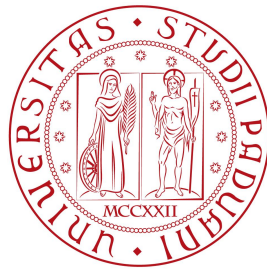


Università degli Studi di Padova  
Dipartimento di Scienze Statistiche  
Corso di Laurea Triennale in

Statistica per le tecnologie e le scienze



## **Robustezza e distorsione delle stime italiane dell'eccesso di mortalità nel 2020 e nel 2021**

Relatore: prof. Stefano Mazzuco  
Dipartimento di Scienze Statistiche

Laureando: Andrea Valentina Pertile  
Matricola n. 1217193

Anno Accademico 2021/2022

# Indice

Elenco delle tabelle	1
<b>1 Introduzione</b>	<b>2</b>
<b>2 Modelli e metodi</b>	<b>4</b>
2.1 Eccesso di mortalità . . . . .	4
2.2 Modelli . . . . .	5
2.2.1 5 year average death model . . . . .	6
2.2.2 Serfling model . . . . .	6
2.3 Convalida incrociata . . . . .	7
<b>3 Dati</b>	<b>9</b>
<b>4 Risultati</b>	<b>11</b>
<b>5 Conclusione</b>	<b>24</b>
<b>A Tabelle</b>	<b>26</b>
<b>B Grafici</b>	<b>31</b>
<b>Bibliografia</b>	<b>80</b>

# Elenco delle tabelle

A.1	Strati per cui il quantile 0.05 della distribuzione del P-score cumulativo all'ultima settimana del 2021 risulta maggiore di zero per tutti i modelli. . . . .	26
A.2	Valore massimo del P-score per il 2020, modello da cui sono stati ottenuti i valori dei decessi attesi e data del lunedì della settimana corrispondente. . . . .	27
A.3	Valore massimo dell'eccesso di mortalità per il 2020, modello da cui sono stati ottenuti i valori dei decessi attesi e data del lunedì della settimana corrispondente. . . . .	28
A.4	Valore massimo del P-score per il 2021, modello da cui sono stati ottenuti i valori dei decessi attesi e data del lunedì della settimana corrispondente. . . . .	29
A.5	Valore massimo dell'eccesso di mortalità per il 2021, modello da cui sono stati ottenuti i valori dei decessi attesi e data del lunedì della settimana corrispondente. . . . .	30

# Capitolo 1

## Introduzione

In Italia il primo decesso per Covid-19 è stato registrato a febbraio del 2020 in Veneto. A partire da quel momento in Italia e nel resto del mondo si sono susseguiti periodi di aumento e diminuzione dei contagi e, nel biennio 2020-2021, le regioni italiane sono state colpite in modo diverso e non in modo sincrono da questo aumento dei contagi.

L'intento è di verificare come varia la stima dell'eccesso di mortalità nelle diverse regioni italiane usando diversi modelli per prevedere i decessi. L'eccesso di mortalità viene definito come i decessi osservati meno i decessi attesi, quindi si possono ottenere valori negativi nel caso in cui i decessi osservati siano minori di quelli attesi. Considerando i decessi per tutte le cause di morte, si ottengono informazioni anche sui casi non diagnosticati e sull'impatto indiretto della pandemia [2, 8].

Nonostante l'uso diffuso dei modelli basati sulle medie degli anni precedenti, grazie alla loro semplicità, questi modelli hanno lo svantaggio di non tener conto del trend della mortalità. Un aggiustamento in questa direzione è l'uso di modelli di tipo Serfling che tengono conto del trend della mortalità e dell'effetto stagionale che la mortalità può avere.

Per tutti i modelli scelti si testa la tendenza a produrre stime maggiori o minori del previsto a livello settimanale e annuale e si osserva se e come la classifica delle regioni italiane fatta in base al valore dell'eccesso di mortalità percentuale stimato per il 2020-2021 varia usando modelli diversi. Si valuta anche se esistono regioni dove i modelli non sono concordi sull'esistenza significativa di un eccesso di mortalità.

Per verificare quale modello sia il più affidabile, viene usata la tecnica della convalida incrociata, ogni modello viene adattato usando i dati dei decessi settimanali dei cinque anni precedenti per poi prevedere i decessi delle 52 settimane successive. L'errore di ogni modello può essere stimato confrontando la previsione del modello con i dati osservati, verificando

così se un modello sottostima o sovrastima i decessi osservati.

## Capitolo 2

# Modelli e metodi

Per stimare l'eccesso di mortalità e le relative statistiche si selezionano cinque modelli di due tipologie diverse, il *5 year average death model* sia tenendo conto della struttura della popolazione che non e il *Serfling model* sia tenendo conto della struttura per età che non e adattando il modello solo su alcune settimane. Tutti i modelli prevedono i decessi settimanali suddivisi per sesso ed età e i parametri del modello vengono stimati separatamente per ognuna delle 20 regioni italiane. Si considera il totale dei decessi settimanali anziché considerare i decessi giornalieri per smorzare le fluttuazioni casuali [11] e ridurre il numero di zeri nelle settimane rispetto alle diverse combinazioni delle variabili di stratificazione.

### 2.1 Eccesso di mortalità

Per eccesso di mortalità si intende quanto i decessi di un determinato periodo superano i decessi attesi per lo stesso periodo. Un primo metodo per calcolare questo valore è di sottrarre ai decessi della settimana  $i$  i decessi registrati nella stessa settimana dell'anno precedente; il problema di questa prima soluzione è la possibilità che anche l'anno precedente sia un anno anomalo. Un altro metodo è di stimare l'eccesso di mortalità come se fosse il residuo di un modello, quindi l'eccesso di mortalità  $D^*$  per una data settimana  $t$  per lo strato  $j$  viene definito come  $D_{tj}^* = D_{tj} - \lambda_{tj}$  dove  $D$  sono i decessi osservati e  $\lambda$  sono i decessi attesi sotto uno specifico modello. L'eccesso di mortalità per un periodo  $[a, b]$  è dato da  $D_{[a,b],j}^* = D_{[a,b],j} - \lambda_{[a,b],j}$  dove  $D_{[a,b],j} = \sum_{t=a}^b D_{tj}$  e  $\lambda_{[a,b],j} = \sum_{t=a}^b \lambda_{tj}$ . Per stimare l'eccesso di mortalità ad un livello di aggregazione maggiore si possono sommare le morti attese e osservate

nei singoli strati[9].

Per poter confrontare la stima dell'eccesso di mortalità tra diverse regioni si usa la percentuale settimanale di incremento delle morti osservate sulle morti attese, questa quantità viene chiamata *P-score* e si calcola come  $P_{tj} = \frac{D_{tj}^*}{\lambda_{tj}}$ . Come per l'eccesso di mortalità si può calcolare il P-score

cumulativo per una sequenza di settimane  $[a, b]$  come  $P_{[a,b],j} = \frac{D_{[a,b],j}^*}{\lambda_{[a,b],j}}$ .

Gli intervalli di previsione relativi alle misure di eccesso di mortalità sono calcolati tramite simulazione. Vengono prima campionati 500 gruppi di stime dei parametri,  $s = 1, \dots, 500$ , da una normale multivariata con distribuzione  $N(\beta, \Sigma)$ , con  $\beta$  vettore dei parametri stimati dal modello e  $\Sigma$  matrice di varianza e covarianza del modello stimato. Per ogni gruppo vengono calcolati i decessi attesi settimanali  $\lambda_{tj}^s$ . Successivamente viene campionato un singolo valore simulato dei decessi attesi  $d_{tj}^s$  da una distribuzione binomiale negativa con media pari a  $\lambda_{tj}^s$  e varianza  $\lambda_{tj}^s \phi$ , corrispondente al momento caratteristico della distribuzione quasi-Poisson. In caso di sotto dispersione si campiona da una Poisson con tasso  $\lambda_{tj}^s$ . Le morti attese simulate  $d_{tj}^s$  vengono poi sostituite nelle formule dell'eccesso di mortalità al posto delle morti attese  $\lambda_{tj}$  e le distribuzioni risultanti sono riassunte dai loro quantili[9].

Considerando i decessi per tutte le cause di morte, per stimare l'eccesso di mortalità nel biennio 2020-2021, si elimina il problema legato alle diverse registrazioni delle morti per Covid-19 e inoltre si tiene conto dell'impatto indiretto della pandemia, come i decessi per altre cause le quali possono essere correlate ad un ritardo nell'accesso alle cure[2].

## 2.2 Modelli

Tutti i modelli scelti sono specificati come una regressione quasi-Poisson sui decessi settimanali  $D_i$  della forma

$$D_i \sim Pois(\lambda_i, \phi)$$

$$\lambda_i = exp(\eta_i)$$

dove  $\eta_i$  sono i decessi attesi,  $\phi$  è il fattore di dispersione e  $\eta_i$  è il predittore lineare su scala logaritmica. Quello che cambia nei modelli scelti è il predittore lineare.

### 2.2.1 5 year average death model

Il modello *5 year average* usa come baseline i decessi settimanali dei cinque anni precedenti ma non tiene conto del trend generale che la mortalità può avere.

Il predittore lineare per il modello *5 year average death count* (AVGc5) è il seguente:

$$\eta_i = \alpha_{j[i]w[i]} \quad (2.1)$$

dove  $\alpha_{j[i]w[i]}$  è il coefficiente per la settimana  $w$  e lo strato sesso-età  $j$  per l'osservazione  $i$ -esima di una specifica regione. Il modello è stimato usando i dati relativi a un periodo di cinque anni.

Il logaritmo delle settimane-persona di esposizione  $\log(E_i)$  viene aggiunto al predittore lineare 2.1 come un offset, cambiando il modello nel *5 year average death rate* (AVGr5)

$$\eta_i = \alpha_{j[i]w[i]} + \log(E_i). \quad (2.2)$$

Quest'ultimo modello tiene conto della struttura della popolazione nello stimare i decessi settimanali.

### 2.2.2 Serfling model

Un modello che tiene conto della stagionalità è il modello Serfling il quale prevede un termine lineare, per descrivere il trend, e dei termini di una serie di Fourier che introducono un andamento armonico, per tenere conto della stagionalità [10].

Il modello *Serfling con esposizione* (SRFr) è specificato come

$$\eta_i = \alpha_{j[i]} + \beta_{j[i]}t_i + \gamma_{1,j[i]} \sin\left(\frac{2\pi}{52}w_i\right) + \gamma_{2,j[i]} \cos\left(\frac{2\pi}{52}w_i\right) + \gamma_{3,j[i]} \sin\left(\frac{2\pi}{26}w_i\right) + \gamma_{4,j[i]} \cos\left(\frac{2\pi}{26}w_i\right) + \delta_{j[i]k[i]} + \log(E_i). \quad (2.3)$$

Il modello 2.3 è caratterizzato dal trend lineare del tempo  $\beta_{j[i]}$  per la settimana  $t_i$  dalla prima osservazione  $i$ , dall'effetto stagionale parametrico  $\gamma_{j[i]}$  e dall'effetto delle festività  $\delta_{j[i]k[i]}$ , dove  $k$  indica il tipo di festività, sui decessi attesi. Il modello è stratificato per i gruppi sesso-età  $j$ . Per ogni strato viene selezionato il numero di termini stagionali (nessun termine stagionale,  $\gamma_{1,2}$  o  $\gamma_{1,2,3,4}$ ) attraverso il confronto dell'AIC usando la verosimiglianza da un'iniziale regressione Poisson senza sovra

dispersione.

Il modello *Serfling senza esposizione* (SRFc) si ottiene eliminando dal predittore lineare 2.3 il termine  $\log(E_i)$  relativo alle settimane-persona di esposizione.

Per avere il modello *Serfling stile Euromomo* (SRFcem) si adatta il modello non considerando le settimane relative all'influenza stagionale, quindi usando i dati dalla 15esima alla 26esima settimana e dalla 36esima alla 45settimana di ogni anno. Il predittore lineare è lo stesso del modello Serfling senza esposizione.

**Nota:** In un primo momento era stato considerato anche il modello generalizzato additivo senza la variabile relativa alla temperatura come nell'articolo [2], dal quale sono stati selezionati i modelli scelti. A causa di un errore del software R relativo ad un numero errato di predittori lineari non si è potuto procedere con le analisi relativamente a questo modello.

## 2.3 Convalida incrociata

Per verificare la precisione delle stime prodotte dai modelli si usa la tecnica della convalida incrociata. La serie temporale dei decessi settimanali osservati e le covariate dei modelli vengono separati in quattro gruppi. Il periodo di training dura cinque anni e dopo aver adattato il modello si prevedono le morti settimanali per le 52 settimane successive. Gli anni di inizio dei gruppi per la convalida incrociata vanno dal 2011 al 2014. Quindi si confrontano le morti attese  $\lambda_i$  con quelle osservate  $D_i$  nell'insieme dei dati test. L'errore di previsione viene quantificato attraverso l'errore medio percentuale

$$MPE = 1/n \sum_{i=1}^n \frac{D_i - \lambda_i}{D_i}$$

mentre l'accuratezza della previsione usando l'errore assoluto medio percentuale

$$MAPE = 1/n \sum_{i=1}^n \frac{|D_i - \lambda_i|}{D_i},$$

dove  $i$  varia sulle  $n$  previsioni ottenute per i quattro insiemi di dati test. Valori positivi del MPE indicano che il modello sottostima i decessi osservati quindi il modello considerato tende a predire un eccesso di

mortalità maggiore rispetto a quello effettivamente verificatosi.

Tutte le previsioni sono fatte per regione, settimana, sesso ed età. Le morti attese e osservate vengono aggregate prima per sesso, età e settimana, ottenendo i decessi annuali attesi per regione, e poi per età e sesso, ottenendo i decessi settimanali attesi per regione. L'errore di previsione e l'accuratezza vengono calcolati per regione e la loro distribuzione viene riassunta dai quantili.

# Capitolo 3

## Dati

Per implementare i modelli di regressione scelti sono serviti i dati relativi ai decessi settimanali e alla popolazione residente stratificati per sesso ed età nelle regioni italiane. Si è resa necessaria l'aggiunta delle informazioni sulle festività pubbliche e sulle settimane-persona di esposizione.

I dati acquisiti sono relativi agli anni dal 2011 al 2021, i dati dal 2015 al 2019 vengono utilizzati per stimare i decessi previsti per il 2020 e il 2021.

I dati relativi ai decessi settimanali, della popolazione residente, suddivisi per regione, sesso ed età sono stati ricavati dai dati sui decessi giornalieri per ogni singolo comune italiano forniti dall'ISTAT [6]. I comuni sono stati aggregati a livello regionale mentre il dato relativo al giorno è stato trasformato nella corrispondente settimana ISO, per poi procedere all'aggregazione settimanale del dato. Dall'insieme dei dati sono state eliminate le 53esime settimane del 2015 del 2020 in modo tale da avere lo stesso numero di settimane negli anni analizzati. I dati forniti dall'ISTAT sono suddivisi in 22 classi di età quinquennali che vanno da 0 a 100+ anni compiuti al momento del decesso. Successivamente le classi di età sono state aggregate in sei classi da 0 a 14, da 15 a 44, da 45 a 64, da 65 a 74, da 75 a 84 e 85+ [1]. Ottenendo così 12 combinazioni delle variabili di stratificazione sesso ed età.

Per il calcolo delle settimane-persona di esposizione è stato ricavato il dato relativo alla popolazione residente al primo gennaio suddivisa per regione, sesso ed età dal sito dell'ISTAT per gli anni dal 2011 al 2019 [7] e dal sito Demo-Geodemo per gli anni 2020 e 2021 [3, 4]. Le settimane-persona di esposizione sono state calcolate come la popolazione per una specifica settimana, età e sesso meno i decessi relativi alla medesima settimana moltiplicati per 0.5.

Per determinare per ogni settimana di ogni anno la tipologia di festività si è fatto riferimento ai dati presenti nel sito [Nager.Date:Worldwide Public Holiday](#) [5]. La distinzione è stata fatta per Natale, Pasqua e Capodanno. Le altre festività: Epifania, Lunedì dell'Angelo, Liberazione, festa del Lavoro, festa della Repubblica, Ferragosto, Ognissanti, Immacolata e Santo Stefano (negli anni in cui non cade la stessa settimana del Natale) sono state categorizzate come "altro" [9].

# Capitolo 4

## Risultati

Le stime italiane mostrano un aumento dell'eccesso di mortalità percentuale cumulativo nella primavera e nell'autunno del 2020 (Figura B.1) in linea con l'andamento dei casi di Covid-19 e tutti i modelli concordano sull'esistenza significativa di un eccesso di mortalità per il biennio 2020-2021 ( $P_{2020-2021} > 0$  con probabilità  $\geq 0.95$ ). L'eccesso di mortalità stimato per il biennio 2020-2021 in Italia varia, in base al modello, dal 7%, modello AVGr5, al 16%, modello SRFcem, rispetto al periodo considerato, mentre le stime basate sugli altri modelli sono comprese tra questi due valori.

Analizzando i risultati stratificati per sesso ed età i modelli sono tutti concordi nell'esistenza significativa di un eccesso di mortalità nel biennio 2020-2021 per la popolazione femminile e maschile di età superiore ai 65 anni. Le differenze nella popolazione femminile e maschile (figure B.2 e B.3), a parità di classe di età, risiedono nel livello dell'eccesso di mortalità percentuale, infatti le stime per la popolazione femminile sono minori di quelle maschili in tutte le classi di età eccetto che nella prima, dove la stima dell'eccesso di mortalità percentuale cumulativo per i maschi è negativo, e nella classe di età dai 15 ai 44 anni dove il valore massimo dell'eccesso di mortalità percentuale cumulativo coincide. L'andamento settimanale dell'eccesso di mortalità percentuale nelle classi di età  $[0,15)$  e  $[15,44)$  ha un andamento diverso rispetto a quello delle altre classi di età, difatti per quanto riguarda la prima classe di età non si notano aumenti nell'eccesso di mortalità percentuale cumulativo mentre per la seconda si nota un aumento nel 2021, ma non secondo tutti i modelli. Da notare che per gli ultra ottantacinquenni il modello che prevede l'eccesso di mortalità percentuale maggiore per il 2020-2021 non è il Serfling Euromomo ma il *5 year average death count*.

