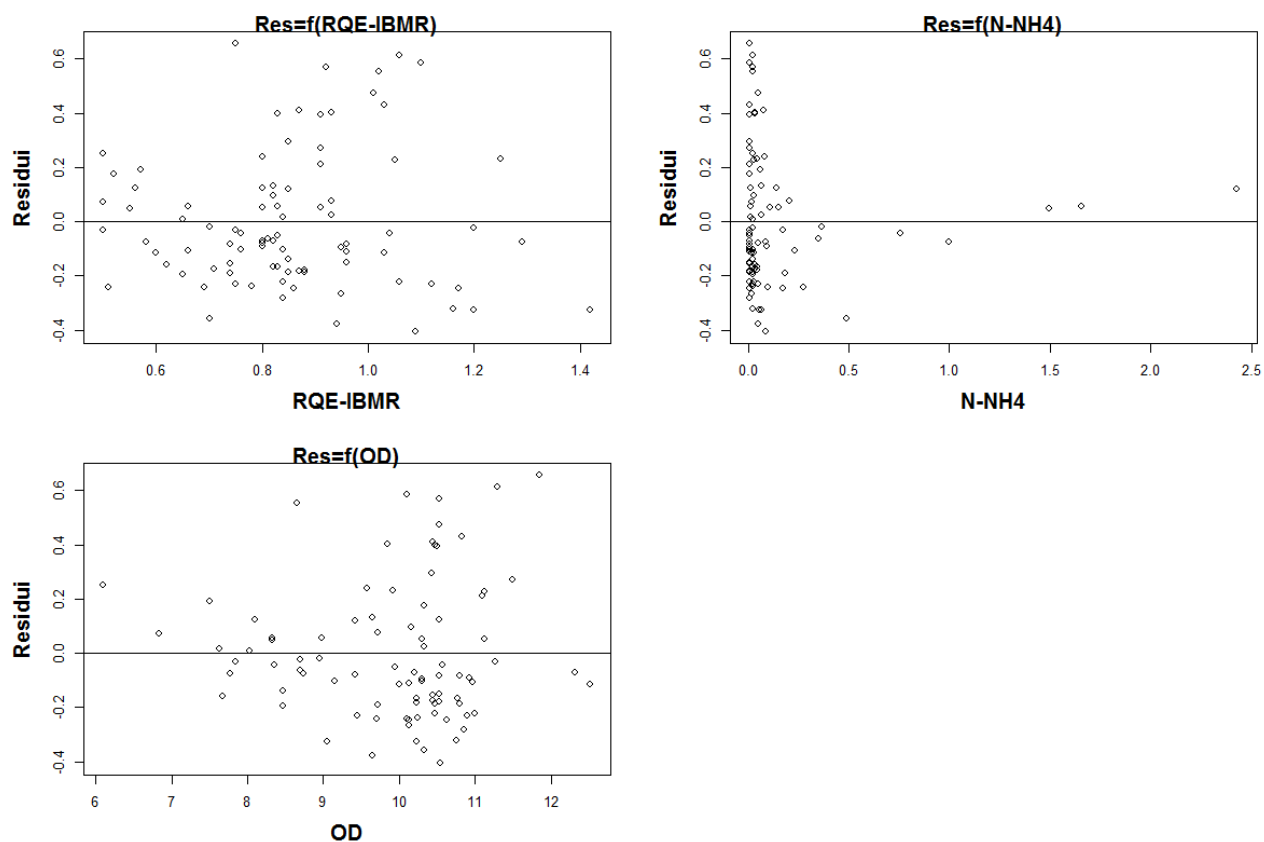


Figura 4.93. Grafici dei residui contro le relative variabili esplicative della regressione del RQE-ICMi dopo applicazione della procedura backward.

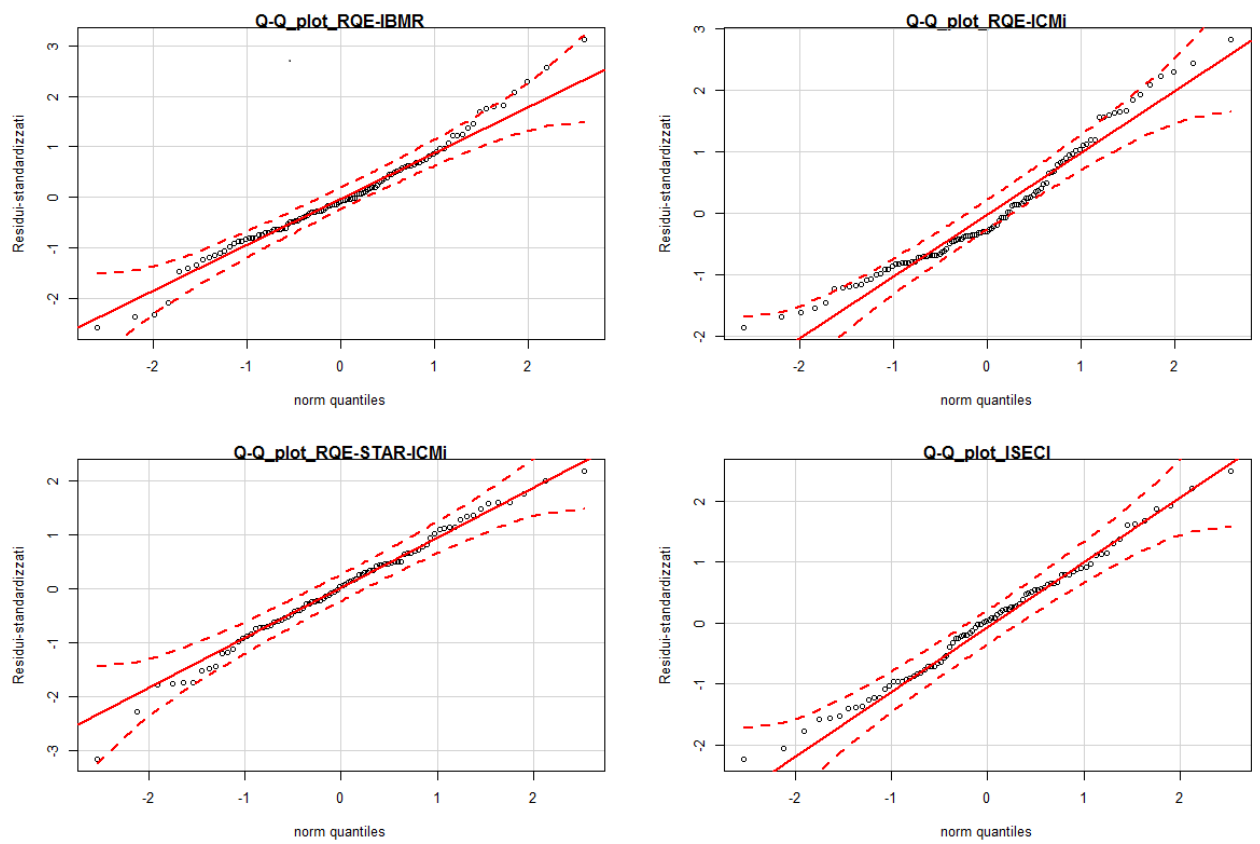


Figura 4.94. QQ plots dei residui standardizzati contro i quantili della distribuzione normale per ciascun RQE. Le linee rosse e tratteggiate sono i limiti degli intervalli di confidenza al 95%.

5. Discussioni

Per le seguenti discussioni, è importante considerare che il modello di regressione indagato stima la variabilità spaziale degli indici RQE espressi come valori medi annuali in punti singolari di interesse (costituiti dalle stazioni di monitoraggio); questa circostanza implica che sia necessario indicare gli impatti in valori medi annuali. E' anche importante notare che i RQE sono espressi come valori numerici compresi tra 0 e 1: valori prossimi a 1 tendono allo stato ecologico elevato, quelli prossimi allo zero allo stato ecologico cattivo. E' noto che la disponibilità dei dati di monitoraggio limita fortemente lo sviluppo dei modelli ed è ovvio che un punto di campionamento non è sufficiente a caratterizzare lo stato fisico-chimico di un corpo idrico soprattutto se sottoposto ad alti livelli di impatto antropico (Halasz et al., 2007) e a rappresentare la variabilità spaziale e temporale dei parametri di qualità dell'acqua; ma, un corretto campionamento sui corpi idrici richiederebbe un numero di campioni che non sono finanziariamente convenienti. La mancanza di un quadro conoscitivo, sia iniziale che a seguito delle modellazioni sviluppate, spazialmente distribuito evidenzia un forte limite dell'approccio utilizzato inerente al problema relativo all'eterogeneità del bacino. In effetti, la rappresentazione a parametri distribuiti, si basa sull'idea di considerare il sistema come un insieme discretizzato di unità omogenee che tengono conto dell'eterogeneità spaziale facendo riferimento alla rappresentazione analitica dei processi fisici agenti su ciascuna unità. Ma modelli che si basano su tale rappresentazione richiederebbero una grande disponibilità di dati e uno sforzo computazionale intenso.

Inoltre, la scelta del particolare set di variabili d'ingresso è legata all'approccio che la WFD richiede per la classificazione dello stato ecologico. Infatti, in Italia, la valutazione dello stato ecologico si basa principalmente sui parametri biologici e sui parametri chimico-fisici che supportano gli elementi biologici (in Italia: LIMeco). Altri parametri come il BOD e la temperatura non sono stati presi in considerazione per il loro forte legame con il DO. Il sub-modello PARBLEU mira a interpretare la procedura di classificazione di stato ecologico: collega i RQE agli elementi chimico-fisici, che possono essere facilmente monitorati attraverso l'attuazione di misure per contrastare gli effetti delle pressioni. In particolare, nella regione Abruzzo, le pressioni significative consistono in prelievi di acqua, carico di inquinamento da impianti di trattamento delle acque reflue e attività agricole: queste pressioni portano ad impatti significativi sulle sostanze nutritive e sul'OD.

Inoltre, le seguenti discussioni riflettono la posizione espressa da Dybas (2001) sulla necessità di investigare le risposte della comunità biologica alle differenti sollecitazioni ambientali secondo il nuovo paradigma della biocomplexità.

“Earth is a living, ever-changing planet with interconnecting threads everywhere. Complexity is a defining part of these threads. All levels of biological organization, from molecules to cells to organisms to ecosystems, exhibit properties that are more than the sum of their parts. Understanding how complex systems emerge from the interaction of biological entities at all levels and with the external environment is critical to ecosystem functioning” (Dybas, 2001).

5.1. Le violazioni delle assunzioni nella regressione del RQE-ICMi

I risultati mostrano che i test diagnostici per normalità ed omoschedasticità applicati alle quattro regressioni sono soddisfatti eccetto per la regressione ottenuta per predire la valutazione di stato basata sulle comunità diatomiche. Per questa regressione, i plots dei residui suggerivano eteroschedasticità collegata alla variabile N-NH₄. Così, come mostrato nella tabella, i risultati per questo regressore sono più influenzati dalle correzioni per l'eteroschedasticità. Questa violazione potrebbe essere relazionata al fatto che le diatomee presentano uno dei cicli di vita più brevi fra tutti gli elementi biologici (Rott, 1991) e che rispondono ai parametri chimici a differenti scale rispetto agli altri. E' generalmente noto che esse riflettono i cambiamenti ambientali rapidamente e segnalano velocemente sia un aumento del grado di inquinamento e sia il successo del ripristino di un habitat (Stevenson et al., 2010). Infatti i risultati di uno studio condotto da Taylor et al. (2007) suggerivano che gli indici basati sulle diatomee, in generale, hanno la migliore correlazione con i dati chimici mediati per un mese.

5.2. L'interpretazione delle doppie frecce

Le doppie frecce indicate nella specificazione iniziale del sotto-modello sono confermate dai risultati della regressione. Le regressioni hanno mostrato il forte peso delle interconnessioni tra diatomee e piante acquatiche nella valutazione degli indici corrispondenti. Sulla base del ruolo delle macrofite e delle diatomee nell'ecosistema, i risultati potrebbero essere interpretati in questo modo: le macrofite contribuiscono alla struttura fisica degli habitat acquatici

colonizzati e la loro complessità morfologica influenza la produttività primaria delle microalghe in un habitat (Fernandes et al., 2016). D'altra parte, le diatomee sono la base della catena trofica e alterazioni delle loro comunità può avere ripercussioni a livelli trofici più alti. Questa forte interconnessione potrebbe essere dovuta al fatto che i due indici danno informazioni ridondanti (Schneider et al., 2012) poiché pattern di comunità per macrofite e diatomee potrebbero essere strettamente relazionati in questa area di studio.

I risultati evidenziano anche la forte interconnessione tra ISECI e RQE-STAR-ICMi: essi mostrano una risposta positiva della valutazione di stato basata sull'indice RQE-STAR-ICMi, all'aumentare della valutazione di stato espressa dall'ISECI ed una risposta positiva della valutazione di stato basata sull'ISECI all'incrementare della valutazione di stato basata sul RQE-STAR-ICMi.

I macroinvertebrati costituiscono un'importante fonte di cibo per numerosi pesci ma essi hanno un ruolo chiave sulla decomposizione della materia (Wallace and Webster, 1996); in effetti, il materiale organico morto è una delle principali fonti di energia per le specie bentoniche in acque poco profonde (Covich, 1999). I risultati mostrano soprattutto l'influenza di pesci e macrofite su di loro. I macroinvertebrati sono influenzati dalla sostanza organica derivante dagli scarti fecali dei pesci e corpi di animali morti e dalle macrofite per l'alto tasso di produzione di biomassa. Infatti, le macrofite rappresentano un'importante risorsa di cibo per gli organismi acquatici, fornendo sia materiale organico vivo che morto (Thomaz and Cunha, 2010). In aggiunta, essi offrono un habitat agli organismi acquatici; le caratteristiche architettoniche di questo habitat possono influenzare la diversità, densità e distribuzione delle specie di invertebrati (Carpenter and Lodge 1986).

5.3. L'effetto di N-NH₄ ed N-NO₃ sugli indici biotici

L'indice IBMR è un metodo basato sulle macrofite per valutare lo stato trofico dei fiumi (Haury et al., 2006). Questa nuova variabile ottenuta sommando i valori di N-NH₄ con i corrispondenti valori di N-NO₃ era significativa. Il partial-residual plot suggeriva una relazione polinomiale per la variabile indipendente TN. In questa regressione, i valori del RQE-IBMR decrescono all'incrementare delle concentrazioni di TN (ma ad un tasso decrescente), mostrando che la valutazione di stato basata sulle macrofite sta peggiorando. Infatti, è generalmente noto che concentrazioni elevate di azoto provenienti da attività umane possono stimolare o favorire lo sviluppo e la proliferazione dei produttori primari

(fitoplancton, alghe bentoniche, macrofite), contribuendo al diffuso fenomeno della eutrofizzazione degli ecosistemi acquatici (Camargo and Alonso, 2006).

Per la regressione inerente al RQE-STAR_ICMi, è stato preferito adottare una funzione di tipo lineare in ciascun termine sebbene per l' N-NH_4 , l'approssimazione lineare non sembrerebbe adattarsi bene ai dati. La relazione tra nutrienti e macroinvertebrati è mostrata in molti studi (Wang et al., 2007; Miltner, 2010). Il risultato della regressione indica una risposta negativa dell'indice STAR-ICMi (espresso come RQE) all'incrementare delle concentrazioni di N-NH_4 e N-NO_3 . Gli effetti tossici dovuti agli alti livelli di composti di azoto inorganici sugli invertebrati sono descritti da Camargo ed Alonso (2006).

5.4. L'effetto del PT sugli indici biotici

La variabile PT non è risultata essere mai significativa tranne in alcuni modelli non selezionati aventi come variabile dipendente il RQE-ICMi. Questa mancanza di significatività potrebbe essere dovuta alla forte correlazione con il N-NO_3 . Graham (2003) dimostra che anche bassi livelli di collinearità causano l'esclusione di variabili importanti durante la creazione del modello.

Ma il fosforo costituisce un componente vitale degli ecosistemi acquatici e gioca un ruolo fondamentale nell'eutrofizzazione dei corpi idrici recettori (Correll, 1998). Negli ecosistemi acquatici, esso si accumula nei sedimenti dei corpi idrici ed il suo ciclo differisce da quello dell'ossigeno e dell'azoto che presentano cicli molto più veloci (Holtan et al., 1998). In aggiunta, processi ecologici, come quelli che si verificano nel substrato e la velocità dell'acqua possono modificare la disponibilità di PT e allora la nutrizione delle piante (Thiebaut, 2008). Così, dati relazionati ad altri parametri, ora non disponibili, dovrebbero essere presi in considerazione.

5.5. L'effetto dell'OD sugli indici biotici

La variabile OD era significativa per tutti gli elementi. Tutte le regressioni mostrano una risposta positiva dell'indice corrispondente all'incrementare dell'ossigeno disciolto (in alcune però, ad un tasso decrescente). Dopotutto, gli effetti avversi dovuti al declino nelle concentrazioni di ossigeno disciolto sono generalmente noti (Killgore and Hoover, 2001;

Camargo and Alonso, 2006). Il partial-residual-plot suggeriva una relazione polinomiale dell'ossigeno disciolto nelle regressioni del RQE-IBMR e ISECI. Questo significa che l'effetto marginale dell'OD non è costante in queste regressioni; le regressioni mostrano che l'effetto di un cambiamento è più grande a bassi che ad alti livelli di ossigeno. In altre parole, un eccesso di ossigeno apparirebbe ingiustificato, in relazione ai costi necessari per raggiungerlo. E' interessante osservare come in tutte le regressioni selezionate, rimuovendo l'intercetta, l'OD o aumentava di significatività (RQE_IBMR) o diventava significativo (RQE-ICMi, RQE-STAR-ICMi, ISECI). Questo risultato è perfettamente ovvio.

5.6. L'effetto della variabile IDRO-ECOREGIONE sugli indici biotici

La variabile IDRO-ECOREGIONE, introdotta con il fine di differenziare i corpi idrici appartenenti alla zona interna di montagna da quelli che fluiscono nella pianura costiera non era significativa per tutti gli elementi biologici tranne per il RQE-ICMI. In tale regressione l'introduzione di tale variabile sembra ridurre l'eteroschedasticità, lasciando ipotizzare anche la possibilità di due campioni provenienti da due diverse popolazioni. E' stato preferito, però, per il RQE-ICMi, non includerla nella regressione poiché la sua introduzione avrebbe contrastato con la definizione di RQE; quest'ultimo già prende in considerazione la specificità della stazione campionata. Per i restanti (RQE-STAR-ICMi, RQE-IBMR, ISECI), la mancanza di significatività rispecchia il significato di RQE. Però, il confronto tra modelli ha anche lasciato ipotizzare una forte correlazione tra la variabile N-NO₃ ed IDRO-ECOREGIONE che potrebbe determinare la mancanza di significatività di quest'ultima. In effetti, è stato verificato che le più alte concentrazioni di N-NO₃ sono registrate nell'area costiera e nei bassi corsi dei fiumi.