Come mostrato in figura 4.1 i modelli sono concordi sull'andamento

settimanale dell'eccesso di mortalità percentuale cumulativo, a variare è il livello dell'eccesso con differenze dai 5 ai 19 punti percentuali tra il P-score cumulativo minore e quello maggiore a seconda della regione considerata. Gli andamenti settimanali variano da regione a regione, ma in tutte le regioni, tranne la Calabria, è riscontrabile un aumento del P-score cumulativo nella primavera e nell'autunno del 2020. In Piemonte e Puglia sono visibili tre aumenti nell'eccesso di mortalità nella primavera e nell'autunno del 2020 e nella primavera del 2021. In alcune regioni come Friuli-Venezia Giulia, Liguria, Piemonte, Sardegna, Trentino Alto Adige e Valle d'Aosta sono chiaramente distinguibili le due ondate di Covid-19 del 2020. Caso diverso la regione Veneto che mostra due massimi dell'eccesso di mortalità cumulativo percentuale settimanale, uno nella primavera del 2020 e il secondo fra dicembre 2020 e gennaio 2021. Per Sardegna, Abruzzo, Basilicata, Toscana, Calabria, Trentino Alto Adige, Friuli-Venezia Giulia, Molise, Lazio, Valle d'Aosta, Umbria e Marche i modelli risultano in disaccordo sull'esistenza significativa di un eccesso di mortalità percentuale per il 2020-2021.

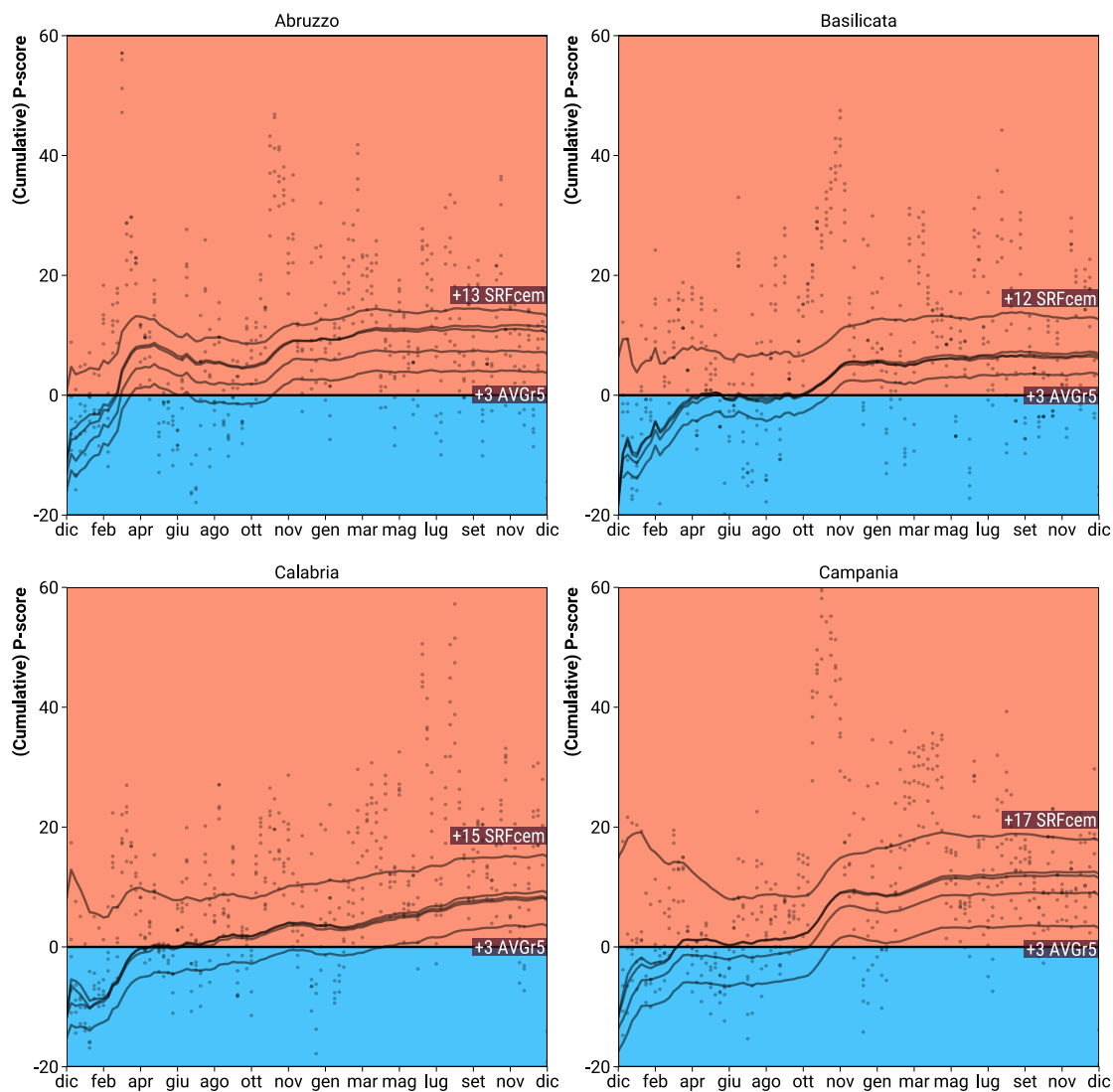
Per 15 regioni su 20 il modello che ha prodotto la stima massima dell'eccesso di mortalità percentuale per i due anni analizzati è il Serfling stile Euromomo (SRFcem), mentre la stima minima per tutte le regioni è stata ottenuta col modello *5 year average death rate* (AVGr5), la cui curva cumulata ha la distanza maggiore rispetto alle altre in quasi tutte le regioni. In quasi tutte le regioni l'eccesso di mortalità percentuale cumulato è maggiore di zero già a gennaio del 2020 perché questo modello è stato adattato non tenendo conto delle settimane tipiche dell'influenza stagionale, come le settimane di gennaio.

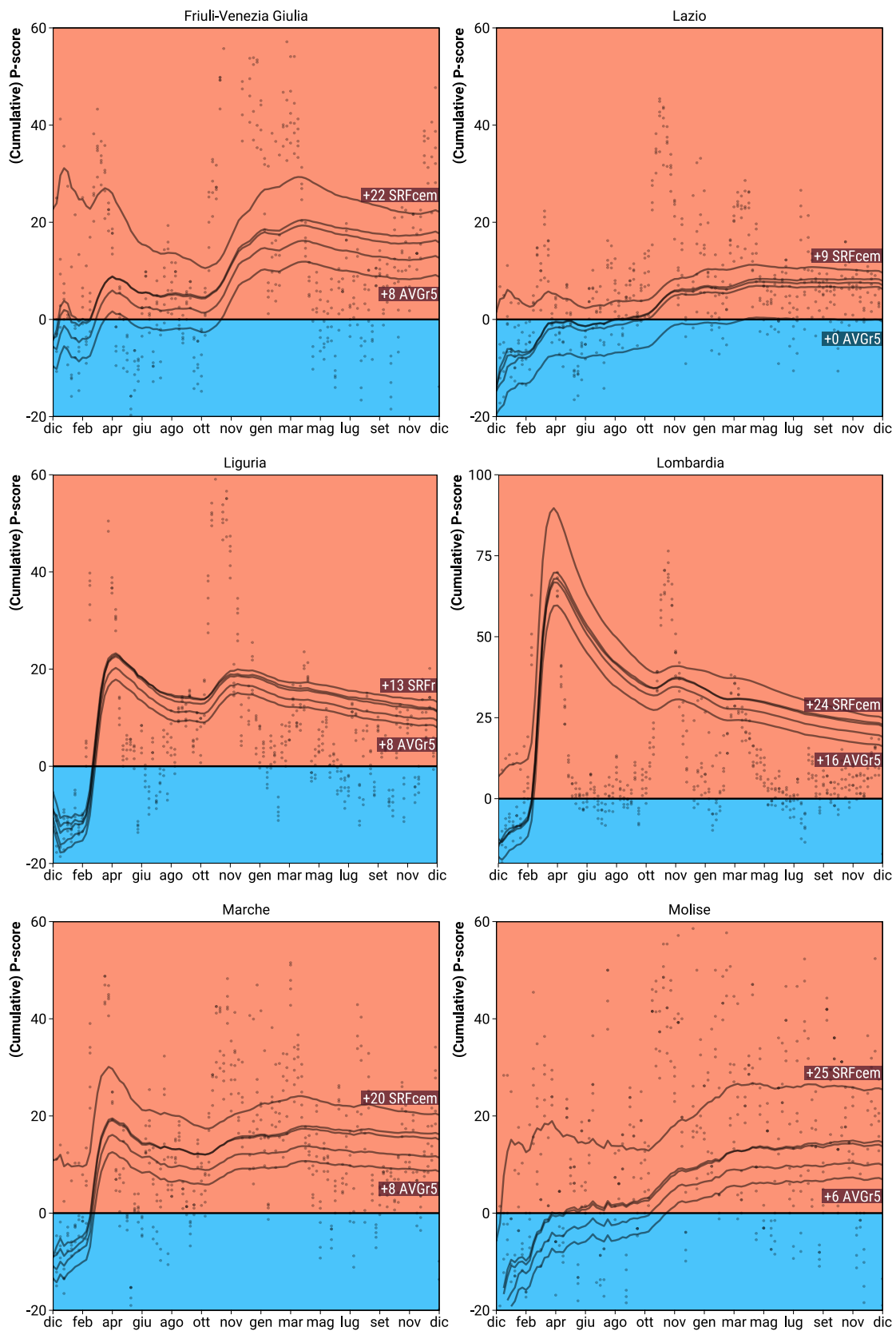
La Lombardia è la regione che ha registrato il valore più elevato di eccesso di mortalità percentuale settimanale, il 311% di decessi in più rispetto alle previsioni per la stessa settimana del periodo di riferimento. Dalla Tabella A.2 emerge che 6 regioni hanno registrato il valore più alto dell'eccesso di mortalità percentuale settimanale la settimana del 23 marzo 2020, altre due regioni la settimana successiva mentre per le restanti regioni il massimo è avvenuto fra novembre e dicembre. La Calabria con un massimo dell'eccesso di mortalità percentuale settimanale pari al 29% è la regione col P-score massimo meno elevato del 2020 in Italia. Nel 2021 (tabella A.4) i valori massimi dell'eccesso di mortalità percentuale settimanale sono minori rispetto al 2020 in tutte le regioni tranne che in Calabria, Molise e Sicilia. La regione che ha registrato il valore massimo di eccesso di mortalità percentuale settimanale del 2021 è il Friuli-Venezia Giulia con un incremento del 91% di decessi rispetto alla

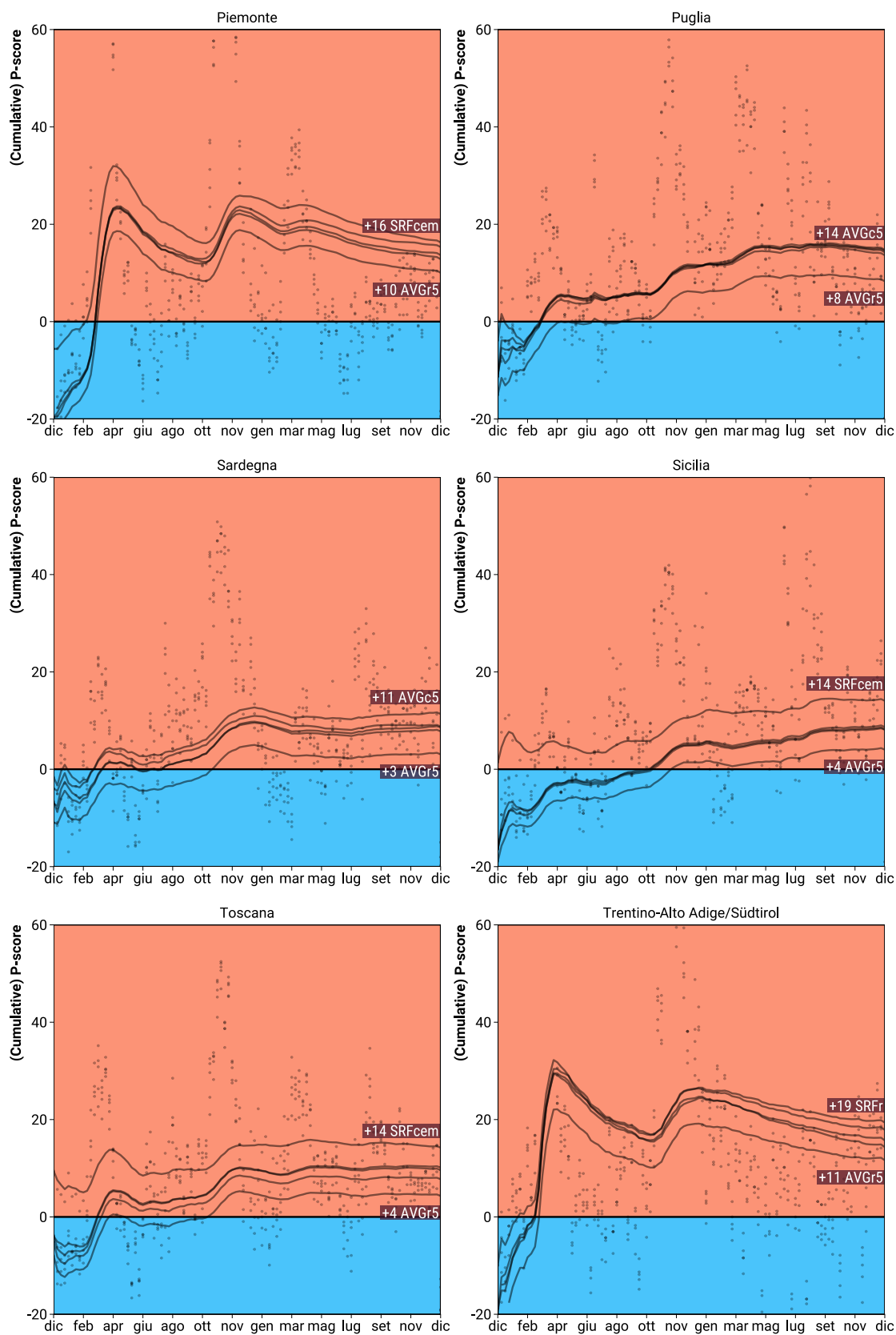
stessa settimana del periodo di riferimento. I valori massimi dell'eccesso di mortalità percentuale settimanale di ogni regione sono stati raggiunti in tre periodi diversi: gennaio, marzo-aprile e agosto.

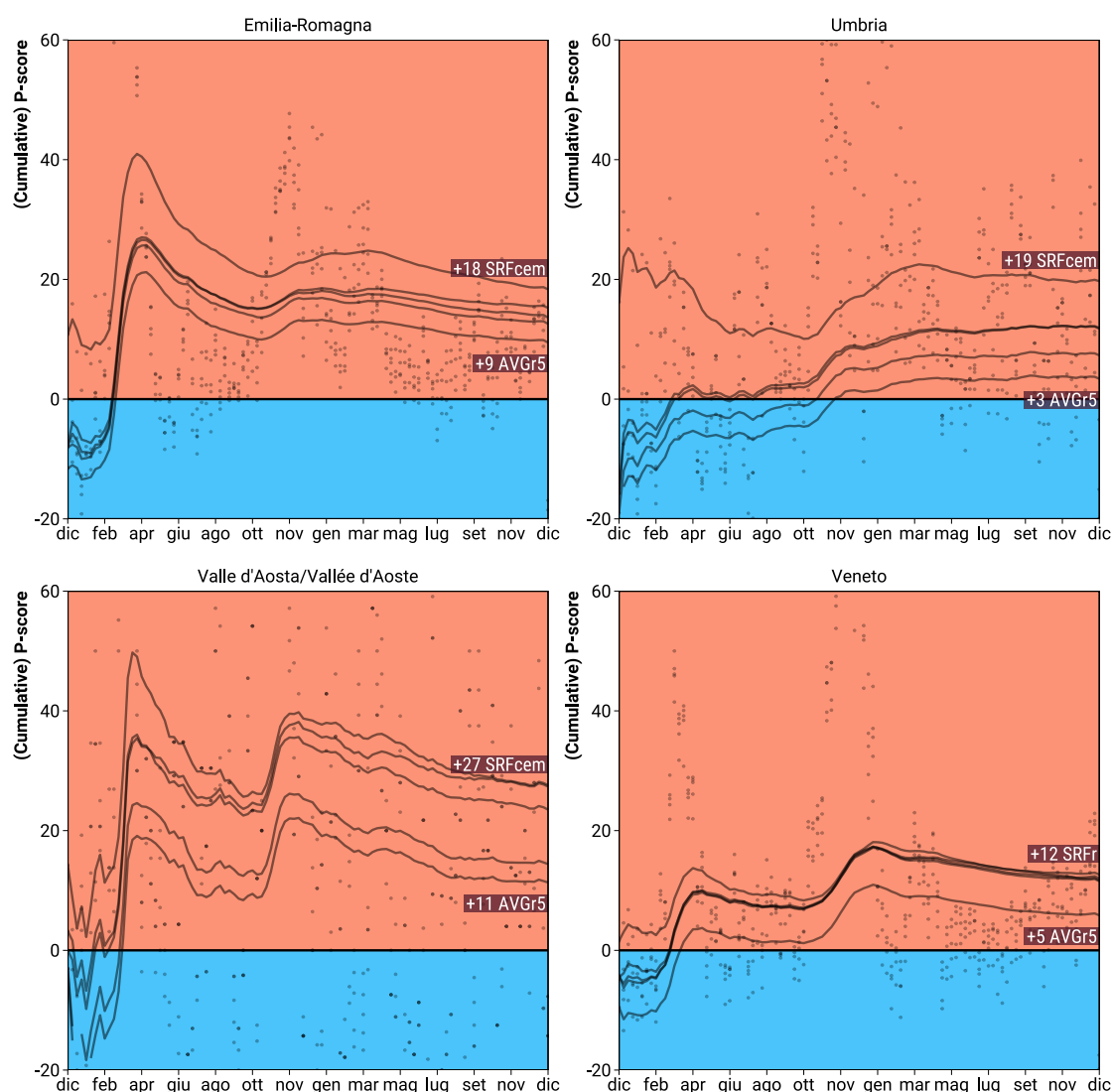
I grafici B.4 - B.43 riportano l'eccesso di mortalità percentuale settimanale stratificato per sesso ed età nelle regioni italiane. Le regioni con meno di 600 000 abitanti (Valle d'Aosta, Molise e Basilicata) mostrano andamenti "anomali" o errori per la popolazione con meno di 45 anni di età, questo è dovuto alla presenza di molte settimane con zero decessi per queste età. Per lo stesso motivo Umbria, Toscana, Trentino-Alto Adige, Sardegna, Marche, Liguria, Friuli-Venezia Giulia e Abruzzo mostrano andamenti "anomali" nella popolazione tra 0 e 14 anni. Stratificando per sesso ed età troviamo che tutti i modelli sono concordi nell'esistenza significativa di un eccesso di mortalità percentuale per il 2020-2021 nella popolazione con un'età maggiore di 45 anni di entrambi i sessi in Piemonte, Emilia Romagna e Lombardia. Diversamente in Friuli-Venezia Giulia, Liguria, Marche e Trentino Alto Adige i modelli sono concordi per entrambi i sessi a partire dai 65 anni di età mentre in Calabria, Campania, Puglia e Sicilia sono concordi solamente nelle classi di età [45,65) e [65, 75) per entrambi i sessi, come mostrato in Tabella A.1. Stratificando per sesso ed età appare ancora più evidente che la scelta del modello influisce notevolmente sulla stima dell'eccesso di mortalità. Infatti in base al sesso, all'età e alla regione i valori massimi e minimi sono ottenuti dalle stime di modelli diversi.