5.7. L'utilità del sotto-modello

Dopo aver calibrato il modello di regressione e dopo che le maggiori assunzioni sono state verificate, differenti scenari sono stati simulati. Come già detto, la risoluzione delle regressioni stimate è stata svolta in modo iterativo usando un metodo che convergeva alla soluzione attraverso successive simulazioni con controllo dei residui. I risultati sono stati riportati con una stima puntuale di ciascun RQE per ciascuna stazione, insieme con l'associato intervallo di confidenza e con il gap espresso in percentuale tra i valori osservati e simulati. Per meglio

cogliere la specificità di tali risultati, è stato ritenuto opportuno studiare le interrelazioni esistenti in termini spaziali: i risultati sono stati analizzati e visualizzati su mappa in ambiente Gis. Le figure 5.1., 5.3., 5.5., 5.7. mostrano gli effetti su ciascun indice implementando la seguente misura: incremento del 10% di OD e riduzione di N-NH₄ e N-NO₃ del 10%. Il confronto tra i valori osservati e simulati è stato effettuato prendendo in considerazione i limiti di classe per ciascun RQE. La valutazione dell'incertezza, quantificata attraverso gli intervalli di confidenza per la media di ciascun RQE, è stata svolta solo per le stazioni in cui il valore simulato era incluso nei limiti di classe buono o alto; in effetti l'obiettivo finale della WFD è rappresentato dal raggiungimento dello stato ecologico "buono" per tutti i corpi idrici. Per tale motivo, l'intervallo di confidenza al 95% associato a ciascuna stazione (il cui valore simulato ricadeva nei limiti di classe "buono" o "alto") è stato analizzato ed usato per la seguente valutazione: se il limite inferiore appartiene alla classe moderata o povera o cattiva allora la valutazione di stato basata su ciascun elemento biologico non è soddisfatta (le stazioni di monitoraggio in rosso nelle figure 5.2., 5.4., 5.6., 5.8.).

In definitiva, il sotto-modello permette di classificare lo stato ecologico di un corpo idrico simulando le variazioni nei componenti del LIMeco attraverso l'implementazione del Programma di Misure. In presenza di scarsità di risorse la decisione di investire dovrebbe prendere in considerazione l'incertezza e non il valore medio degli indici biotici. Quando il limite inferiore del valore simulato di ciascun RQE è lontano dal limite inferiore dello stato ecologico buono, l'incertezza di conseguire l'obiettivo è alta e quindi ulteriori misure (qui non prese in considerazione) dovrebbero essere applicate per ripristinare il corso d'acqua.

Tutte le figure mostrano che nei bassi corsi dei fiumi e della zona costiera si addensano le stazioni di monitoraggio caratterizzate da valori dei RQE appartenenti alle classi: Sufficiente, Scarsa e Cattiva. L'applicazione della misura di incremento del 10% di OD e riduzione di N-NH₄ e N-NO₃ del 10% sembra essere sufficiente solo per il RQE-ICMi. Per i restanti RQE, sarebbe necessario effettuare altre simulazioni per individuare la misura che determinerebbe il raggiungimento dello stato buono sia per il valore medio stimato sia per il limite inferiore del corrispondente intervallo di confidenza.

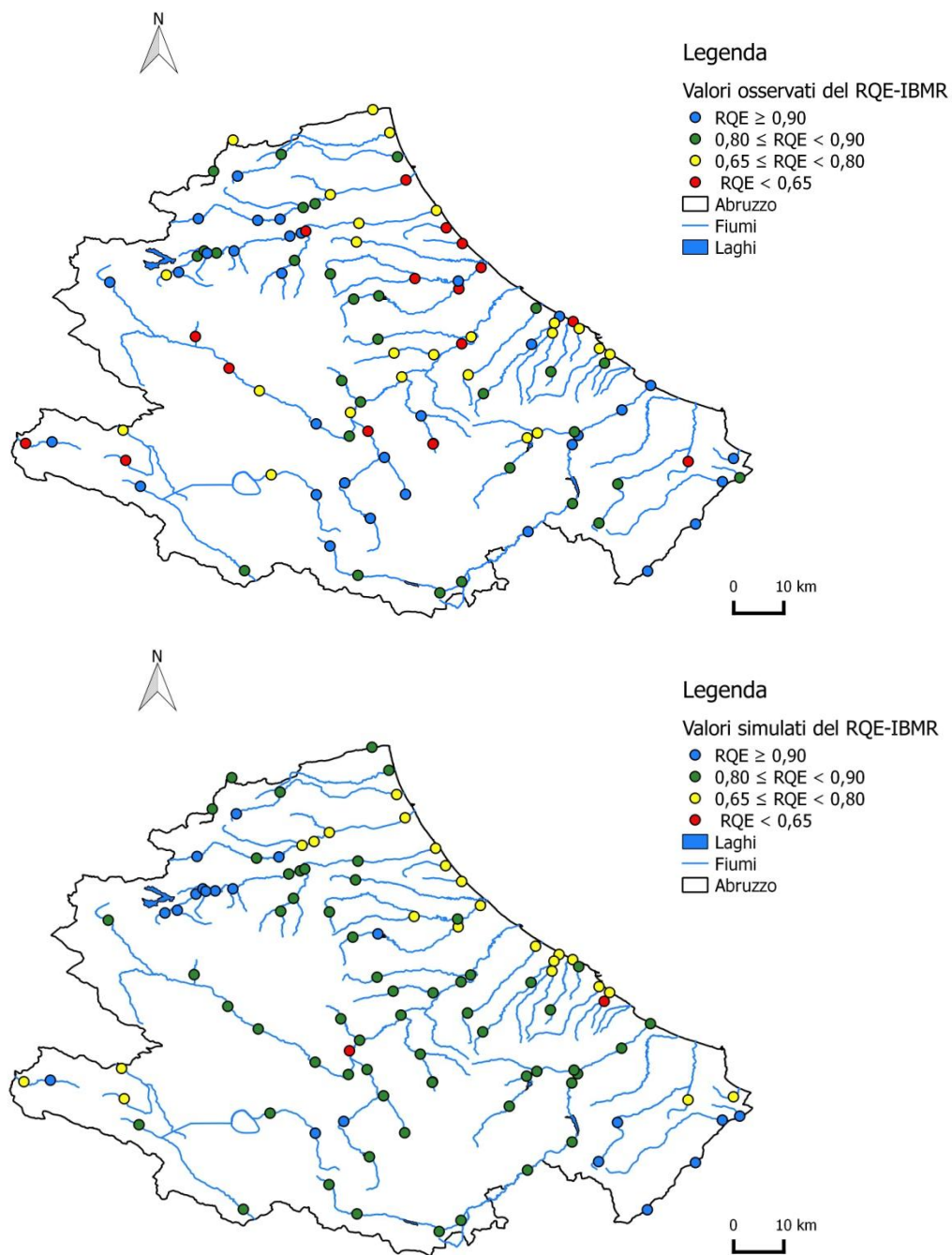


Figura 5.1. Confronto tra i valori osservati e simulati del RQE-IBMR, tenendo conto dei limiti tra le classi: Elevata ($\geq 0,90$), Buona ($0,80 \leq \text{RQE} < 0,90$), Sufficiente ($0,65 \leq \text{RQE} < 0,80$), Scarsa e Cattiva ($\text{RQE} < 0,65$). I valori simulati sono stati ottenuti implementando la seguente misura: incremento del 10% di OD e riduzione del 10% di N-NH4 ed N-NO3.

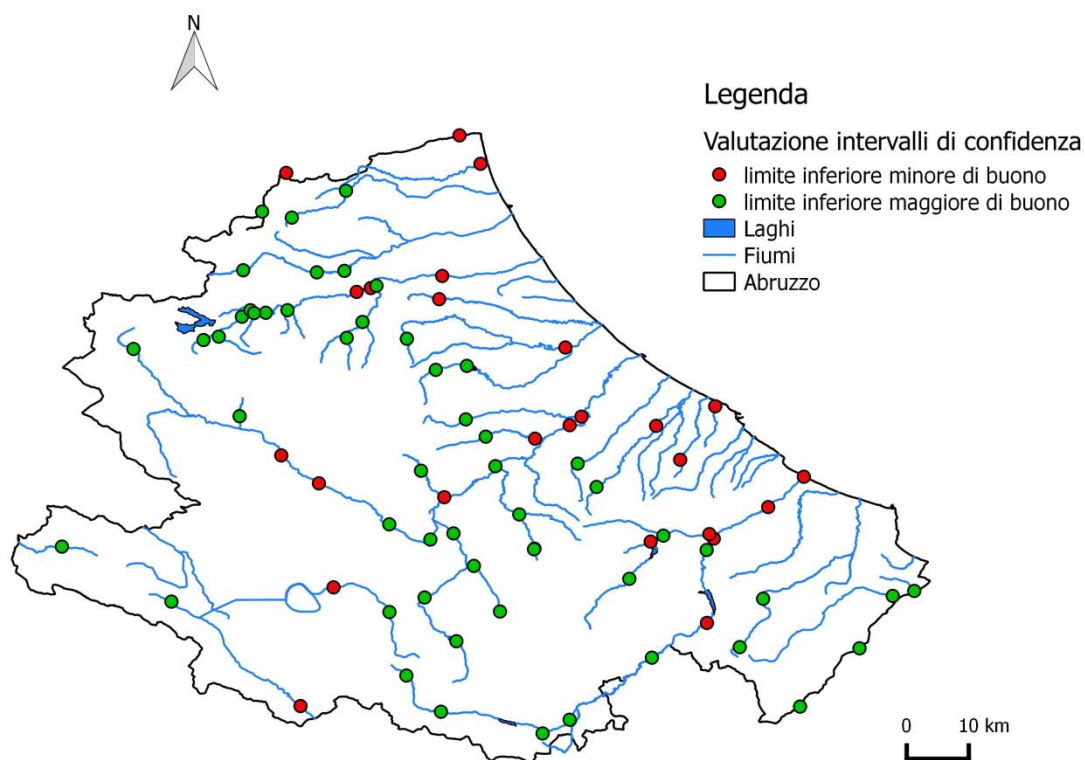


Figura 5.2. Analisi degli intervalli di confidenza dei valori simulati del RQE-IBMR ricadente nei limiti di classe buono o alto. Il colore rosso significa che il limite inferiore dell'intervallo di confidenza appartiene alla classe o sufficiente o scarsa o cattiva. Il colore verde significa che il limite inferiore dell'intervallo di confidenza appartiene alla classe buona o alta.

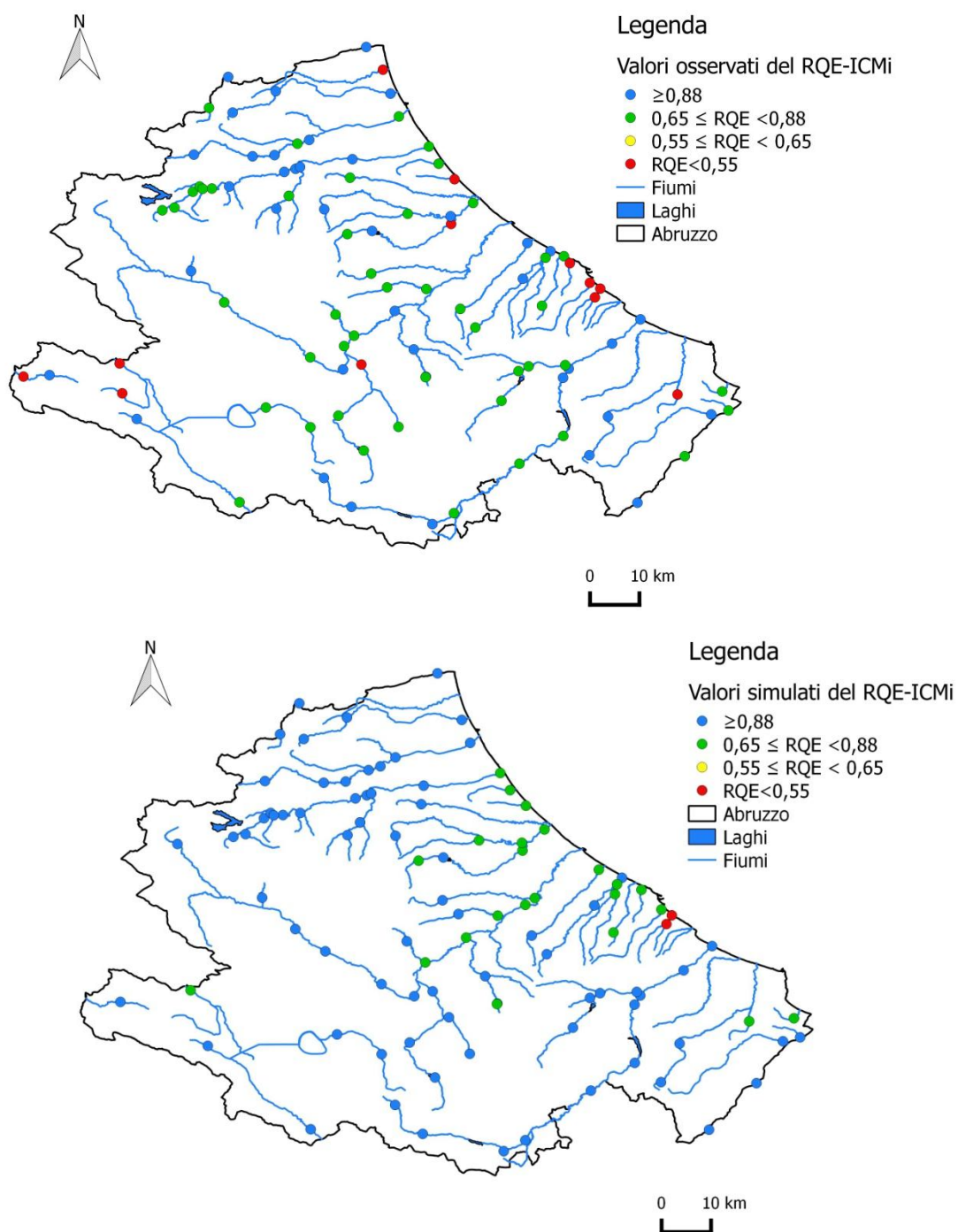


Figura 5.3. Confronto tra i valori osservati e simulati del RQE-ICMi, tenendo conto dei limiti tra le classi: Elevata ($\geq 0,88$), Buona ($0,65 \leq RQE < 0,88$), Sufficiente ($0,55 \leq RQE < 0,65$), Scarsa o Cattiva ($RQE < 0,55$). I valori simulati sono stati ottenuti implementando la seguente misura: incremento del 10% di OD e riduzione del 10% di N-NH4 ed N-NO3.

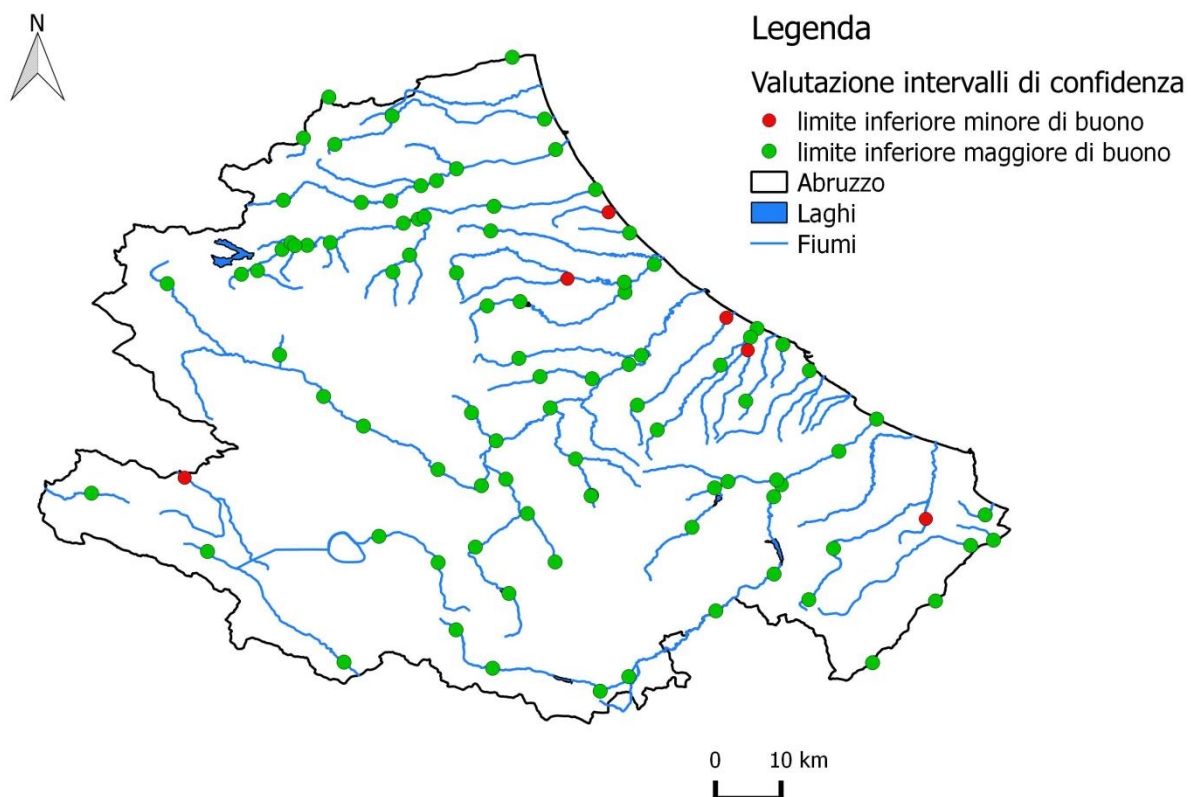


Figura 5.4. Analisi degli intervalli di confidenza dei valori simulati del RQE-ICMi ricadente nei limiti di classe buono o alto. Il colore rosso significa che il limite inferiore dell'intervallo di confidenza appartiene alla classe o sufficiente o scarsa o cattiva. Il colore verde significa che il limite inferiore dell'intervallo di confidenza appartiene alla classe buona o alta.