Figura 4.1: Eccesso di mortalità percentuale settimanale cumulativo predetto da 5 modelli diversi per il biennio 2020-2021 nelle regioni italiane. *I punti indicano l'eccesso di mortalità percentuale settimanale mentre la curva rappresenta quello cumulativo. Le etichette si riferiscono al valore dell'eccesso di mortalità percentuale cumulato all'ultima settimana del 2021.*







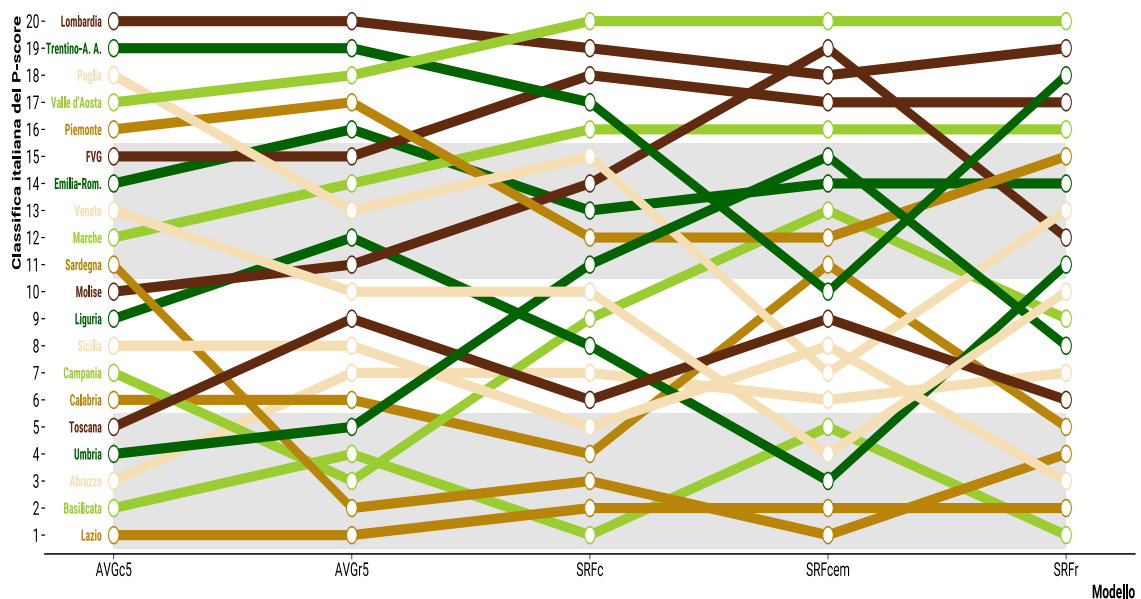


La classifica delle regioni fatta rispetto al valore del P-score cumulativo per il 2020-2021 cambia molto in base al modello usato per la stima (Figura 4.2). I modelli mostrano un buon accordo per quanto riguarda Lombardia e Valle d'Aosta che si trovano tra le regioni con l'eccesso di mortalità percentuale cumulativo maggiore e nella regione col valore minore (Lazio). I modelli *5 year average* mostrano un buon grado di accordo in quasi la metà delle regioni. Si ha un buon accordo anche tra i modelli Serfling con e senza esposizione nelle posizioni più alte e basse della classifica. La classifica fatta usando le stime del modello Serfling stile Euromomo varia molto rispetto alle altre, per esempio il Veneto si trova alla decima posizione basandosi sulle stime dei modelli AVGr5, SRFc e SRFr mentre col modello SRFcem passa alla quarta posizione, situazione inversa col Molise che nella classifica basata sulle stime del

modello Serfling stile Euromomo si ritrova ad essere la seconda regione con l'eccesso di mortalità percentuale maggiore. Quindi le variazioni maggiori nelle classifiche si hanno tenendo conto o meno delle settimane relative all'influenza stagionale e considerando o meno il trend e la stagionalità della mortalità. Tenere conto della struttura della popolazione a parità delle altre variabili esplicative non sembra influire molto sui risultati nella maggior parte delle regioni.

Stratificando per sesso ed età i modelli di tipo *5 year average* e Serfling mostrano un buon livello di accordo nella popolazione tra i 45 e i 64 anni (Figura B.44). Anche nelle classifiche fatte nei diversi strati il modello che cambia maggiormente le posizioni è il modello Serfling stile Euromomo.

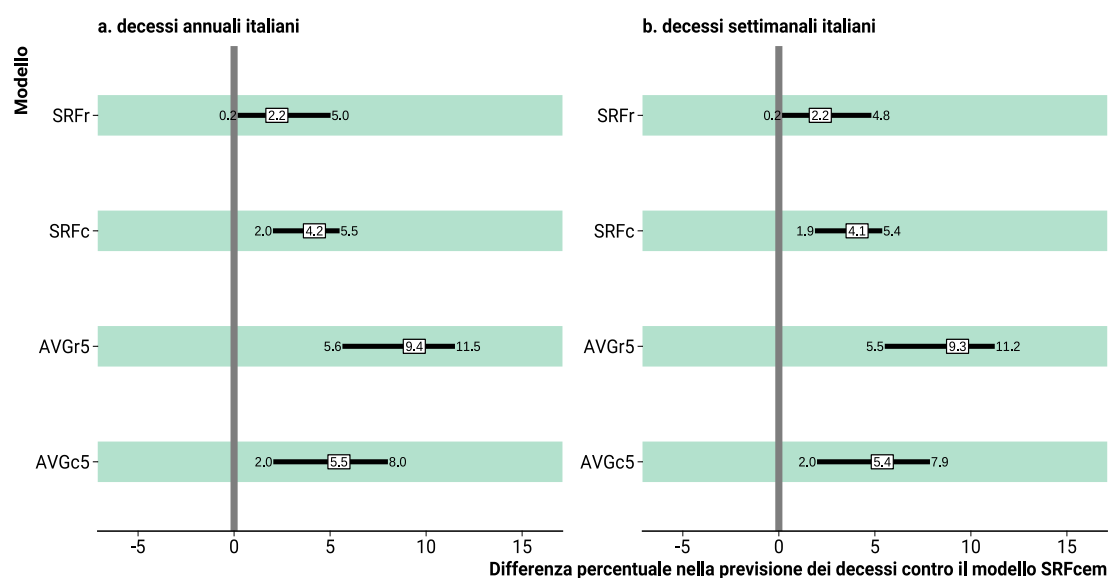
Figura 4.2: Classifica delle regioni italiane fatta rispetto all'eccesso di mortalità percentuale per il 2020-2021 rispetto a cinque modelli diversi. Il valore 1 indica l'eccesso di mortalità minore.



In figura 4.3 vengono confrontate le morti predette dai vari modelli con le morti predette dal modello Serfling Euromomo (SRFcem), preso come modello di riferimento. Tutti i modelli tendono a prevedere un maggior numero di decessi annuali e settimanali rispetto al modello di riferimento, senza differenze nel passare dai dati annuali ai dati settimanali. Il modello *5 year average death rate* (AVGr5) ha la differenza percentuale maggiore, con previsioni settimanali maggiori dal 5.5% all'11.2% rispetto al riferimento per metà delle regioni italiane. Quindi

il modello Serfling prevede un numero di decessi molto diverso rispetto agli altri modelli, soprattutto rispetto ai modelli *5 year average*. Stratificando i risultati variano (figure B.45 e B.46) tranne per il modello *5 year average death rate* il quale predice sempre meno decessi rispetto al modello di riferimento. Nella prima classe di età i modelli hanno una differenza percentuale negativa o lo zero è compreso tra i quantili 0.25 e 0.75 della distribuzione, ad indicare che in alcune regioni la previsione dei decessi è uguale. È da tenere presente l'elevata presenza di zeri in molte settimane di diverse regioni relativamente a questa classe di età. Per la popolazione femminile con un'età compresa tra i 15 e i 44 anni i modelli di tipo Serfling hanno una differenza percentuale negativa, tuttavia lo zero è compreso tra i quantili 0.25 e 0.75 della distribuzione, succede lo stesso per i decessi previsti dal modello *5 year average death count* per gli ultra ottantacinquenni sia fra i maschi che tra le femmine. Mentre nei restanti strati continuano a verificarsi le tendenze già osservate ad un livello di aggregazione maggiore.

Figura 4.3: Differenza percentuale delle morti predette da diversi modelli contro le previsioni del modello Serfling tipo Euromomo. *Le differenze percentuali nelle previsioni italiane dei decessi sono riassunte dai quantili 0.25, 0.5 e 0.75.*

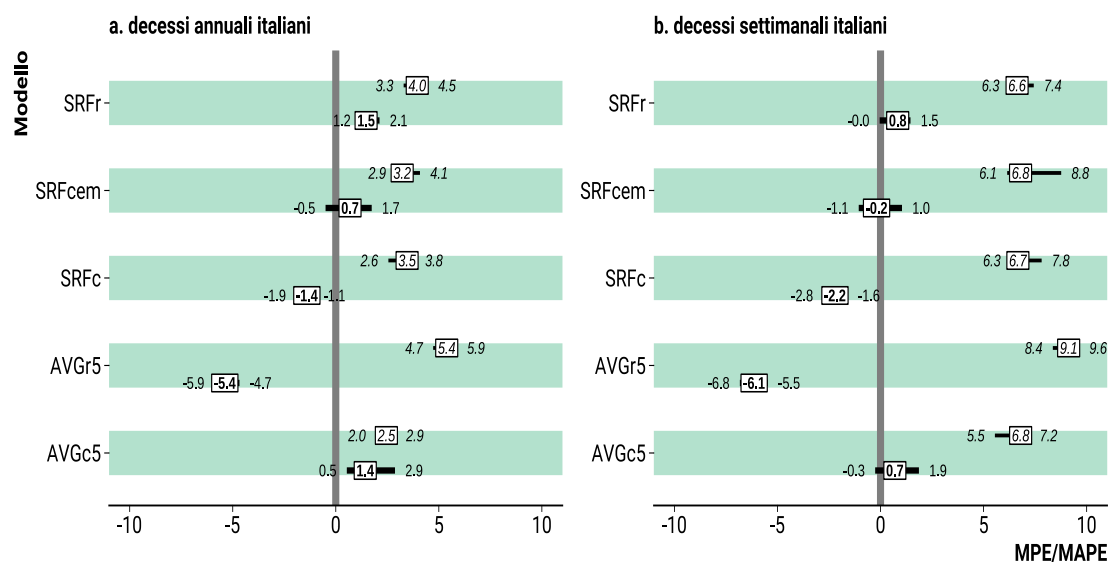


Le analisi eseguite con la tecnica della convalida incrociata hanno portato ai risultati seguenti. Il modello che mostra la propensione maggiore a sovrastimare i decessi, sia annuali che settimanali, è il modello *5 year average death rate*, il quale sovrastima i risultati annuali in 19

regioni su 20 (figura 4.4). Dimostrando quanto già osservato nei grafici precedenti di come questo modello porti a stimare un eccesso di mortalità molto minore rispetto agli altri modelli. I modelli che tendono a sottostimare i decessi annuali sono i modelli Serfling con esposizione (SRFr), Serfling stile Euromomo (SRFcem) e *5 year average death count* (AVGc5). Il modello che commette l'errore di previsione minore è il modello Serfling stile Euromomo sia a livello annuale che settimanale. Passando alla suddivisione in strati (figure da B.47 a B.49) il modello *5 year average death count* cambia il verso della stima passando agli ultra ottantacinquenni, quindi passa da sovrastimare a sottostimare i decessi. Nelle due classi di età che comprendono la popolazione tra i 45 e i 74 anni il modello con l'errore di previsione minore è il modello Serfling senza esposizione, mentre per gli ultra settantacinquenni il modello che commette meno errore di previsione è il modello Serfling stile Euromomo, tranne nello strato femmine - 85+ dove il valore del errore medio percentuale del modello *5 year average death rate* è prossimo allo zero. Tendenzialmente il modello *5 year average death rate* tende a sovrastimare i risultati in tutti gli strati.

L'accuratezza delle stime diminuisce nel passare dai dati annuali a quelli settimanali (figura 4.4). I decessi annuali sono predetti con un MAPE che varia dal 4% al 5.4% aumentando a 6.8-9.1% passando alle previsioni settimanali. A livello annuale il modello con il MAPE minore è il *5 year average death count*. Negli strati (figure da B.47 a B.49) i valori del MAPE aumentano ulteriormente. Non considerando la popolazione tra 0 e 44 anni, i quantili 0.75 della distribuzione dell'errore assoluto medio percentuale sono inferiori al 15% in tutti i modelli considerati. Il modello *5 year average death rate* è il modello che produce gli errori più elevati confermando la scarsità di robustezza di questo modello. Da notare che le distribuzioni degli errori del modello Serfling Euromomo, a livello di strati, sono quelle con la maggiore variabilità. Un modello che da buoni risultati previsivi è il modello Serfling senza esposizione e nell'ultima classe di età il modello Serfling con esposizione.

Figura 4.4: Errore medio percentuale (MPE, in grassetto in basso) ed errore medio assoluto percentuale (MAPE, in corsivo in alto) commessi dai modelli nel prevedere le morti sui dati test. *La distribuzione degli errori è riassunta dai quantili 0.25, 0.5 e 0.75.*



Dal grafico in figura 4.5 emerge che il modello *5 year average death rate* prevede un eccesso di mortalità di 5-6 mila decessi in meno rispetto agli altri modelli per il biennio 2020-2021 in Veneto. Lo stesso succede nelle previsioni per la regione Sicilia (figura 4.6), regione con un numero di abitanti simile al Veneto, dove la differenza si attesta tra i 4 e i 10 mila decessi in meno. Come già osservato in figura 4.4 il modello AVGr5 tende a sovrastimare i decessi dal 5.5% al 6.8% in metà delle regioni, portando a stime dell'eccesso di mortalità minori rispetto a quelle basate su altri modelli. Una spiegazione della differenza nelle previsioni del modello *5 year average death rate* rispetto agli altri modelli può essere il calo demografico in corso in Italia e il conseguente aumento della popolazione nelle classi di età più anziane dove il numero dei decessi è maggiore. Infatti questo modello tiene conto della struttura della popolazione ma non del trend della mortalità. L'effetto della struttura della popolazione non sembra influenzare allo stesso modo il modello Serfling con esposizione che tiene conto del trend della mortalità e della stagionalità. Inoltre è da tenere presente che il modello AVGr5 risulta essere il modello meno accurato nelle previsioni settimanali (figura 4.4).

Figura 4.5: Confronto tra la stima dell'eccesso di mortalità cumulativo previsto da cinque modelli diversi per il 2020-2021 per la regione Veneto. In rosso è riportato l'eccesso di mortalità previsto dal modello, in grigio l'intervallo di previsione al 95%, l'etichetta si riferisce al valore cumulativo dell'eccesso di mortalità previsto nel biennio 2020-2021.

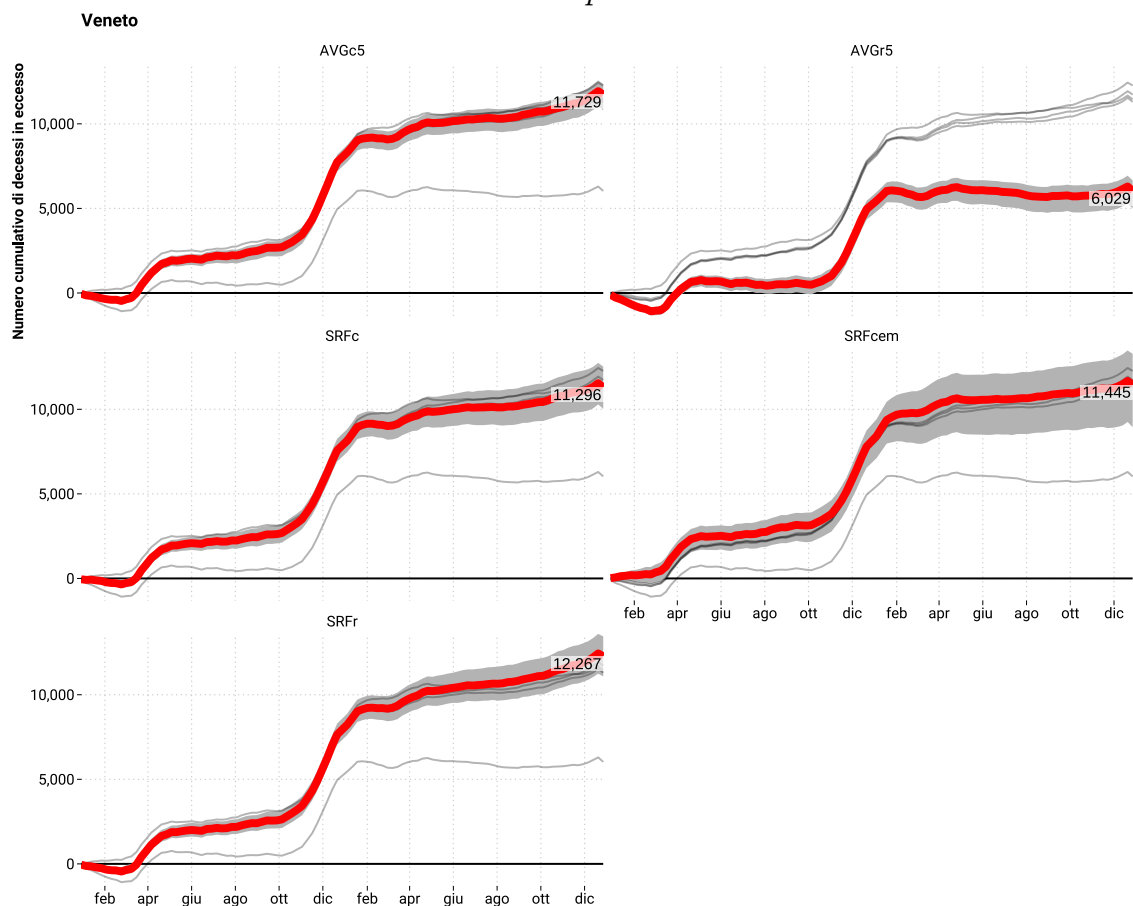
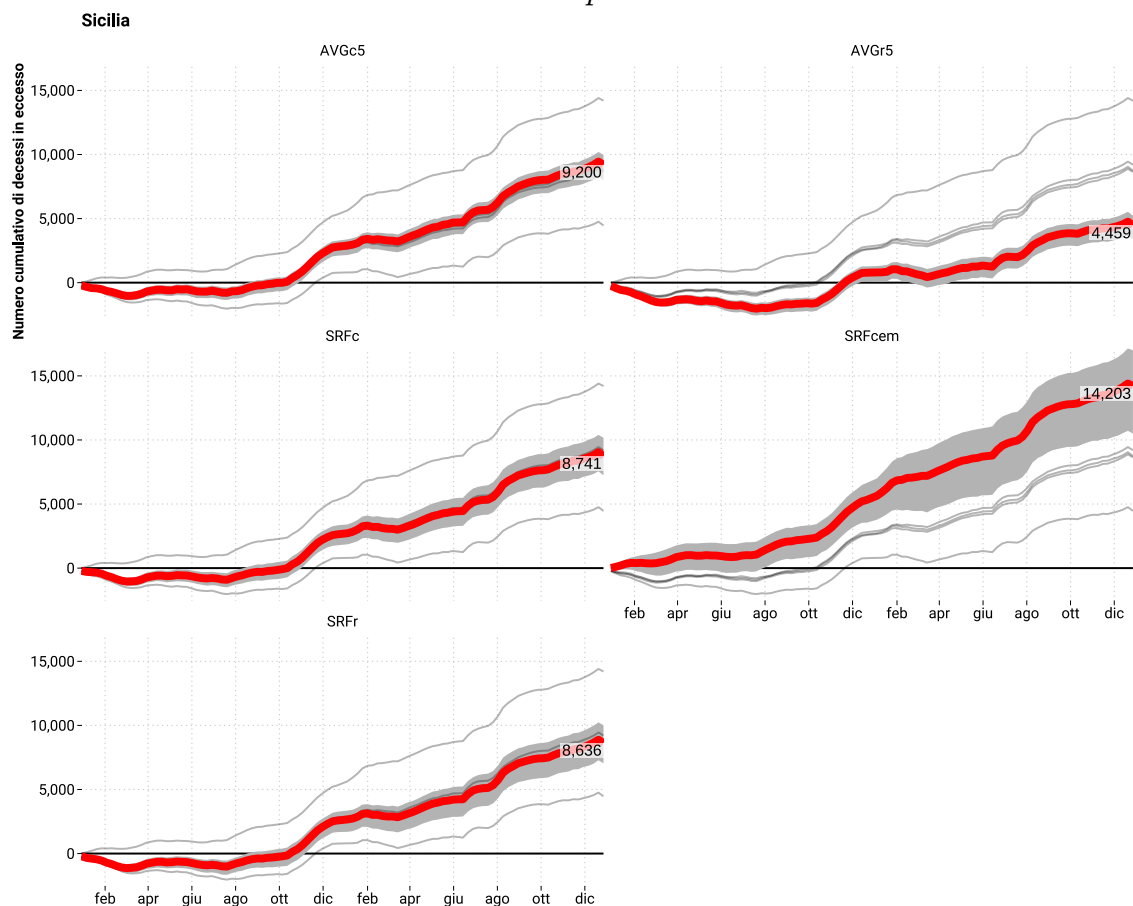


Figura 4.6: Confronto tra la stima dell'eccesso di mortalità cumulativo previsto da cinque modelli diversi per il 2020-2021 per la regione Sicilia. *In rosso è riportato l'eccesso di mortalità previsto dal modello, in grigio l'intervallo di previsione al 95%, l'etichetta si riferisce al valore cumulativo dell'eccesso di mortalità previsto nel biennio 2020-2021.*



# Capitolo 5

## Conclusione

La scelta del modello per stimare l'eccesso di mortalità in un dato anno risulta molto importante infatti, per esempio, scegliendo il modello *5 year average death rate*, che tende a sovrastimare i decessi, si otterrebbe un eccesso di mortalità minore rispetto alle previsioni effettuate con modelli diversi. Dalle analisi svolte con i dati italiani tutti i modelli considerati prevedono un numero di decessi maggiore rispetto al modello Serfling stile Euromomo, il quale è stato adattato senza usare i dati relativi alle settimane dell'influenza stagionale, infatti le stime dell'eccesso di mortalità percentuale cumulativo sono positive già a gennaio del 2020 mentre per gli altri modelli sono negative.

Risultati simili sono stati ottenuti anche su uno studio relativo a 20 nazioni europee dove il modello *5 year average death rate* sovrastima i decessi maggiormente rispetto agli altri modelli ed ha un errore maggiore nella stima annuale e settimanale [9]. Questo modello risulta essere il peggiore in quanto non tenendo conto del trend della mortalità ma solo della struttura della popolazione sovrastima eccessivamente le morti, portando in certi casi a prevedere un eccesso di mortalità cumulato negativo dove altri modelli lo prevederebbero positivo.

La stima di questo eccesso di mortalità varia da modello a modello e da regione a regione andando da un incremento del 27% ottenuto col modello Serfling stile Euromomo in Valle d'Aosta ad un incremento del 0% ottenuto col modello *5 year average death rate* in Lazio. Mentre i modelli non sono concordi nell'affermare l'esistenza significativa di un eccesso di mortalità per il periodo 2020-2021 per metà delle regioni, lo sono a livello nazionale dove il valore massimo dell'eccesso di mortalità percentuale cumulativo è stato raggiunto nella primavera del 2020.

La presenza di molte settimane con zero decessi nelle età più giovani per le regioni meno popolate ha portato all'impossibilità di stimare il valore

dell'eccesso di mortalità percentuale per molte settimane e ad errori più elevati rispetto alle classi di età oltre i 45 anni. Infatti la mediana della distribuzione dell'errore assoluto medio percentuale è superiore al 10% nelle classi di età [0, 15) e [15, 45).

Tutti i modelli sono concordi nell'esistenza significativa di un eccesso di mortalità per gli ultra sessantacinquenni a livello italiano, mentre per alcune regioni l'esistenza significativa di un eccesso di mortalità è verificata già a partire dai quarantacinquenni. Tuttavia l'esistenza o meno è molto variabile tra le regioni stratificando per sesso ed età, come per esempio in Valle d'Aosta dove i modelli non sono concordi sull'esistenza significativa di un eccesso di mortalità per nessuna delle classi d'età di entrambi i sessi.

Sicuramente il Covid-19 ha influito sul numero di decessi avvenuti in Italia nel 2020 e nel 2021 ma l'entità è variabile tra le regioni e in base al modello usato per le stime. Per quanto riguarda il 2021 l'eccesso di mortalità percentuale cumulativo risulta minore rispetto ai valori massimi raggiunti nel 2020 indicando una stabilità e/o meno decessi rispetto a quelli previsti dai modelli in alcune settimane dell'anno.

Quindi se si ha la necessità di valutare una stima dell'eccesso di mortalità per un determinato anno l'uso di un modello di tipo *5 year average death rate* è sconsigliato poiché sovrastima i decessi ed è il modello con l'errore maggiore tra quelli analizzati. Mentre il modello Serfling stile Euromomo, non tenendo conto delle settimane dell'influenza stagionale nell'adattare il modello, prevede eccessi di mortalità contenenti in parte i decessi tipici della stagione influenzale. Un buon compromesso è il modello Serfling senza esposizione, il quale rispetto ad un modello *5 year average* tiene conto del trend della mortalità e della stagionalità.

# Appendice A

## Tabelle

	F45	M45	F65	M65	F75	M75	F85	M85
Abruzzo				x				x
Basilicata				x			x	
Calabria	x	x	x	x				
Campania	x	x	x	x	x			
Emilia-Rom.		x	x	x	x	x	x	x
FVG			x	x	x	x	x	x
Lazio				x		x		
Liguria			x	x		x		x
Lombardia	x	x	x	x	x	x	x	x
Marche			x	x	x	x	x	x
Molise							x	
Piemonte		x	x	x	x	x	x	x
Puglia	x	x	x	x	x			x
Sardegna		x				x		
Sicilia	x	x	x	x		x		
Toscana	x		x	x	x	x		
Trentino-A. A.				x	x	x	x	x
Umbria				x			x	
Valle d'Aosta								
Veneto				x	x	x	x	x

Tabella A.1: Strati per cui il quantile 0.05 della distribuzione del P-score cumulativo all'ultima settimana del 2021 risulta maggiore di zero per tutti i modelli.

	<b>Settimana</b>	<b>Modello</b>	<b>Pscore</b>
<b>Abruzzo</b>	23-3	SRFcem	0.57
<b>Basilicata</b>	30-11	SRFcem	0.48
<b>Calabria</b>	30-11	SRFcem	0.29
<b>Campania</b>	9-11	SRFcem	0.67
<b>Emilia-Romagna</b>	23-3	SRFcem	1.26
<b>Friuli-Venezia Giulia</b>	14-12	SRFcem	1.05
<b>Lazio</b>	2-11	SRFce	0.45
<b>Liguria</b>	30-3	SRFr	1.16
<b>Lombardia</b>	23-3	SRFcem	3.11
<b>Marche</b>	23-3	SRFcem	0.93
<b>Molise</b>	23-11	SRFcem	0.64
<b>Piemonte</b>	23-3	SRFr	0.4
<b>Puglia</b>	16-11	AVGc5	0.61
<b>Sardegna</b>	2-11	AVGc5	0.51
<b>Sicilia</b>	16-11	SRFcem	0.42
<b>Toscana</b>	9-11	SRFr	0.52
<b>Trentino-Alto Adige</b>	23-3	SRFr	1.51
<b>Umbria</b>	7-12	SRFcem	0.61
<b>Valle d'Aosta</b>	30-3	SRFcem	2.56
<b>Veneto</b>	14-12	SRFr	0.7

Tabella A.2: Valore massimo del P-score per il 2020, modello da cui sono stati ottenuti i valori dei decessi attesi e data del lunedì della settimana corrispondente.

	<b>Settimana</b>	<b>Modello</b>	<b>Eccesso</b>
<b>Abruzzo</b>	23-3	SRFcem	161
<b>Basilicata</b>	30-11	SRFcem	57
<b>Calabria</b>	30-11	SRFcem	106
<b>Campania</b>	9-11	SRFcem	662
<b>Emilia-Romagna</b>	23-3	SRFcem	1171
<b>Friuli-Venezia Giulia</b>	14-12	SRFcem	272
<b>Lazio</b>	2-11	SRFc	476
<b>Liguria</b>	30-3	SRFr	467
<b>Lombardia</b>	23-3	SRFcem	5765
<b>Marche</b>	23-3	SRFcem	302
<b>Molise</b>	23-11	SRFcem	41
<b>Piemonte</b>	23-3	SRFr	1200
<b>Puglia</b>	16-11	AVGc5	439
<b>Sardegna</b>	2-11	AVGc5	155
<b>Sicilia</b>	16-11	SRFcem	404
<b>Toscana</b>	9-11	SRFr	426
<b>Trentino-Alto Adige</b>	23-3	SRFr	283
<b>Umbria</b>	7-12	SRFcem	109
<b>Valle d'Aosta</b>	30-3	SRFcem	64
<b>Veneto</b>	14-12	SRFr	757

Tabella A.3: Valore massimo dell'eccesso di mortalità per il 2020, modello da cui sono stati ottenuti i valori dei decessi attesi e data del lunedì della settimana corrispondente.

	<b>Settimana</b>	<b>Modello</b>	<b>Pscore</b>
<b>Abruzzo</b>	15-3	SRFcem	0.42
<b>Basilicata</b>	2-8	SRFcem	0.44
<b>Calabria</b>	9-8	SRFcem	0.57
<b>Campania</b>	9-8	SRFcem	0.39
<b>Emilia-Romagna</b>	4-1	SRFcem	0.45
<b>Friuli-Venezia Giulia</b>	11-1	SRFcem	0.91
<b>Lazio</b>	18-1	SRFcem	0.32
<b>Liguria</b>	11-1	SRFc	0.23
<b>Lombardia</b>	15-3	SRFcem	0.38
<b>Marche</b>	22-3	SRFr	0.51
<b>Molise</b>	8-3	SRFcem	0.86
<b>Piemonte</b>	5-4	SRFr	0.4
<b>Puglia</b>	12-04	SRFr	0.51
<b>Sardegna</b>	9-8	SRFr	0.3
<b>Sicilia</b>	9-8	SRFcem	0.76
<b>Toscana</b>	16-8	SRFcem	0.35
<b>Trentino-Alto Adige</b>	4-1	SRFcem	0.49
<b>Umbria</b>	1-2	SRFcem	0.6
<b>Valle d'Aosta</b>	13-12	SRFcem	0.84
<b>Veneto</b>	4-1	SRFcem	0.6

Tabella A.4: Valore massimo del P-score per il 2021, modello da cui sono stati ottenuti i valori dei decessi attesi e data del lunedì della settimana corrispondente.

	<b>Settimana</b>	<b>Modello</b>	<b>Eccesso</b>
<b>Abruzzo</b>	15-3	SRFcem	120
<b>Basilicata</b>	2-8	SRFcem	46
<b>Calabria</b>	9-8	SRFcem	198
<b>Campania</b>	8-3	SRFcem	378
<b>Emilia-Romagna</b>	4-1	SRFcem	440
<b>Friuli-Venezia Giulia</b>	11-1	SRFcem	237
<b>Lazio</b>	18-1	SRFcem	388
<b>Liguria</b>	11-1	SRFc	109
<b>Lombardia</b>	15-3	SRFcem	719
<b>Marche</b>	22-3	SRFr	168
<b>Molise</b>	8-3	SRFcem	57
<b>Piemonte</b>	5-4	SRFr	379
<b>Puglia</b>	22-3	AVGc5	390
<b>Sardegna</b>	4-1	SRFcem	102
<b>Sicilia</b>	9-8	SRFcem	655
<b>Toscana</b>	22-3	SRFr	267
<b>Trentino-Alto Adige</b>	4-1	SRFcem	99
<b>Umbria</b>	1-2	SRFcem	105
<b>Valle d'Aosta</b>	13-12	SRFcem	21
<b>Veneto</b>	4-1	SRFcem	624

Tabella A.5: Valore massimo dell'eccesso di mortalità per il 2021, modello da cui sono stati ottenuti i valori dei decessi attesi e data del lunedì della settimana corrispondente.