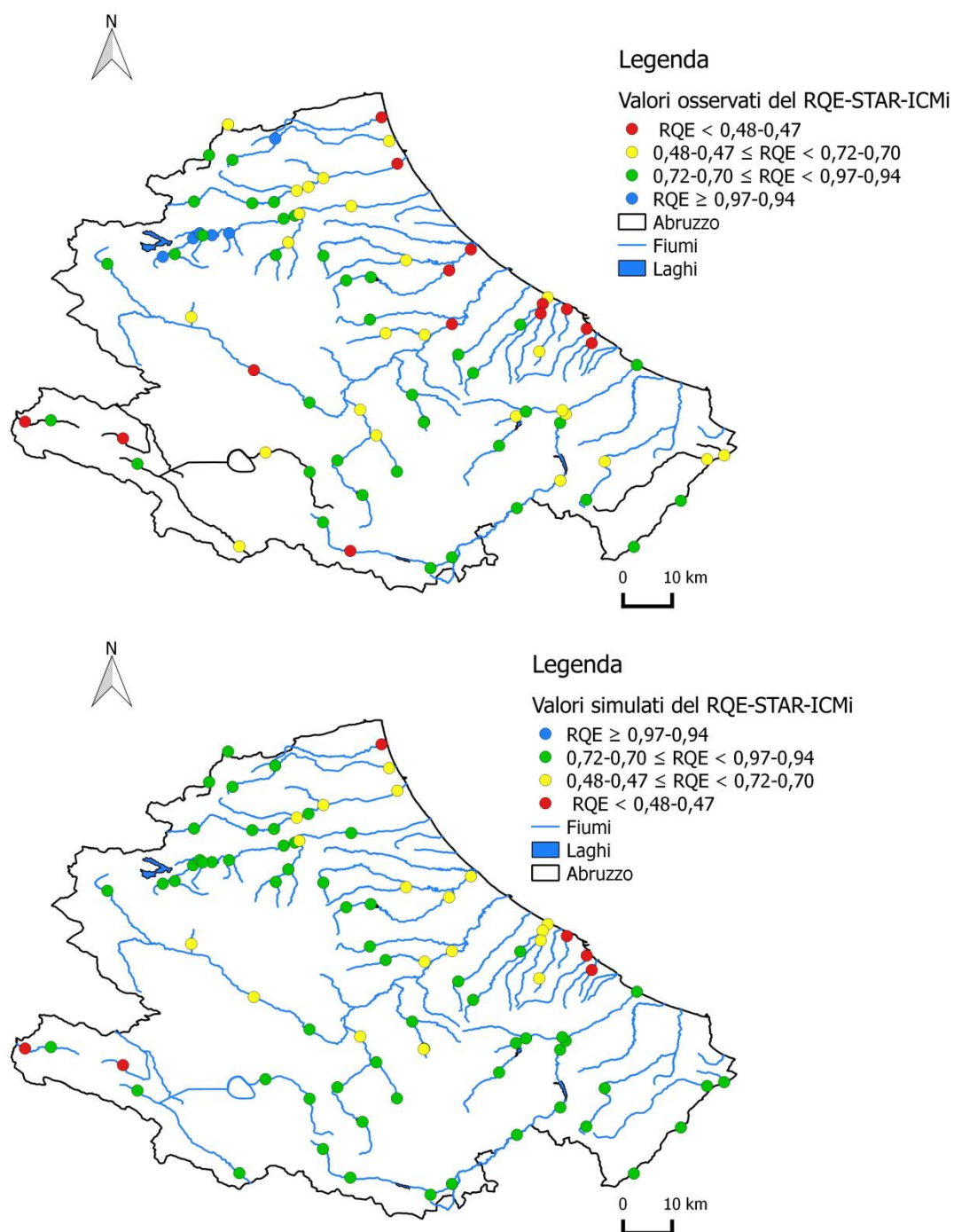


Figura 5.5. Confronto tra i valori osservati e simulati del RQE-STAR-ICMi, tenendo conto dei limiti tra le classi: Elevata ($\geq 0,97-0,94$), Buona ($0,72-0,70 \leq \text{RQE} < 0,97-0,94$), Sufficiente ($0,48-0,47 \leq \text{RQE} < 0,72-0,70$), Scarsa o Cattiva ($\text{RQE} < 0,48-0,47$). I valori simulati sono stati ottenuti implementando la seguente misura: incremento del 10% di OD e riduzione del 10% di N-NH4 ed N-NO3.

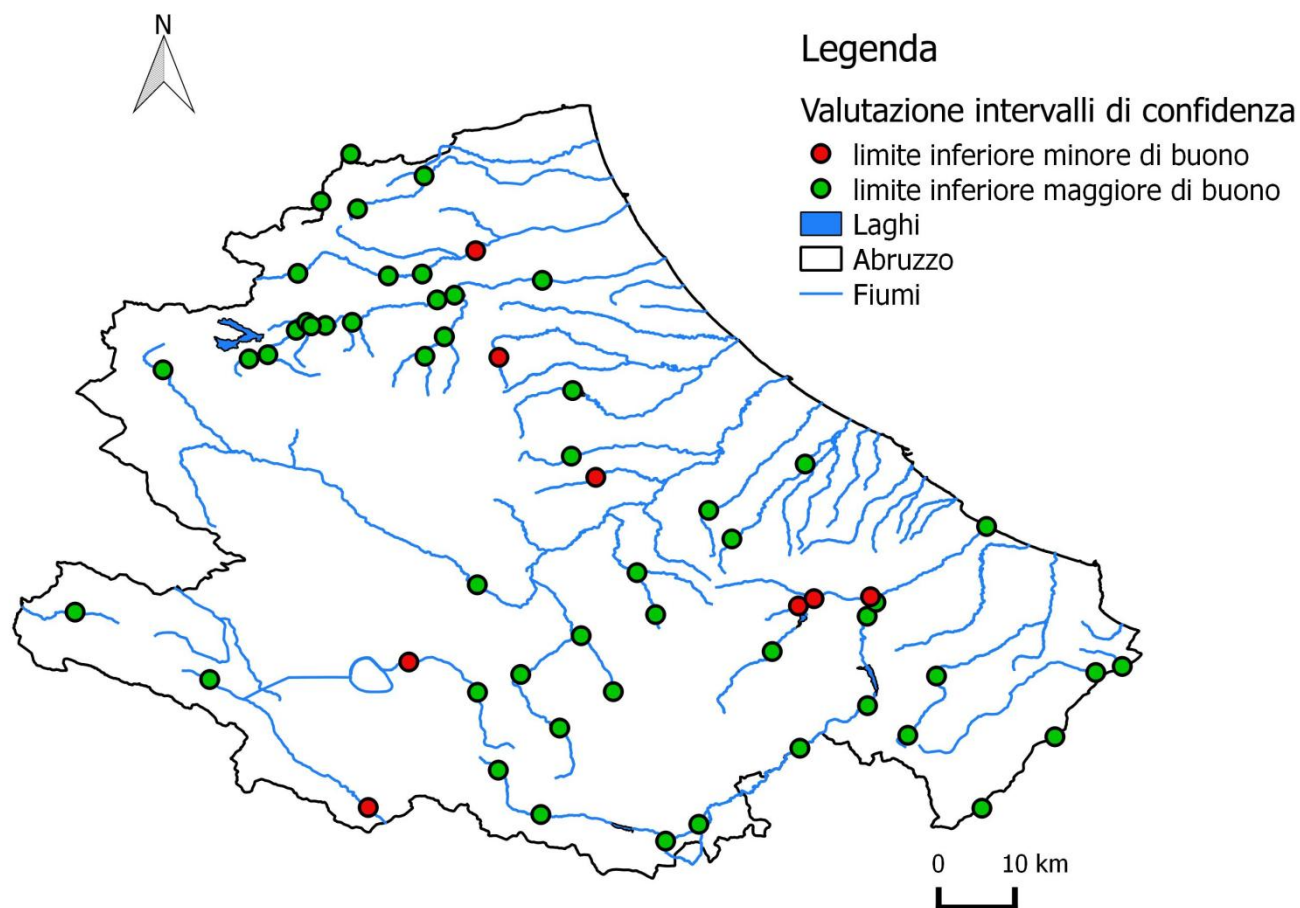


Figura 5.6. Analisi degli intervalli di confidenza dei valori simulati del RQE-STAR-ICMi ricadenti nei limiti di classe buono o alto. Il colore rosso significa che il limite inferiore dell'intervallo di confidenza appartiene alla classe o sufficiente o scarsa o cattiva. Il colore verde significa che il limite inferiore dell'intervallo di confidenza appartiene alla classe buona o alta.