# Appendice B

## Grafici

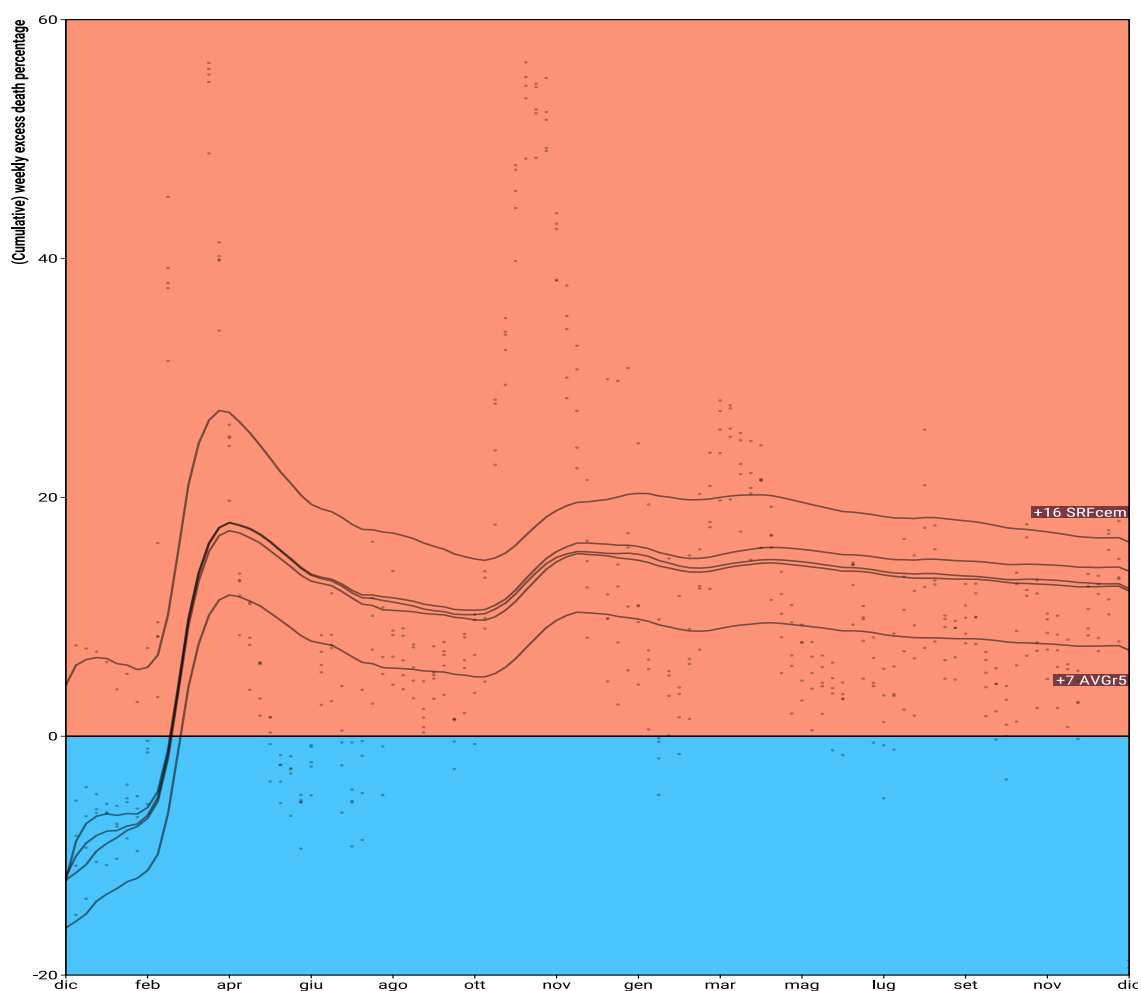


Figura B.1: Eccesso di mortalità percentuale settimanale cumulativo predetto da 5 modelli diversi per il biennio 2020-2021 Italia. *I punti indicano l'eccesso di mortalità percentuale settimanale mentre la curva rappresenta quello cumulativo. Le etichette si riferiscono al valore dell'eccesso di mortalità percentuale cumulato all'ultima settimana del 2021. L'asse Y è troncato al 60%.*

A seguire i grafici dell'eccesso di mortalità percentuale cumulativo settimanale predetto da 5 modelli diversi per il 2020-2021 per le regioni italiane e per l'Italia stratificando per sesso ed età. Le classi di età sono, partendo dalla prima a sinistra fino all'ultima a destra, nell'ordine [0, 14), [15, 44), [45, 64), [65, 74), [75, 84) e 85+. *I punti indicano l'eccesso di mortalità percentuale settimanale mentre la curva rappresenta quello cumulativo.* Le etichette si riferiscono al valore dell'eccesso di mortalità percentuale cumulato all'ultima settimana del 2021.

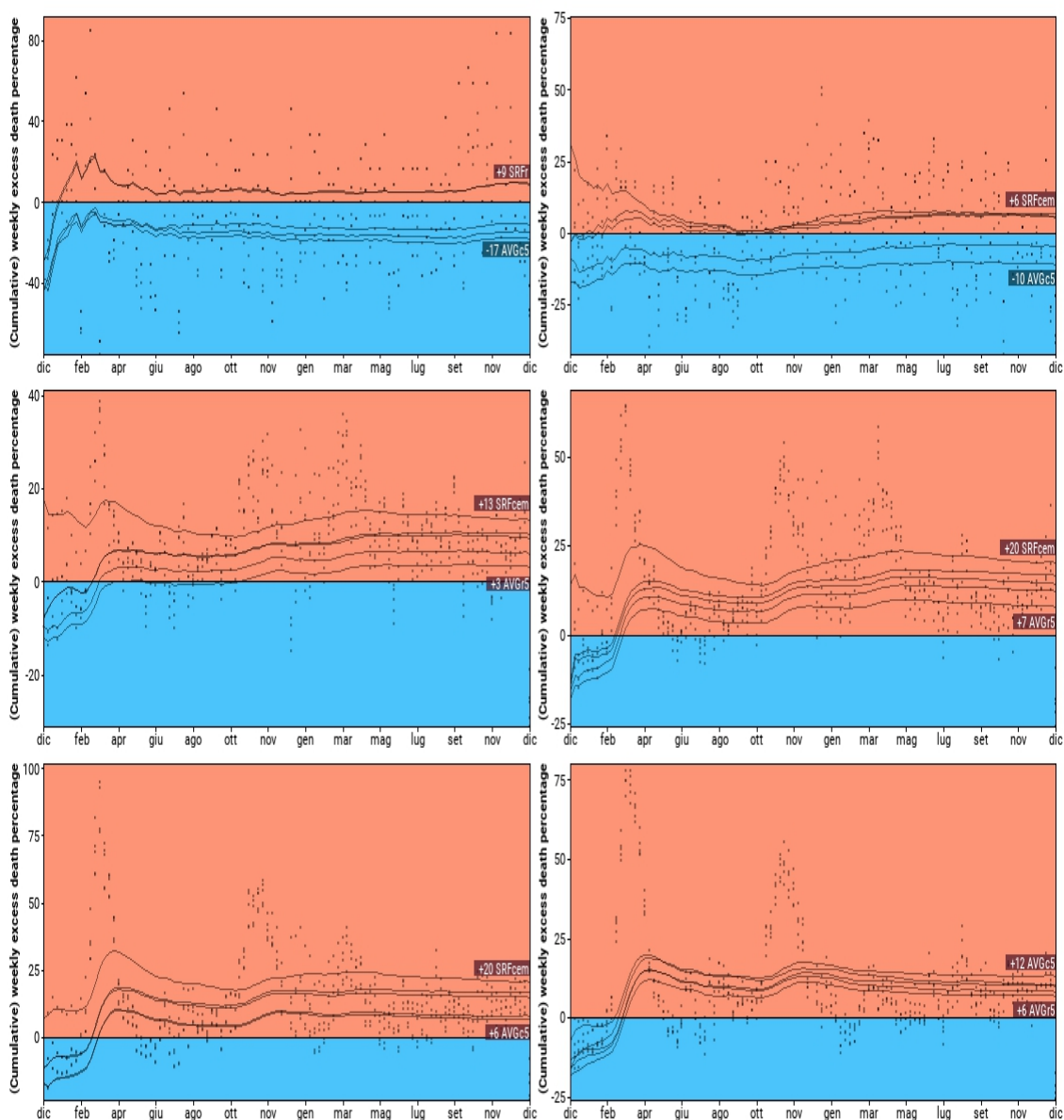


Figura B.2: Popolazione femminile dell'Italia.

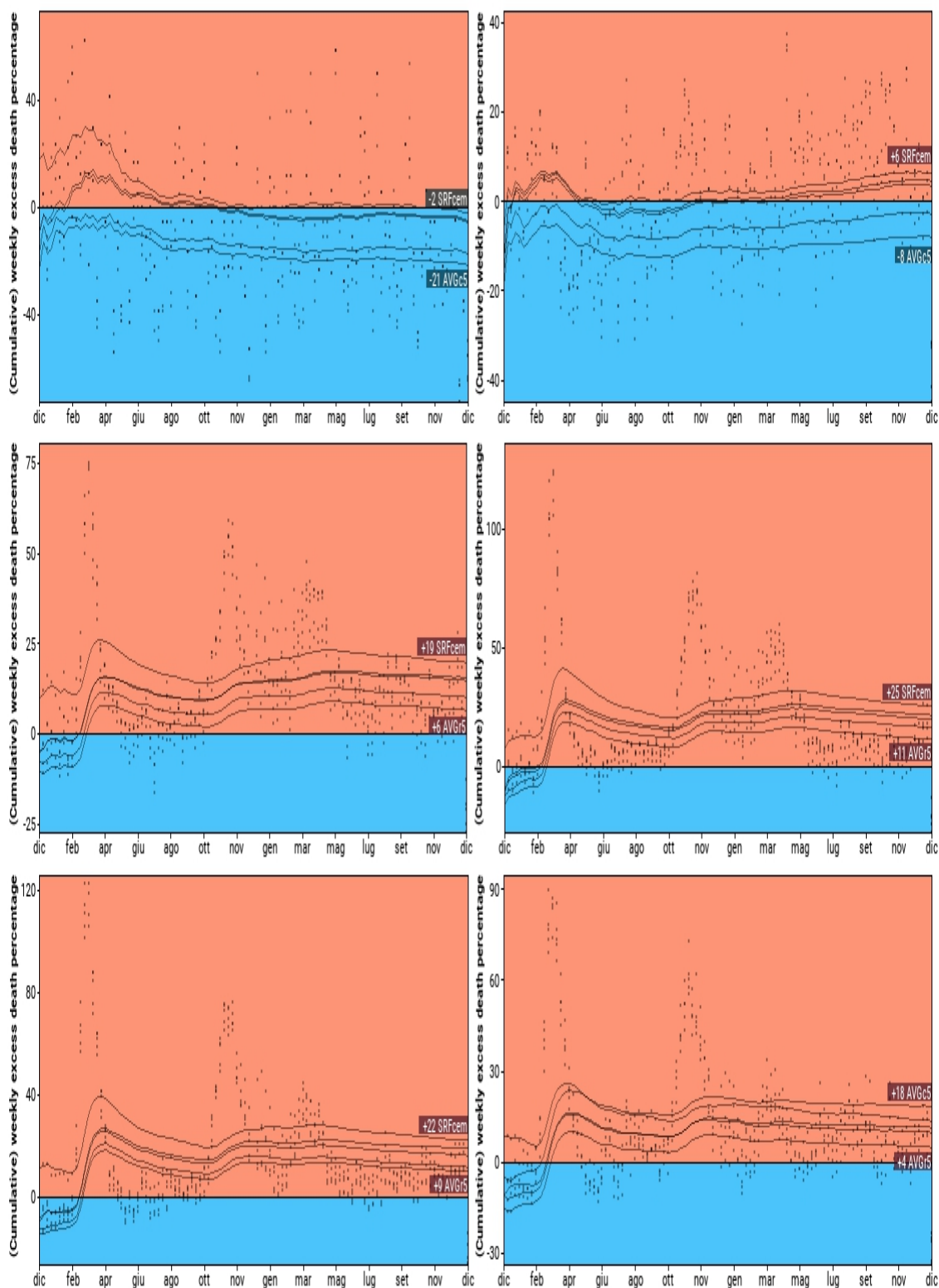


Figura B.3: Popolazione maschile dell'Italia.

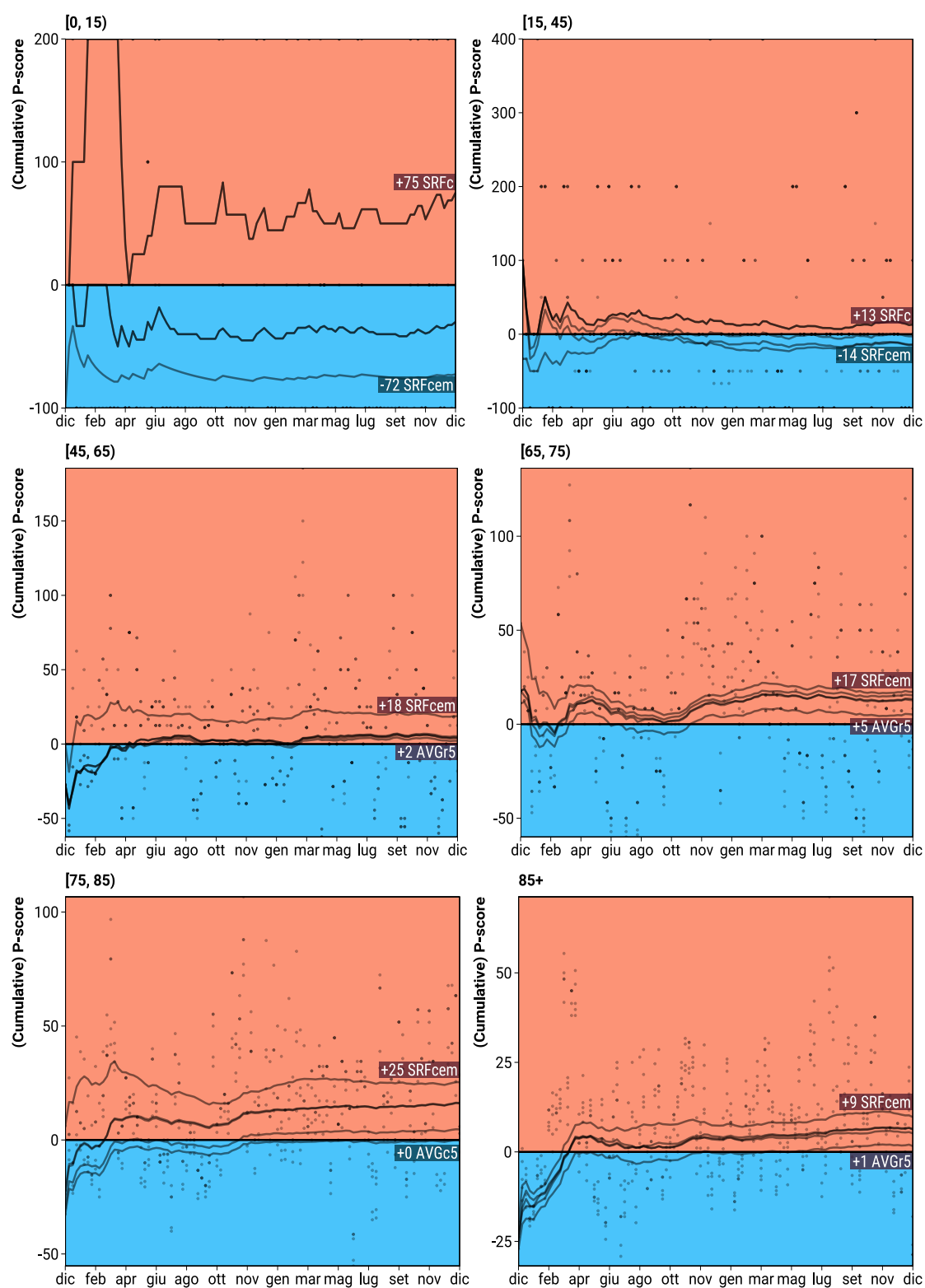


Figura B.4: Popolazione femminile dell'Abruzzo.

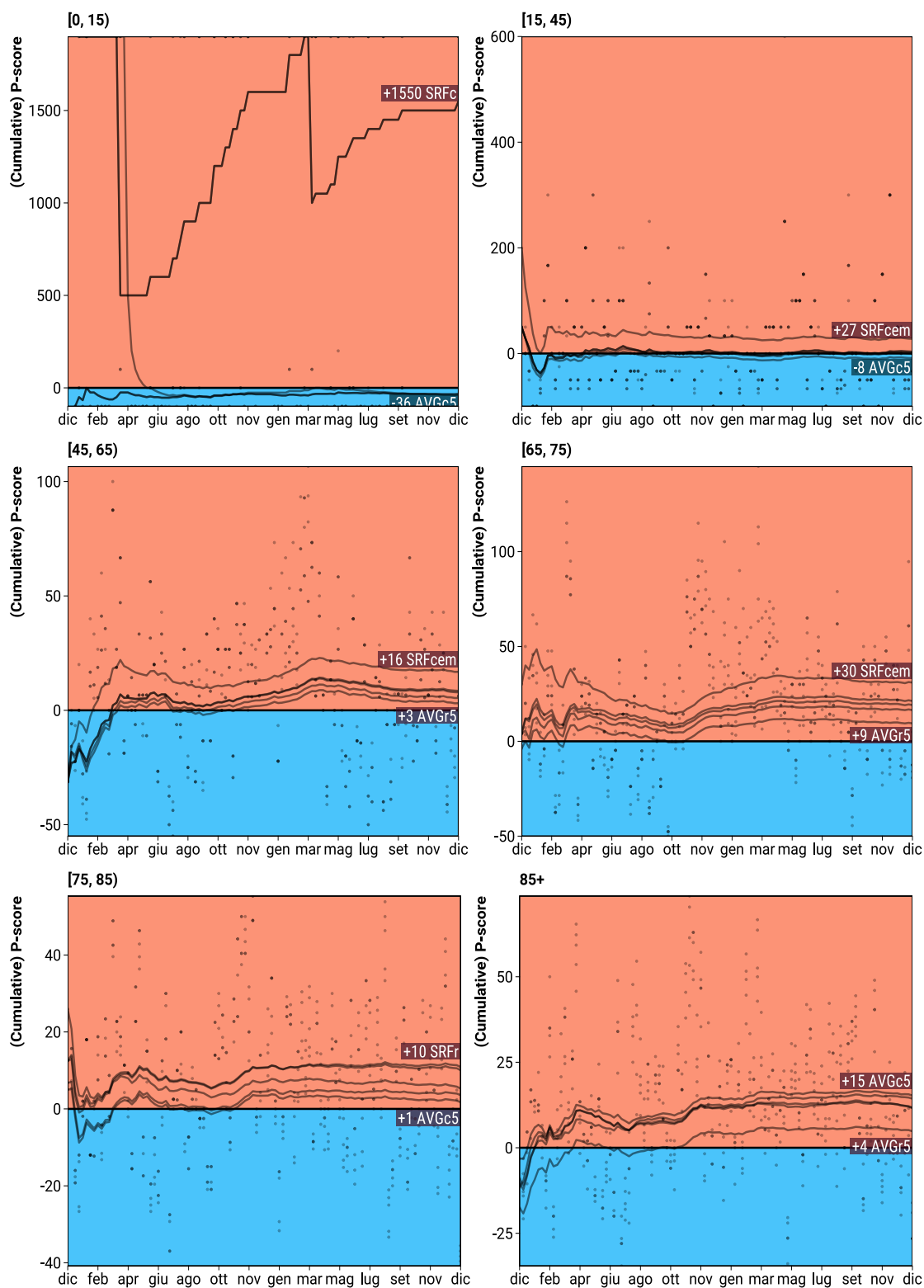


Figura B.5: Popolazione maschile dell'Abruzzo.

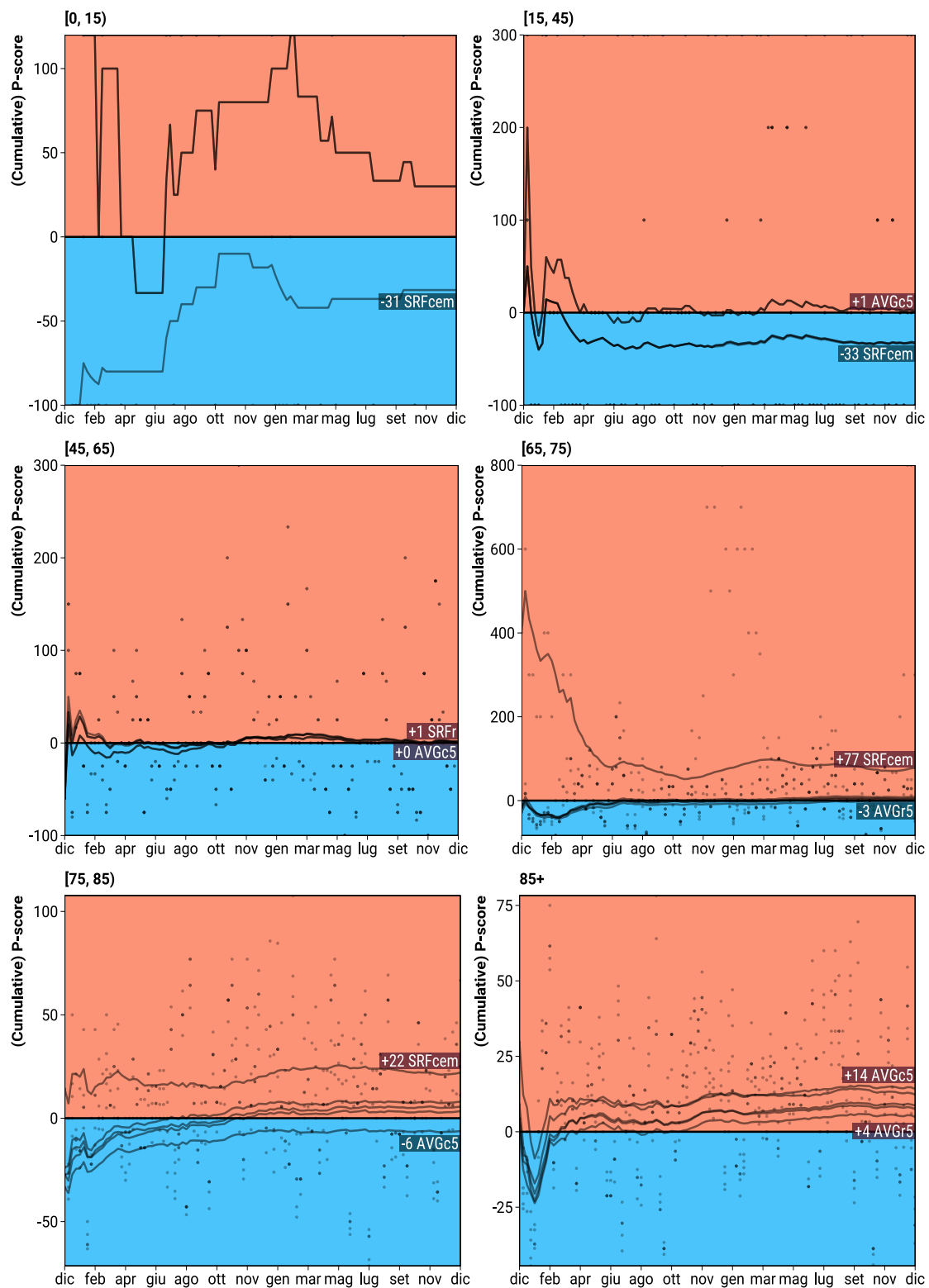


Figura B.6: Popolazione femminile della Basilicata.

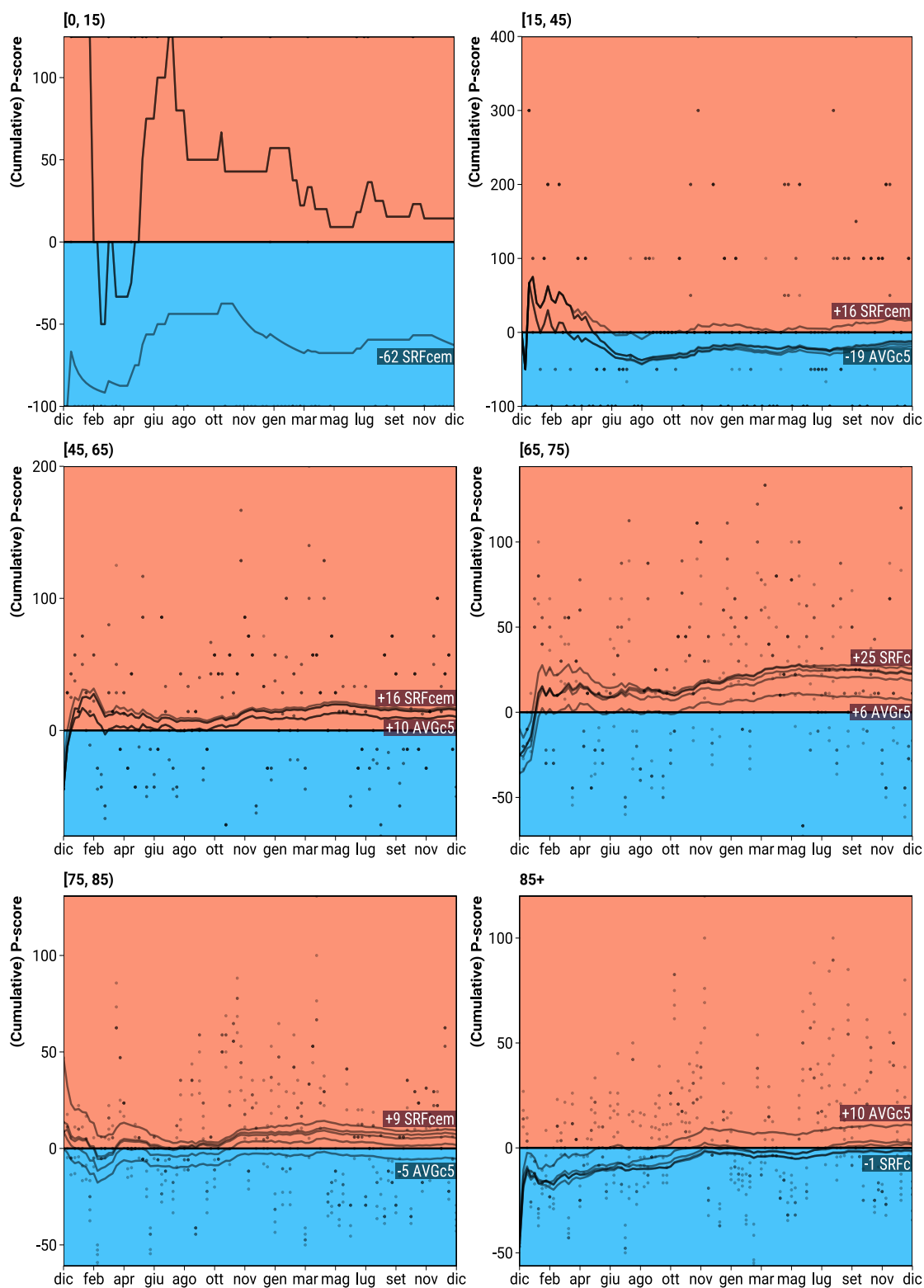


Figura B.7: Popolazione maschile della Basilicata.

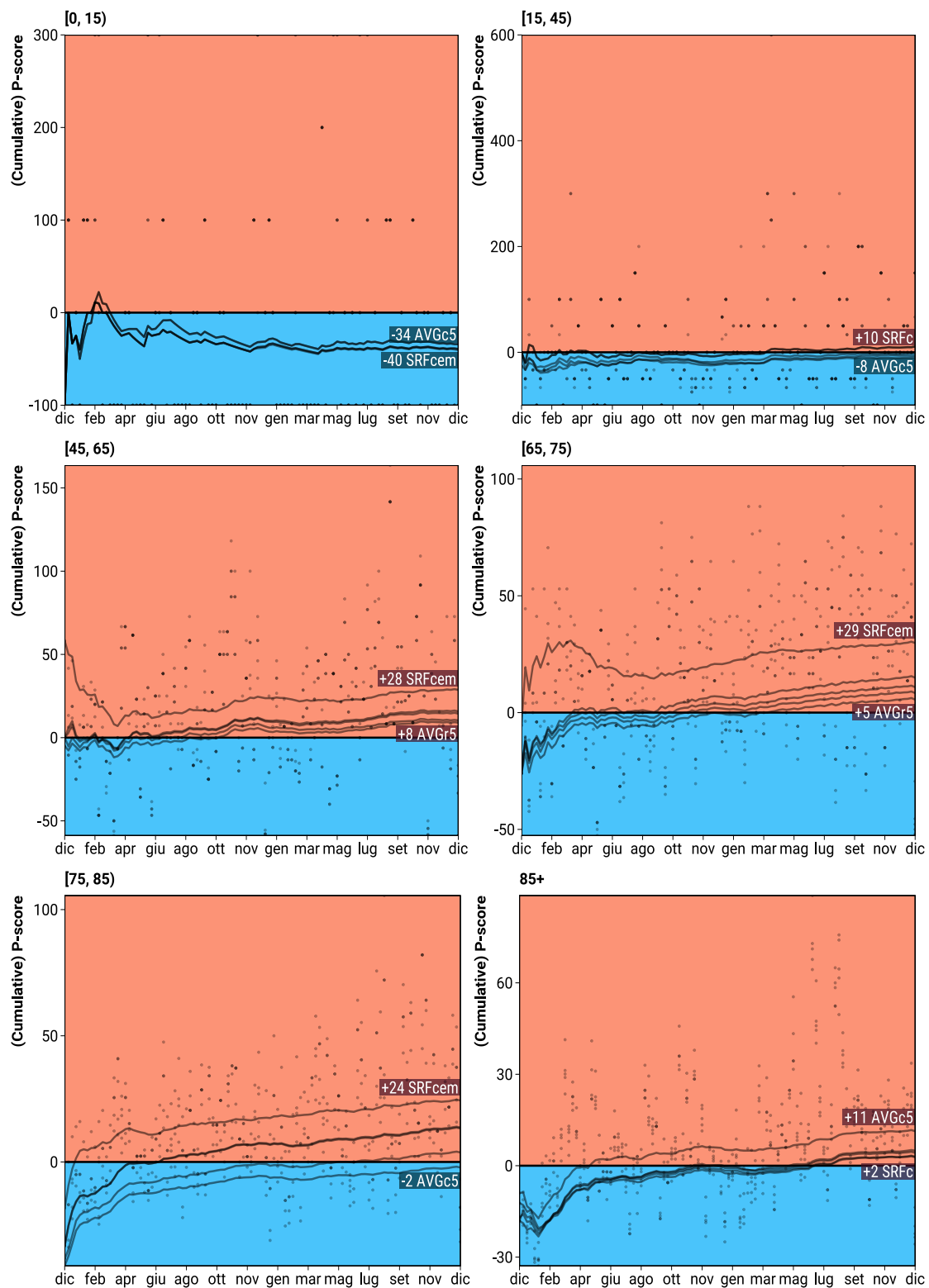


Figura B.8: Popolazione femminile della Calabria.

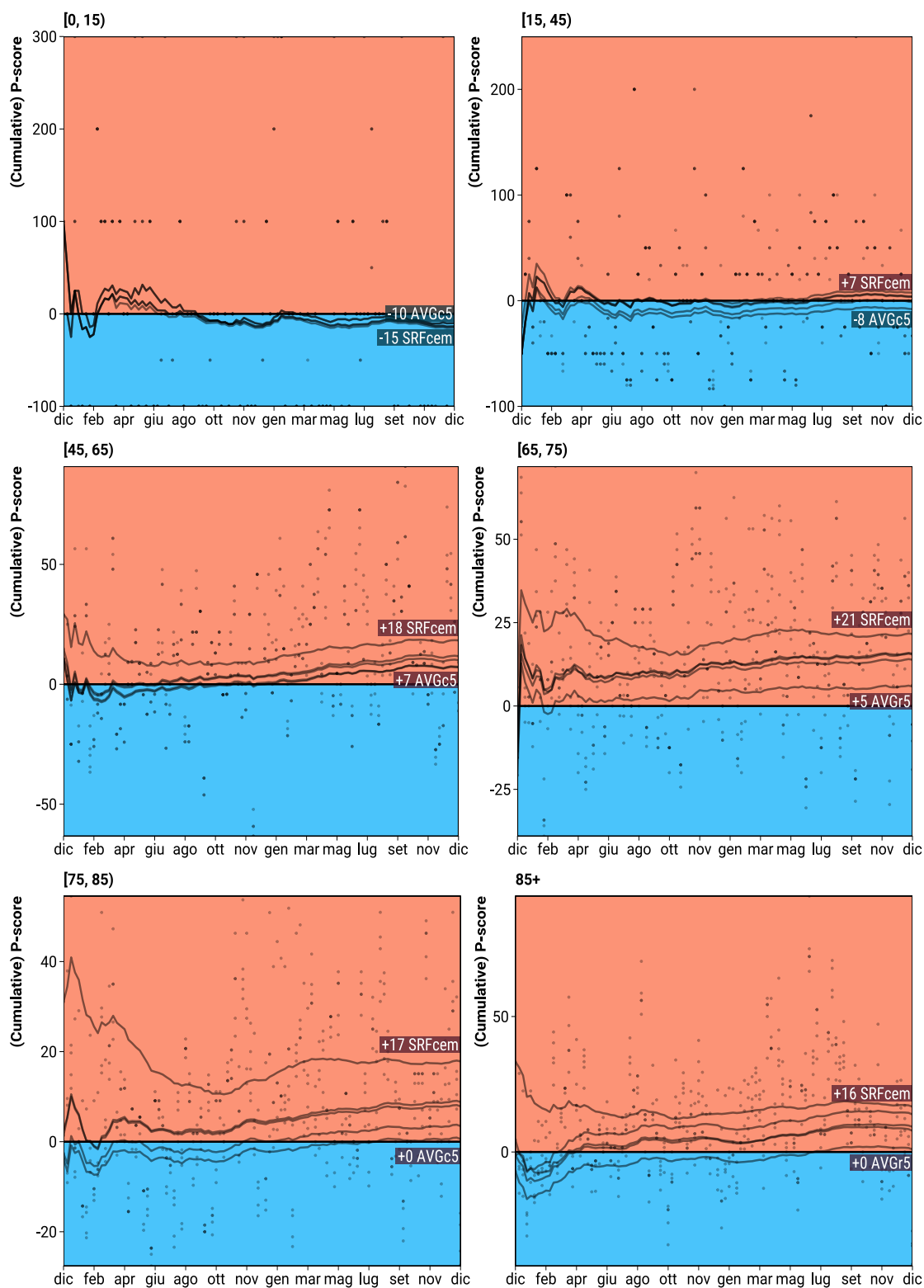


Figura B.9: Popolazione maschile della Calabria.

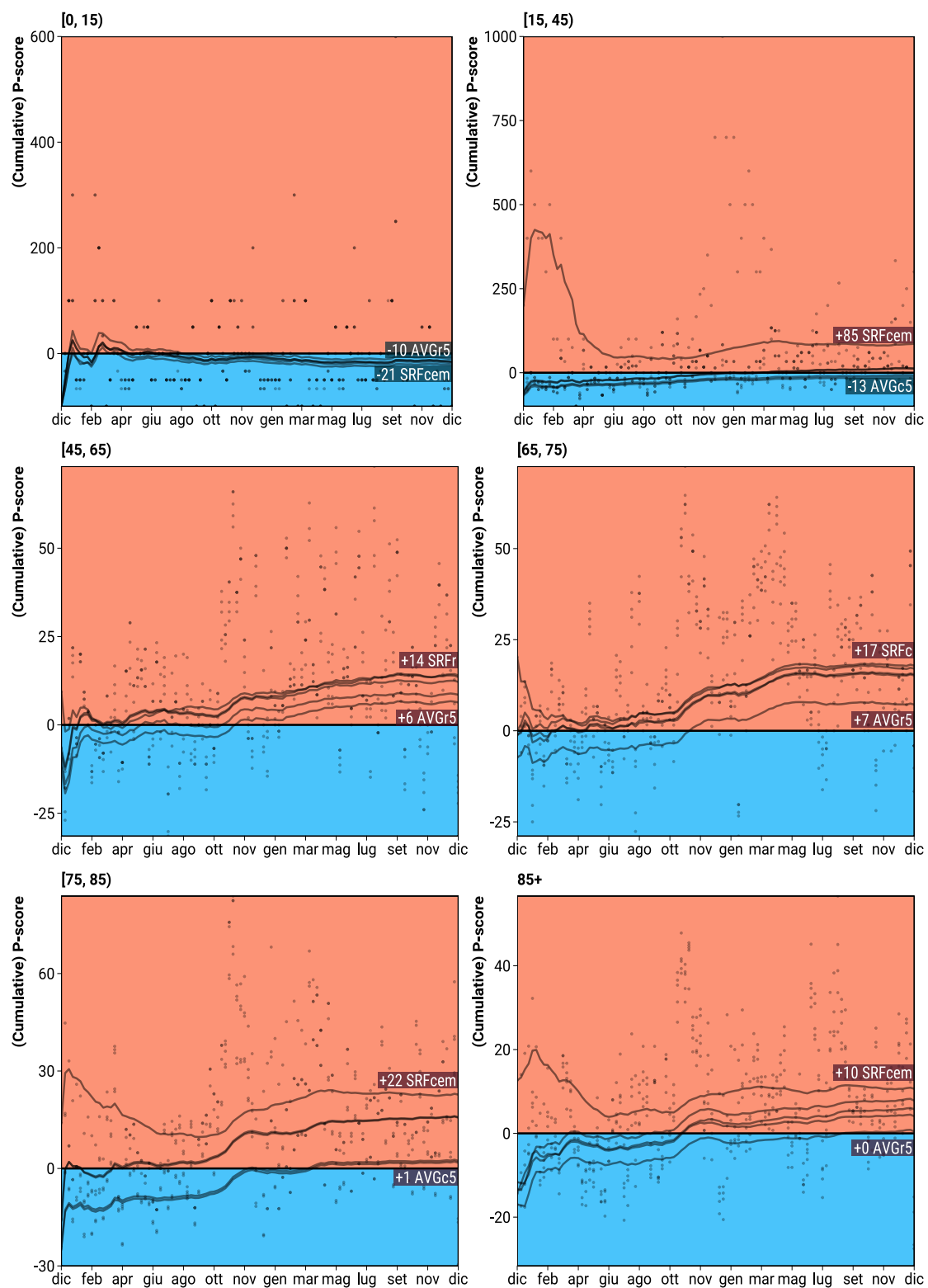


Figura B.10: Popolazione femminile della Campania.