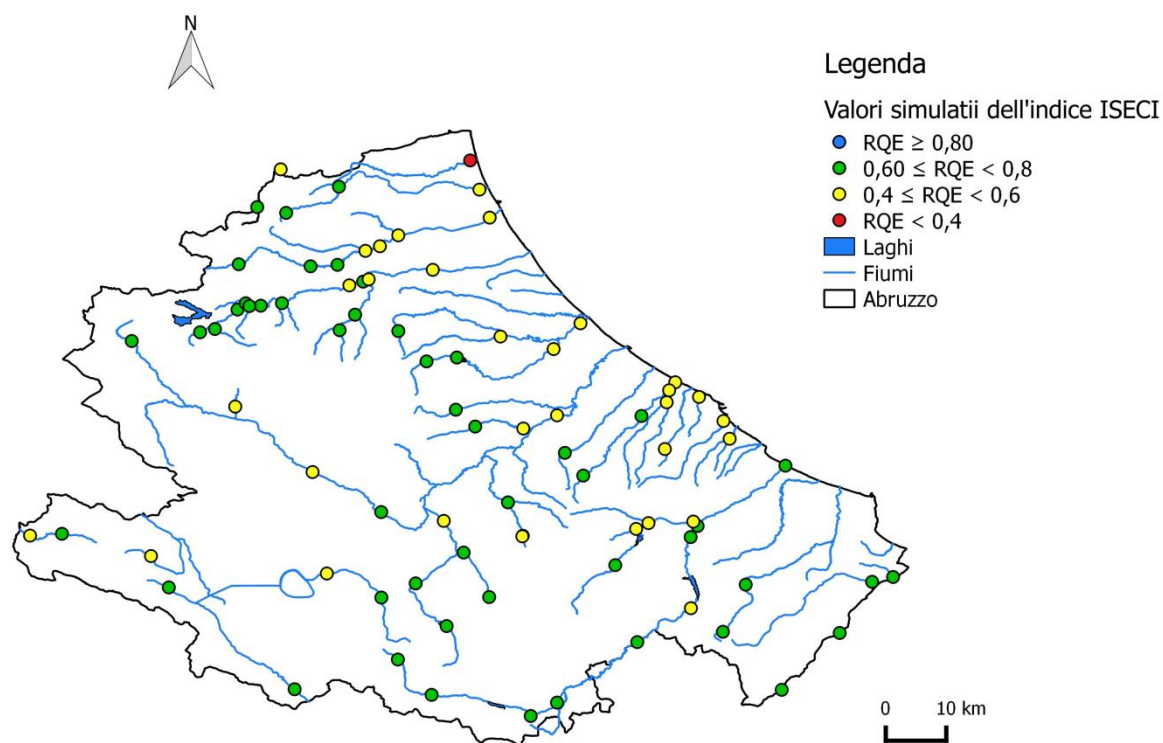
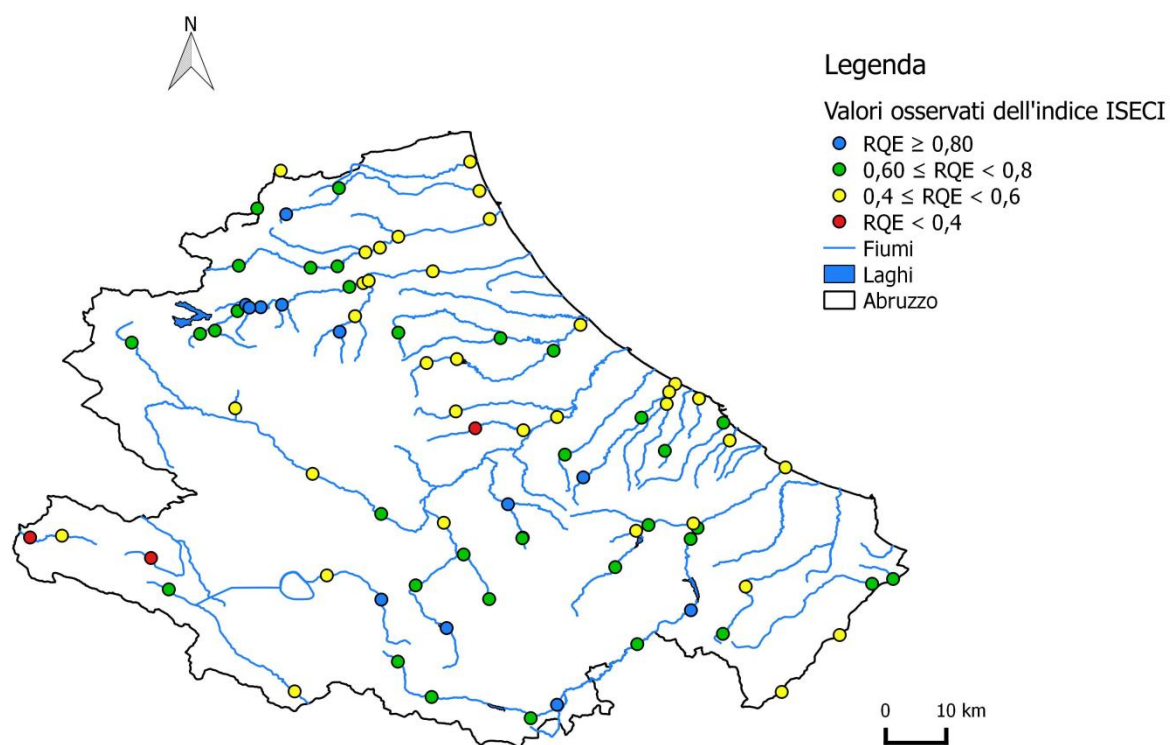


Figura 5.7. Confronto tra i valori osservati e simulati dell'ISECI, tenendo conto dei limiti tra le classi: Elevata ($\geq 0,80$), Buona ($0,60 \leq RQE < 0,80$), Sufficiente ($0,4 \leq RQE < 0,6$), Scarsa e Cattiva ($RQE < 0,4$). I valori simulati sono stati ottenuti implementando la seguente misura: incremento del 10% di OD e riduzione del 10% di N-NH₄ ed N-NO₃.

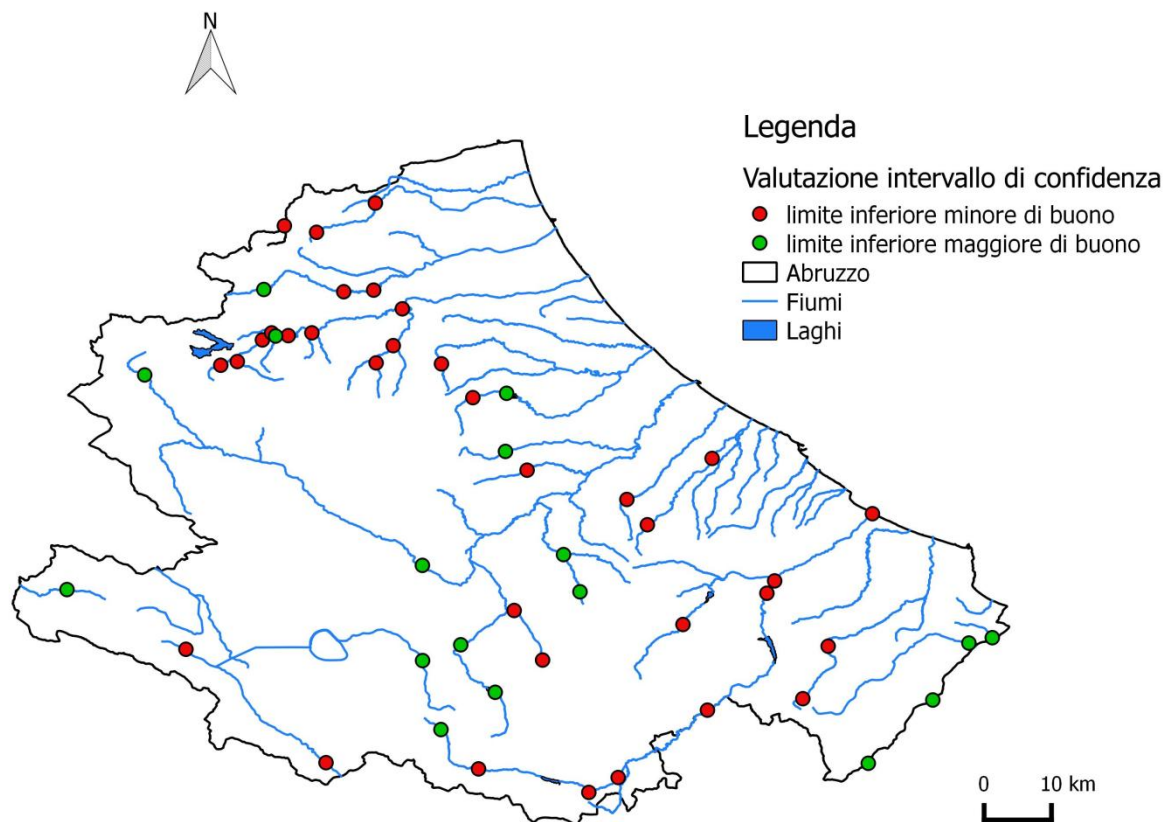


Figura 5.8. Analisi degli intervalli di confidenza dei valori simulati dell'ISECI ricadenti nei limiti di classe buono o alto. Il colore rosso significa che il limite inferiore dell'intervallo di confidenza appartiene alla classe o sufficiente o scarsa o cattiva. Il colore verde significa che il limite inferiore dell'intervallo di confidenza appartiene alla classe buona o alta.

6. Conclusioni

L'obiettivo principale del presente studio era di sviluppare e dimostrare l'uso di un modello di qualità ecologica come strumento per la valutazione di differenti scenari ottenuti dall'applicazione di diverse misure relative alla gestione della risorsa idrica in un distretto idrografico in Italia Centrale. Anche se dettagliati modelli di qualità meccanicistici (Ambrose et al., 1993; Chapra e Runkel, 1998; DHI, 2000; Tsakiris e Alexakis, 2012) sono stati utilizzati in molti casi di studio, la loro applicabilità è limitata nei bacini idrografici su larga scala. Ciò è legato alla mancanza dei dati richiesti e dalla difficoltà di comprendere e calcolare l'incertezza nei risultati modellistici, a causa della complessità intrinseca dei sistemi naturali (Reckhow, 1999).