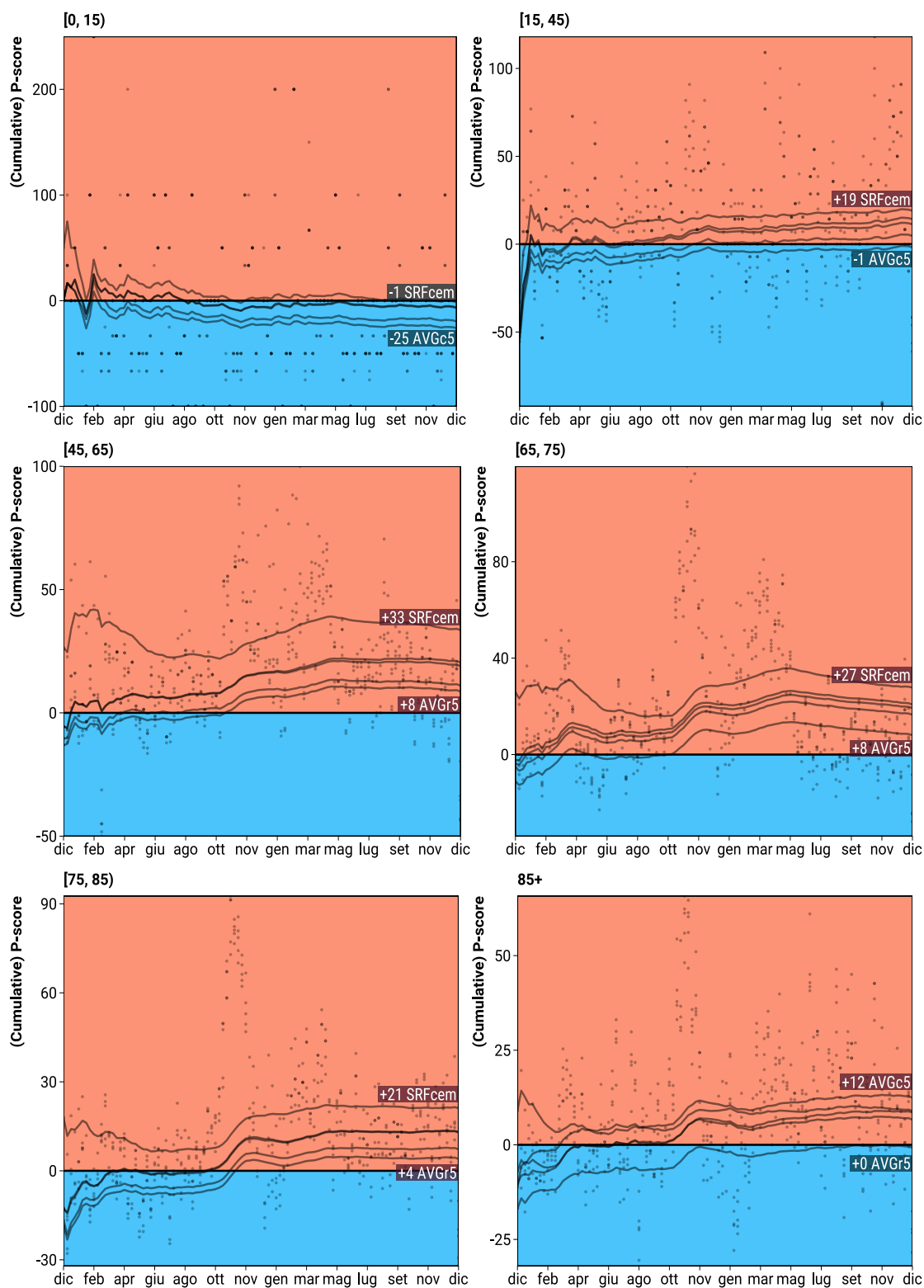


Figura B.11: Popolazione maschile della Campania.

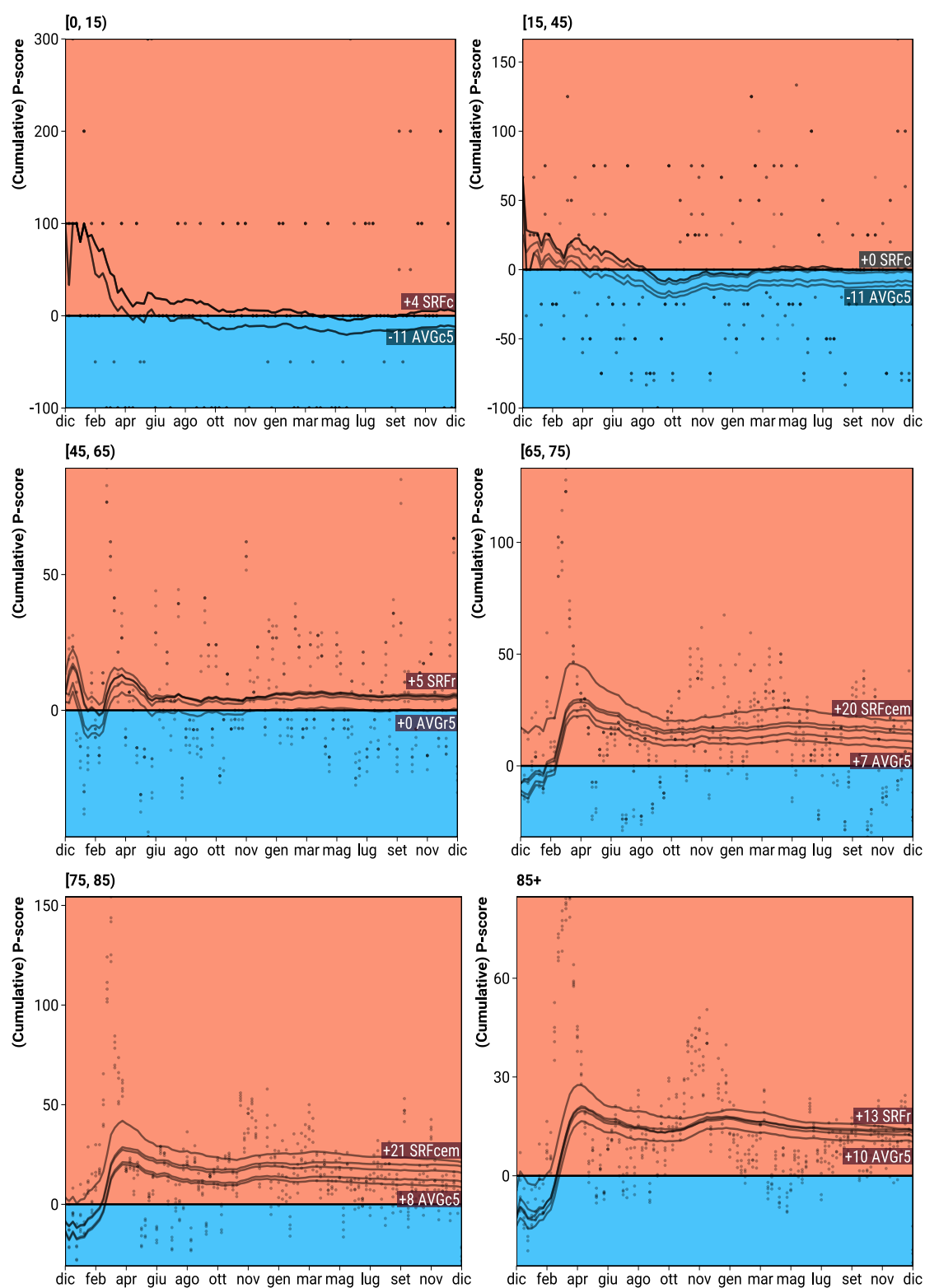


Figura B.12: Popolazione femminile dell'Emilia-Romagna.

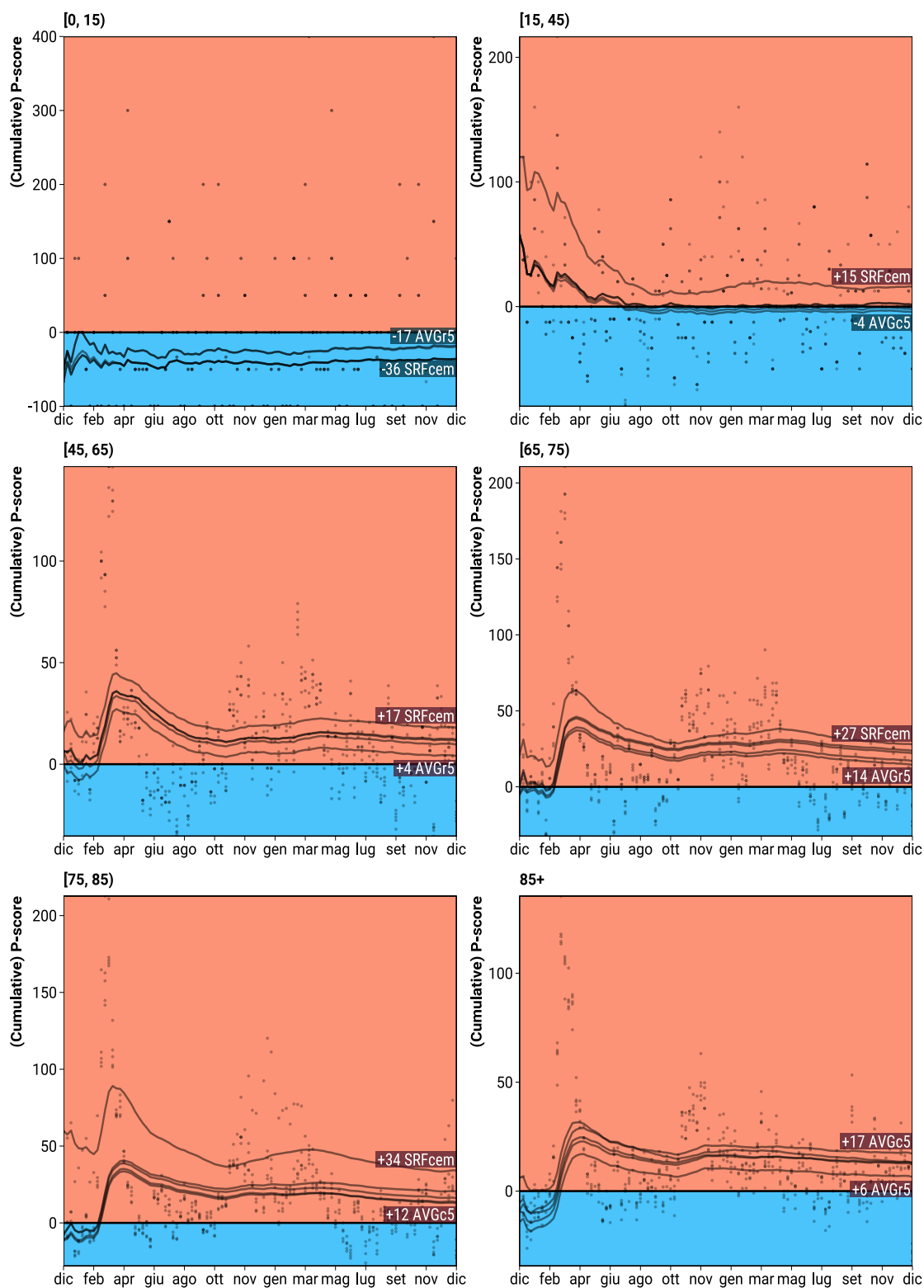


Figura B.13: Popolazione maschile dell'Emilia Romagna.

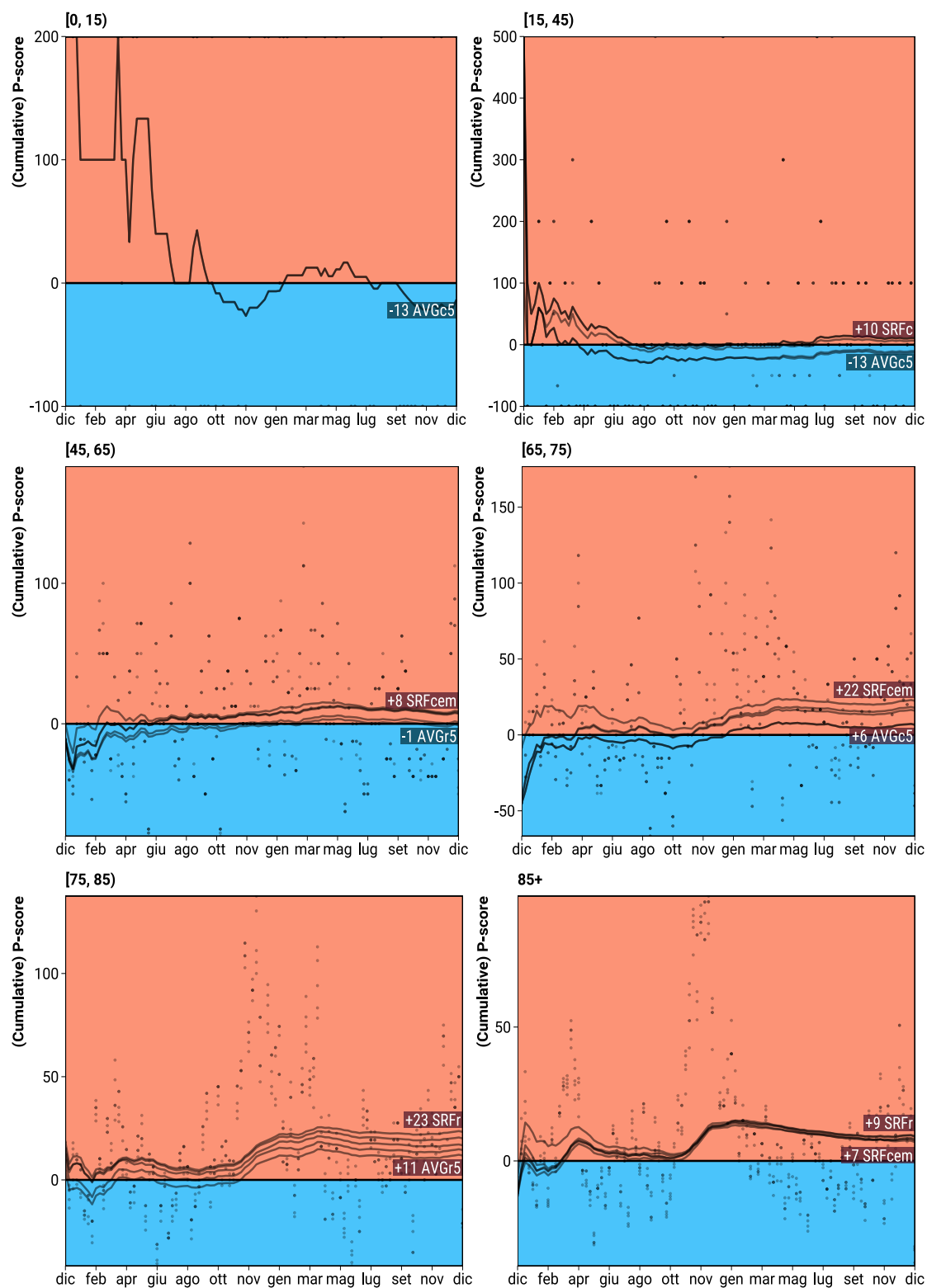


Figura B.14: Popolazione femminile del Friuli-Venezia Giulia.