Il principale vantaggio di utilizzare regressioni multiple, come nel modello proposto, denominato PARBLEU (Pressures/impacts/responses Assessment into River Basins Leading to Environmental Uses.), è la semplice quantificazione dell'incertezza. Tuttavia, l'analisi di regressione multipla può essere ostacolata dalla natura complessa dei dati ecologici, in cui le variabili considerate sono legate a molte variabili esplicative che spesso sono correlate tra loro. Sebbene la collinearità statistica può essere rimossa dai dati, le variabili esplicative restano per la loro natura ancora correlate, se o no funzionalmente (Graham, 2003).

Nel presente lavoro di studio, un totale di 121 stazioni di campionamento fluviali sono state analizzate e i RQE determinati, valutando la qualità ecologica da un punto di vista spaziale e l'incertezza associata alle simulazioni modellistiche relative a diversi scenari. Questa incertezza è espressa attraverso intervalli. Comunque, la necessità di metodi che utilizzino scale ad intervalli per valutare la qualità dell'acqua è stato già fortemente sottolineato da Wang et al. (2014).

I risultati mostrano che l'applicazione di un modello semplice come quello proposto permetterebbe di raggiungere nuove conoscenze non solo nella valutazione delle relazioni tra gli impatti e risposte ecologiche, ma anche tra le pressioni e gli impatti. I sistemi di regressione multipla possono ovviamente essere affetti anche da endogeneità. Le interazioni tra gli elementi biologici sono espresse come relazioni tra gli indici biologici, ma queste relazioni sono bidirezionali, e la causalità bidirezionale è una delle principali fonti di endogeneità (Kendall, 2015). Nonostante ciò, la comprensione delle interazioni umane-

ambientali come base per il processo decisionale della pianificazione territoriale, è una delle principali sfide che la ricerca della bio-complessità sta affrontando (Michener et al., 2001). Lo stimatore OLS è distorto e inconsistente, ma esso è stato utilizzato in questo studio come uno stimatore preliminare da confrontare con quelli derivanti da metodi statistici più avanzati negli sviluppi futuri. Il confronto dello stimatore OLS e di quelli ottenuti da metodi statistici più avanzati con i risultati futuri di monitoraggio potrà indicare i limiti attuali dell'uso del metodo OLS nella modellistica empirica di sistemi complessi.

In questo studio, i dati usati sono rappresentativi della popolazione di stazioni di monitoraggio della regione Abruzzo e non di quella relativa all'intero distretto dell'Appennino Centrale. Non è raccomandabile usare i parametri identificati in altre regioni del Distretto. In effetti, l'incertezza è comparabile alle metriche di valutazione dello stato ecologico ed essa resta un elemento strutturale tipico di ciascuna popolazione di stazioni di monitoraggio relativa a ciascuna regione appartenente al Distretto.

Comunque, come già evidenziato nelle discussioni, la scelta dei parametri di ingresso dovuta alla finalità con cui si sviluppa PARBLEU ed anche alla scarsità dei dati relativi ad altre variabili potrebbe condizionare la stima dei parametri del sotto-modello. In effetti, l'accuratezza dei sistemi di regressione multipla dipende se il modello include tutte le variabili rilevanti e le loro relazioni. Inoltre, la mancanza di un quadro conoscitivo, sia iniziale che a seguito delle modellazioni sviluppate, spazialmente distribuito evidenzia un forte limite dell'approccio utilizzato inerente al problema relativo all'eterogeneità del bacino.

Quando un numero maggiore di dati sarà disponibile nello spazio e nel tempo, alternative fattibili per un sistema di regressione multipla come quello descritto nel presente studio saranno rappresentate da tecniche di reti neurali artificiali (ANN) (Xie et al., 2006; Wu et al., 2010; Chau e Wu, 2010), come per esempio il sistema ANFIS (Adaptive-Network-based Fuzzy Inference System) (Jang, 1993).

Negli sviluppi futuri, queste tecniche potrebbero essere attuate in relazione al livello di specializzazione raggiunto dalle Autorità e dalle strutture tecniche regionali. Quando un database più completo e strutturato delle pressioni sarà disponibile, il legame tra impatti e pressioni sarà oggetto di ricerca.

Bibliografia

- Allen, Michael P. 1998. Understanding Regression Analysis. Plenum Publishing, New York.
- Ambrose, R.B., Wool, T.A., Martin, J.L. 1993. The water quality analysis simulation program WASP5 Version 5.10 Part A: Model Documentation. Environmental Research Laboratory, Office of Research and Development, US EPA, Athens, Georgia, USA.
- Berry, W. D., & Feldman, S., 1985. Multiple Regression in Practice. Sage, London.
- Beven, K. 2002. Towards a coherent philosophy for modelling the environment. Proceedings of the Royal Society of London, A 458 (2026), 2465-2484.
- Beven, K., Binley, A.M. 1992. The future of distributed models, model calibration and uncertainty predictions. Hydrological Processes 6, 279-298.
- Breusch, T. S. and Pagan, A. R., 1979, A simple test for heteroskedasticity and random coefficient variation. Econometrica, 47(5), 1288-1290.
- Box G.E.P. and Cox D.R., 1964. An analysis of transformations. J Roy Stat Soc, Ser B 26:211–252.
- Camargo, J.A., Alonso, Á. 2006. Ecological and toxicological effects of inorganic nitrogen pollution in aquatic ecosystems: a global assessment. Environ. Int. 32, 831–849.
- Carpenter S.R., Lodge D.M. 1986. Effect of submersed macrophytes on ecosystem processes. Aquat. Bot. 26: 341–370.
- Chapra S.C., Runkel R.L. 1998. Modelling impact of storage zones on stream dissolved oxygen. ASCE Journal of Environmental Engineering, 125 (5), 415-419.
- Chau K.W., Wu C.L. 2010. A hybrid model coupled with singular spectrum analysis for daily rainfall prediction. Journal of Hydroinformatics 12 (4): 458-473
- Chatfield, C. and Collins, A. J. 1980. Introduction to Multivariate Analysis. Chapman and Hall, London.
- Correll, D.L. 1998. The role of phosphorus in the eutrophication of receiving waters: A review. Journal of Environmental Quality 27:261-266.
- Covich, A.P, Palmer, M.A, Crowl, T.A. 1999. The role of benthic invertebrate species in freshwater ecosystems. BioScience 49:119-28.
- Cox, B.A, 2003. A review of currently available in-stream water-quality models and their applicability for simulating dissolved oxygen in lowland rivers. The Science of the Total Environment 314 –316 335–377.
- COX, B. A.; WHITEHEAD, P. G., 2005. Parameter sensitivity and predictive uncertainty in a new water quality model, Q2. Journal of Environmental Engineering 131(1): 147-157

Decreto Legislativo 31 marzo 1998, n. 112. Conferimento di funzioni e compiti amministrativi dello Stato alle regioni ed agli enti locali, in attuazione del capo I della legge 15 marzo 1997, n. 59"

Decreto Legislativo 22 gennaio 2004, n. 42 (1). Codice dei beni culturali e del paesaggio, ai sensi dell'art. 10 della Legge 06/07/2002, n. 137.

Decreto n. 260 del 8 novembre 2010, Regolamento recante i criteri tecnici per la classificazione dello stato dei corpi idrici superficiali e per la modifica delle norme tecniche del DLgs 152/06 etc.

DHI. 2000. MIKE 11: A modelling system for rivers and channels: User guide. DHI Software.

Dybas, C.L. 2001. From biodiversity to biocomplexity: a multidisciplinary step toward understanding our environment. *BioScience*. 51, 426–430.

EuropeanEnvironmentAgency. (1999). "Environmental Indicators: Typology and Overview. European Environment Agency, Technical Report No. 25," Copenhagen, 19 pp.

European Union, 2000. Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy. Brussels. (72 pp.)

Everitt, B.S. & Hothorn, T., 2011. An introduction to applied multivariate analysis with R. Springer, New York.

Fernandes, U.L., de Oliveira, E.C.C., Lacerda, S.R. 2016. Role of macrophyte life forms in driving periphytic microalgal assemblages in a Brazilian reservoir. *Journal of Limnology* 75.1, 44-51

Freund, R. J., Wilson, W. J., & Sa, P., 2006. Regression analysis: statistical modeling of a response variable (2nd ed.). Elsevier Academic Press, Burlington, MA.

Fox J., 1991, Regression Diagnostics: An Introduction, (Sage University Paper series on Quantitative Applications in the Social Sciences, series no. 07-79), Sage, Newbury Park, CA.

Graham, M.H. 2003. Confronting Multicollinearity in Ecological Multiple Regression. *Ecology*, 84, 2809-2815.

Gunst, R.F. and Mason R.L., 1980, Regression Analysis and Its Application: A Data- Oriented Approach, Marcel Dekker, Inc., New York.

Hayes, A.F. and Cai, L., 2007 . Using heteroskedasticity-consistent standard error estimators in OLS regression: An introduction and software implementation, *Behavior Research Methods*, 39 – 4, 709-722.

Halasz, G., Szlepak, E., Szilagyi, E., Zagyva, A. e Fekete, I., 2007, Application of EU Water Framework Directive for monitoring of small water catchment areas in Hungary, II. Preliminary study for establishment of surveillance monitoring system for moderately loaded (rural) and heavily loaded (urban) catchment areas, *Microchemical Journal*, vol. 85, 1, 72-79.

Haury, J., Peltre, M.C., Tremolieres, M., Barbe, J., Thiebaut, G., Bernez, I., Daniel, H., Chatenet, P., Haan-Archipof, G., Muller, S., Dutartre, A., Laplace-Tretyure, C., Cazaubon, A., Lambert-Servien, E. 2006. A new method to assess water trophy and organic pollution – The Macrophyte biological index for rivers (IBMR): Its application to different types of river and pollution. *Hydrobiologia*, 570, 153–158.

Holtan, H, Kamp-Nielsen, L, Stuanes, A. 1988. Phosphorus in soil, water and sediment: An overview. *Hydrobiologia*. 170: 19–34. Jakeman, A.J., Letcher, R.A., 2003. Integrated assessment and modelling: features, principles and examples for catchment management. *Environmental Modelling & Software* 18, 491-501.

Jang, J.S.R. 1993. ANFIS: Adaptive-network-based fuzzy inference systems. *IEEE Trans. Syst. Man Cybern.*, 23, 665–685.

Kaufman, Robert L. 2013. *Heteroskedasticity in Regression: Detection and Correction*. Sage, Thousand Oaks, CA.

Kendall, B.E. 2015. A statistical symphony: Instrumental variables reveal causality and control measurement error. Pp. 149-167 in G.A. Fox, S. Negrete-Yankelevich, and V.J. Sosa, eds. *Ecological Statistics: Contemporary Theory and Application* (Oxford University Press, Oxford).

Keupers, I., Willems, P. 2015. Global sensitivity analysis of transformation processes in a river water quality model by means of conceptualization. 36th IAHR World Congress, 28 June -3 July 2015, The Hague.

Killgore, K.J., Hoover, J.J. 2001. Effects of hypoxia on fish assemblages in a vegetated water body. *Journal of Aquatic Plant Management* 39:40–44.

Koenker, R., 1981. A note on studentizing a test for heteroskedasticity. *Journal of Econometrics*, 17(1), 107-112.

Kuan C. M., 2004. *Introduction To Econometric Theory*. Taipei: Institute of Economics, Academia Sinica, Disponibile all'indirizzo: http://homepage.ntu.edu.tw/~ckuan/pdf/et01/et_Ch4.pdf

Levine D. M.; Krehbiel T. C.; Berenson M. L., 2010. *Statistica*. Pearson, Milano.

Liu, T.C., 1960. Underidentification, structural Estimation and Forecasting. *Econometrica*, 28, 855-865.

Loucks, D. P., J. R. Stedinger, and D. Haith, 1981. *Water Resources Systems Planning and Management*. Prentice-Hall, Englewood Cliffs, N.J..

Michener, W.K., Baerwald, T.J., Firth, P., Palmer, M.A., Rosenberger, J.L., Sandlin, E.A., Zimmerman, H. 2001. Defining and Unraveling Biocomplexity: *Bioscience*. 51,12, 1018-1023.

Miltner, R.J. 2010. A method and rationale for deriving nutrient criteria for small rivers and streams in Ohio. *Environmental Management*, 45, 842–845.

Montgomery, D.C. and Peck E. A., 1982. *Introduction to linear regression analysis*. Wiley.

Jaccard, J., Robert T., & Choi K. W., 1990. Interaction effects in multiple regression. Sage Publications, Newbury Park, Calif.

Joseph Ofungwu, 2011, Statistical Applications for Environmental Analysis and Risk Assessment, Wiley.

Nannicini Tommaso , 2010, Università Bocconi, Appunti di Econometria. Disponibile all'indirizzo http://tommasonannicini.eu/media/courses/files/appunti_econometria_2.pdf.

Pahl-Wostl, C., 2002. Towards sustainability in the water sector-the importance of human actors and processes of social learning. *Aquatic Sciences* 64, 394-411.

Pahl-Wostl, C., 2007. The implications of complexity for integrated resources management. *Environmental Modelling & Software* 22, 561-569.

Rawlings, J. O., Pantula, S. G. and Dickey, D. A., 1998. Applied regression analysis: a research tool (second edition). Springer-Verlag, New-York.

Reckhow, K.H., 1999. Water quality predictions and probability network models. *Canadian Journal of Fisheries and Aquatic Sciences*, 56, 1150–1158.

Refsgaard, J.C., Henriksen, H.J., 2004. Modelling guidelines- terminology and guiding principles. *Advances in Water Resources* 27, 71-82.

Refsgaard, J. C., J. P. Van der Sluijs, J. Brown, and P. Van der Keur, 2006. A framework for dealing

with uncertainty due to model structure error. *Adv. Water Resour.*, 29, 1586–1597.

Ricci, V. (2006). Principali tecniche di regressione con R. Lecture notes, vito_ricco@yahoo.it.

Rode, M., Arhonditsis, G., Balin, D., Kebede, T., Krysanova, V., Van Griensven, A., Van der Zee, S.E.A.T.M. 2010. New challenges in integrated water quality modelling. *Hydrological Processes* 24, 3447–3461.

Rott, E. 1991. Methodological aspects and perspectives in the use of periphyton for monitoring and protecting rivers. In B.A. Whitton, E. Rott, and G. Friedrich (editors). *Use of algae for monitoring rivers*. Institut fur Botanik, University of Innsbruck, Austria.

Schneider, S.C., Lawniczak, A.E., Picińska-Faltynowicz, J., Szoszkiewicz, K. 2012. Do macrophytes, diatoms and non-diatom benthic algae give redundant information? Results from a case study in Poland. *Limnologica*, 42, 204–211.

Smith Eric, 2004. Uncertainty analysis, *Encyclopedia of Environmetrics* (ISBN 0471 899976), 4, 2283–2297 .

Sosa-Escudero W., 2009, Universidad de San Andrés, Buenos Aires. Econometric Analysis: Heteroskedasticity and Weighted Least Squares. Disponibile all'indirizzo: http://www.econ.uiuc.edu/~wsosa/econ471/GLS_Heteroskedasticity.pdf.

Stevenson, R.J., Pan Y., Van Dam, H. 2010. Assessing environmental conditions in rivers and streams with diatoms. In Smol JP and Stoermer EF (eds), *The Diatoms: Applications for the*

Environmental and Earth Sciences. 2nd edition, (Cambridge University Press, Cambridge, England) 57-85.

Taylor, J.C., Janse van Vuuren, M.S., Pieterse, A.J.H. 2007. The application and testing of diatom-based indices in the Vaal and Wilge rivers, South Africa. *Water SA*, 33, 51–60. Tao, H, 2008. Calibration, sensitivity and uncertainty analysis in surface water quality modeling (Doctoral Dissertation). Medford: Tufts University. USEPA (United States Environmental Protection Agency), 2009.

Thiebaut, G. 2008. Phosphorus and aquatic plants. *Plant. Ecophysiol.* 7:31-49.

Thomaz, S.M, Cunha, E.R. 2010. The role of macrophytes in habitat structuring in aquatic ecosystems: methods of measurement, causes and consequences on animal assemblages' composition and biodiversity. *Acta Limnol Bras* 22: 218–236.

Tsakiris, G., Alexakis, D. 2012. Water quality models: An overview. *European Water*, 37, 33-46.

Uusitalo, L., Lehtikoinen, A., Helle, I., Myrberg, K., 2015. An overview of methods to evaluate uncertainty of deterministic models in decision support. *Environ. Model. Softw.* 63 (0), 24-31.

Vandenberghe, V., Bauwens, W., Vanrolleghem, P.A., 2007. Evaluation of uncertainty propagation into river water quality predictions to guide future monitoring campaigns. *Environmental Modelling & Software* 22, 725-732.

Vittinghoff E., David V. Glidden, Stephen C. Shiboski y Charles E. McCulloch., 2005. *Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models. Statistics for Biology and Health.* Springer Science+Business Media, Inc., New York, NY, USA.

Wallace, J.B., Webster, J.R. 1996. The role of macroinvertebrates in stream ecosystem function. *Annual Review of Entomology*, 41, 115-139.

Wang W.C., Xu D.M., Chau K.W., Lei G.J. 2014. Assessment of river water quality based on theory of variable fuzzy sets and fuzzy binary comparison method. *Water Resour. Manag.*, 28, 4183–4200.

Wang, L., Robertson, D.M., Garrison, P.J. 2007. Linkages between nutrients and assemblages of macroinvertebrates and fish in Wadeable streams: implications for nutrient criteria development. *Environmental Management*, 39, 194–212.

Wellen, C., Kamran-Disfani, A.R., Arhonditsis, G.B. 2015. Evaluation of the current state of distributed watershed nutrient water quality modeling. *Environ. Sci. Technol.* 49, 3278–3290.

Wooldridge J. M., 2006. *Introductory econometrics: A modern approach.* Mason, OH, Thomson/South-Western.

Wu C.L. , Chau K.W., Fan C., 2010, Prediction of Rainfall Time Series Using Modular Artificial Neural Networks Coupled with Data Preprocessing Techniques, *Journal of Hydrology*, 389, (1-2), 146-167.

Xie J. X., Cheng C. T., Chau K. W. et al., 2006. A hybrid adaptive time-delay neural network model for multi-step-ahead prediction of sunspot activity, *International Journal of Environment and Pollution*, 28 (3), 364-381.

Yan, X., and Su, X., 2009. *Linear regression analysis: Theory and computing*. World Scientific Publishing Company Inc., Singapore.

Young, P.C. 1983. The validity and credibility of models for badly defined systems, in *Uncertainty and Forecasting of Water Quality*, eds: M.B. Beck and G. Van Straten. Berlin, Springer-Verlag, 69–100.