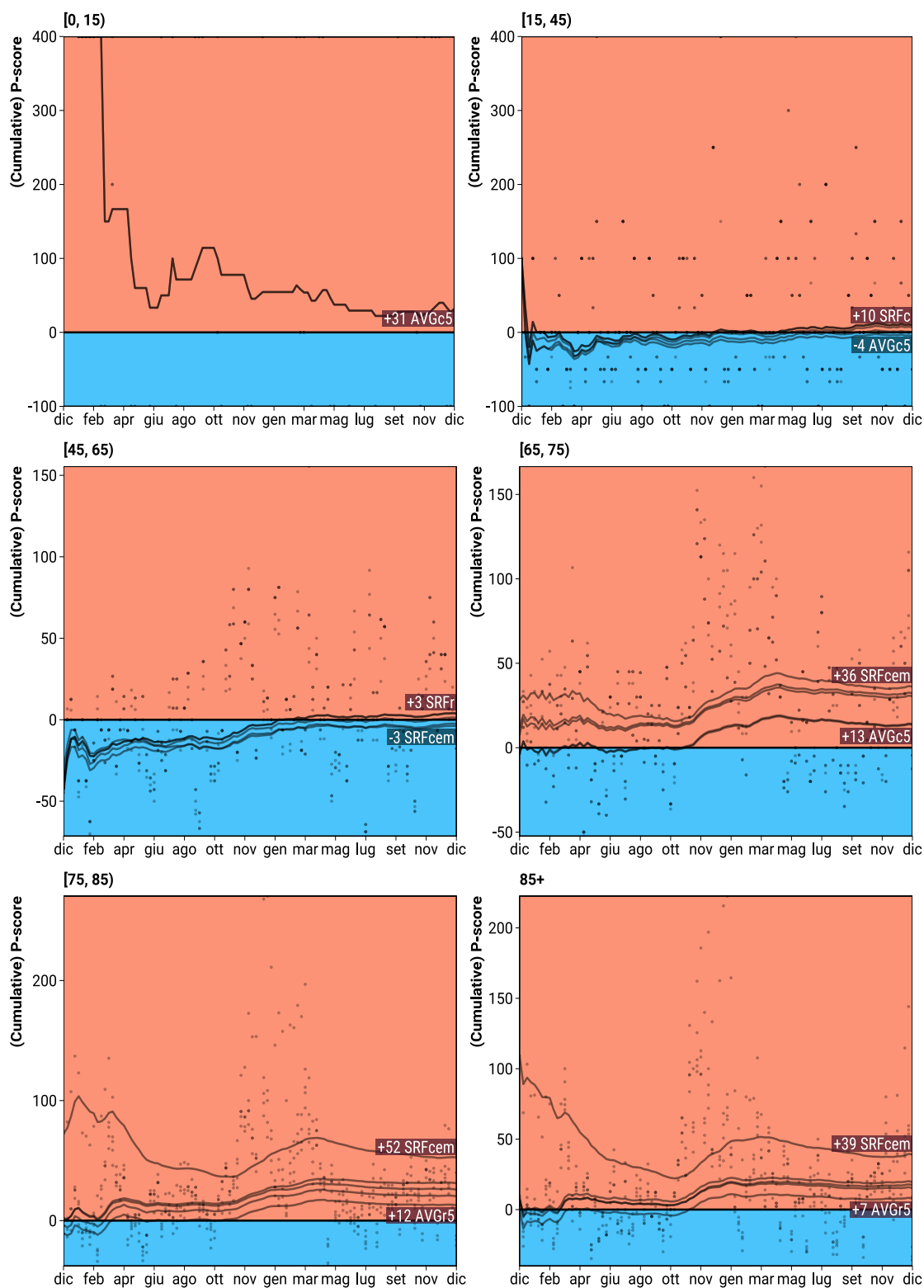


Figura B.15: Popolazione maschile del Friuli-Venezia Giulia.

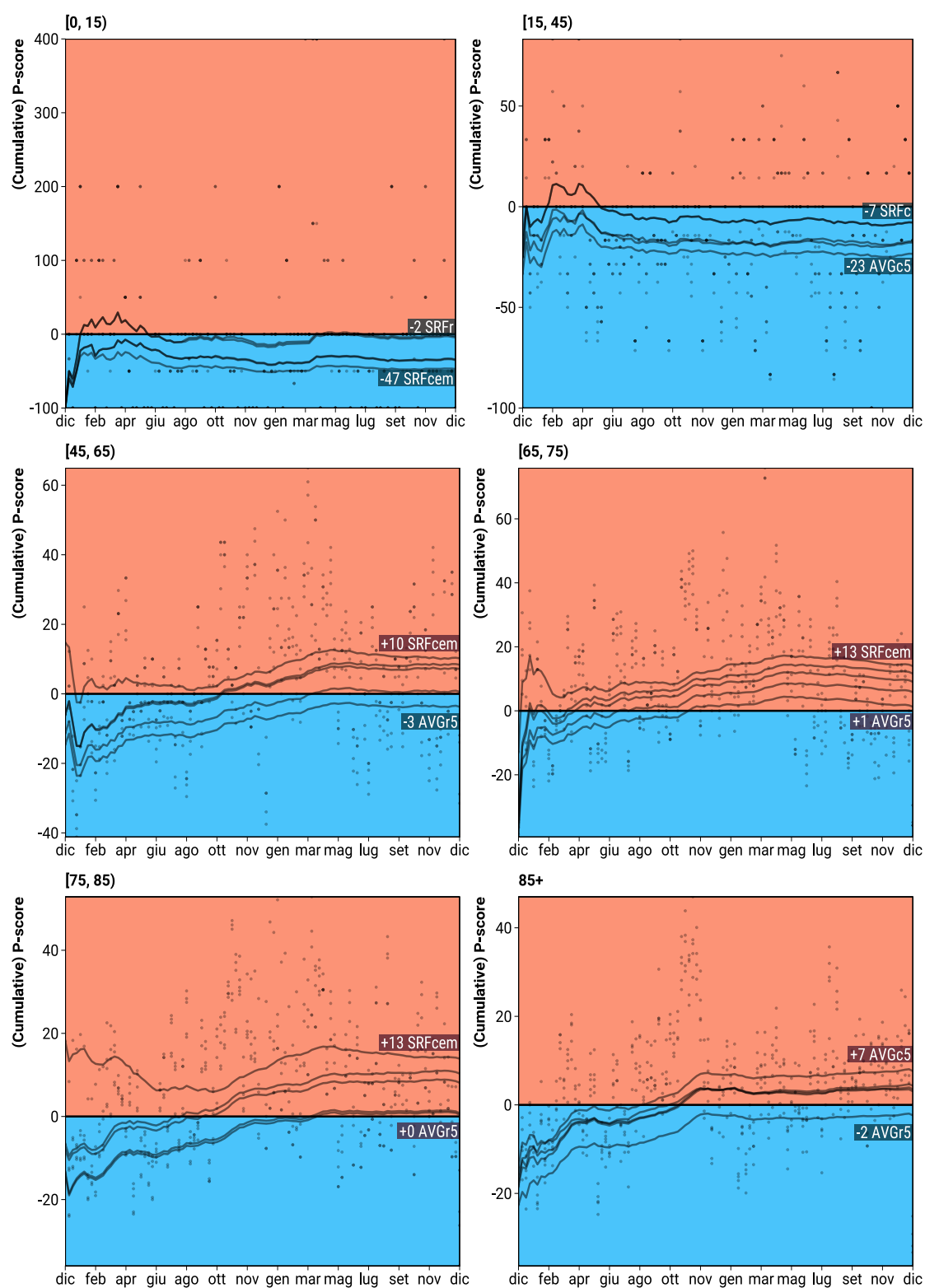


Figura B.16: Popolazione femminile del Lazio.

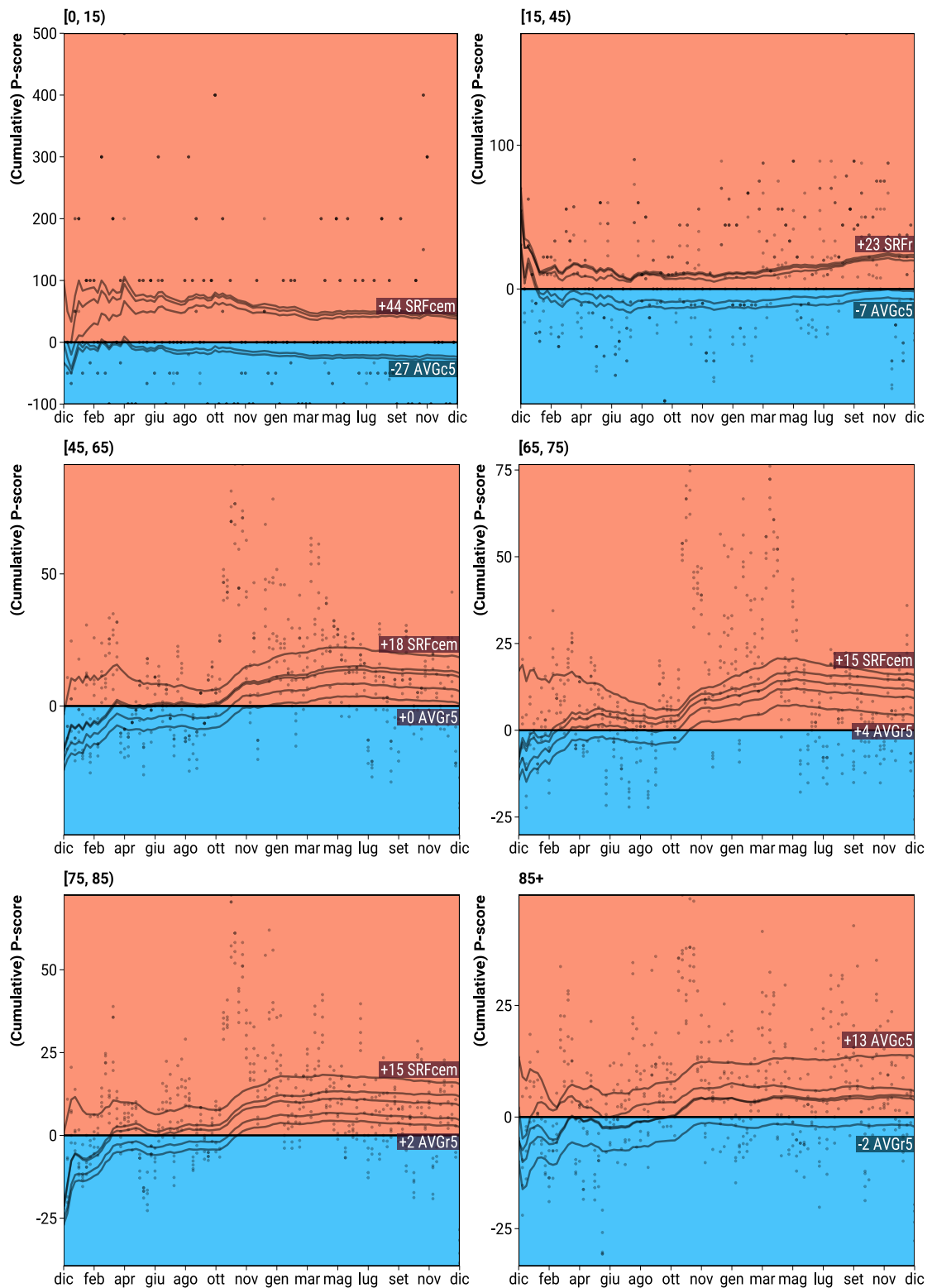


Figura B.17: Popolazione maschile del Lazio.

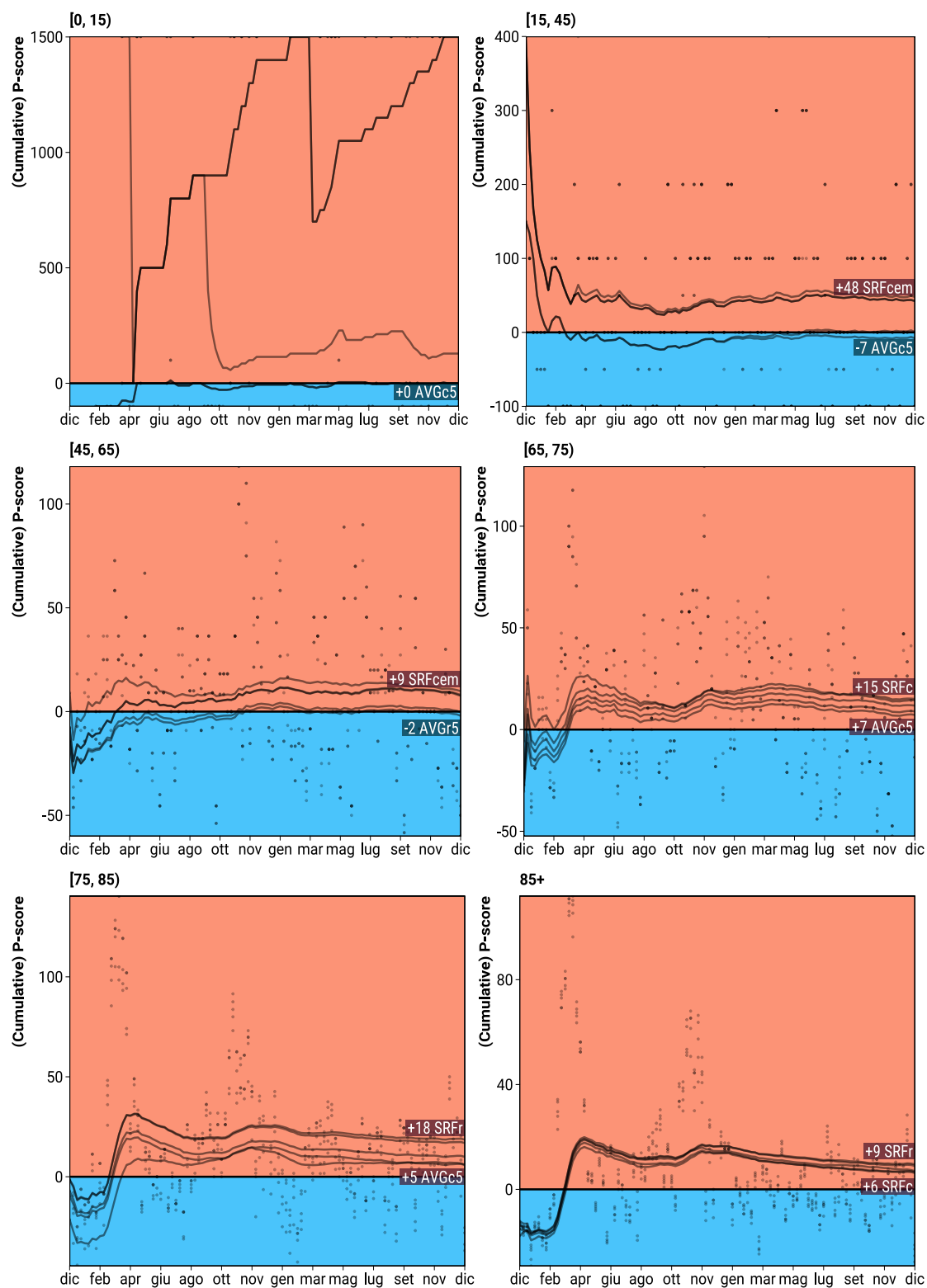


Figura B.18: Popolazione femminile della Liguria.

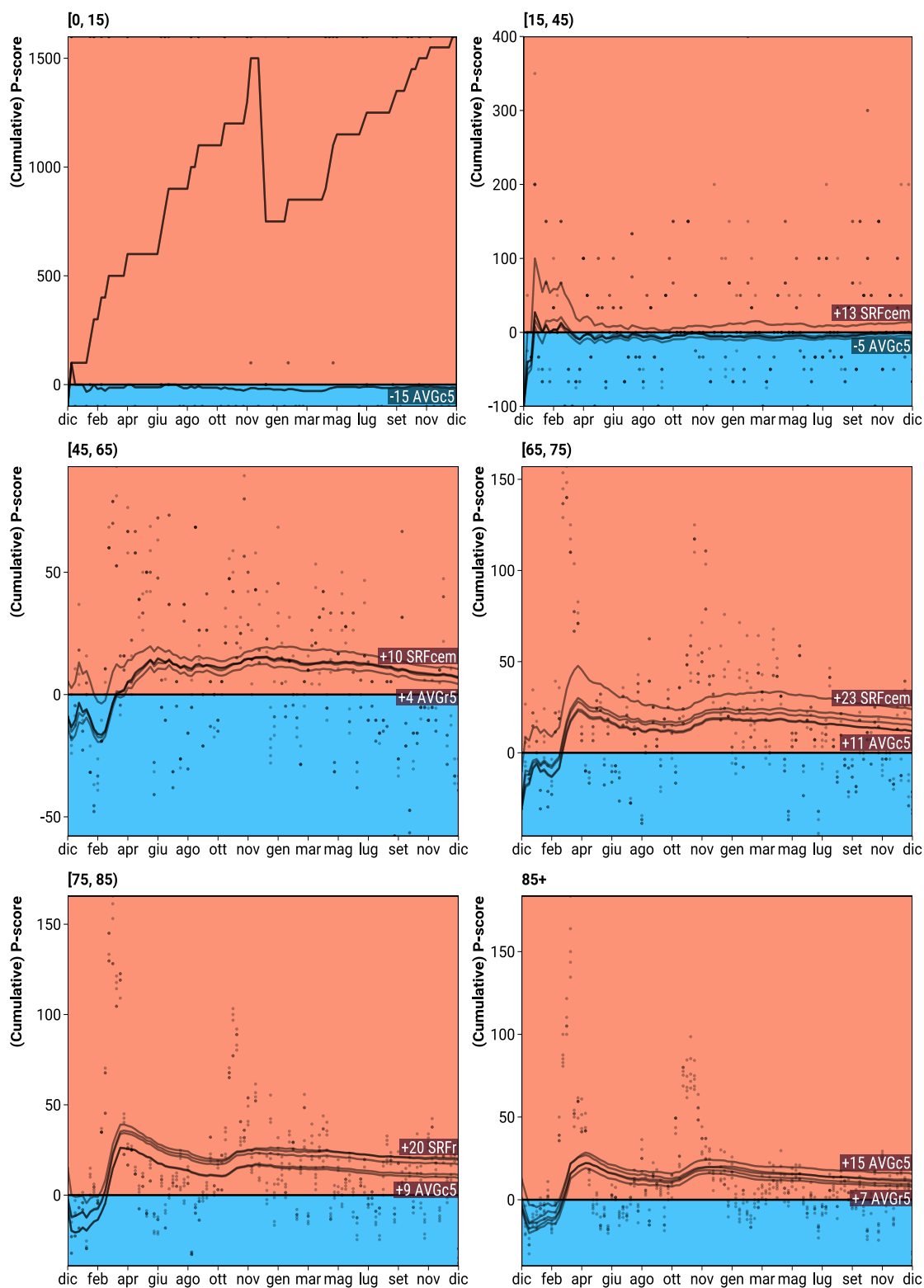


Figura B.19: Popolazione maschile della Liguria.

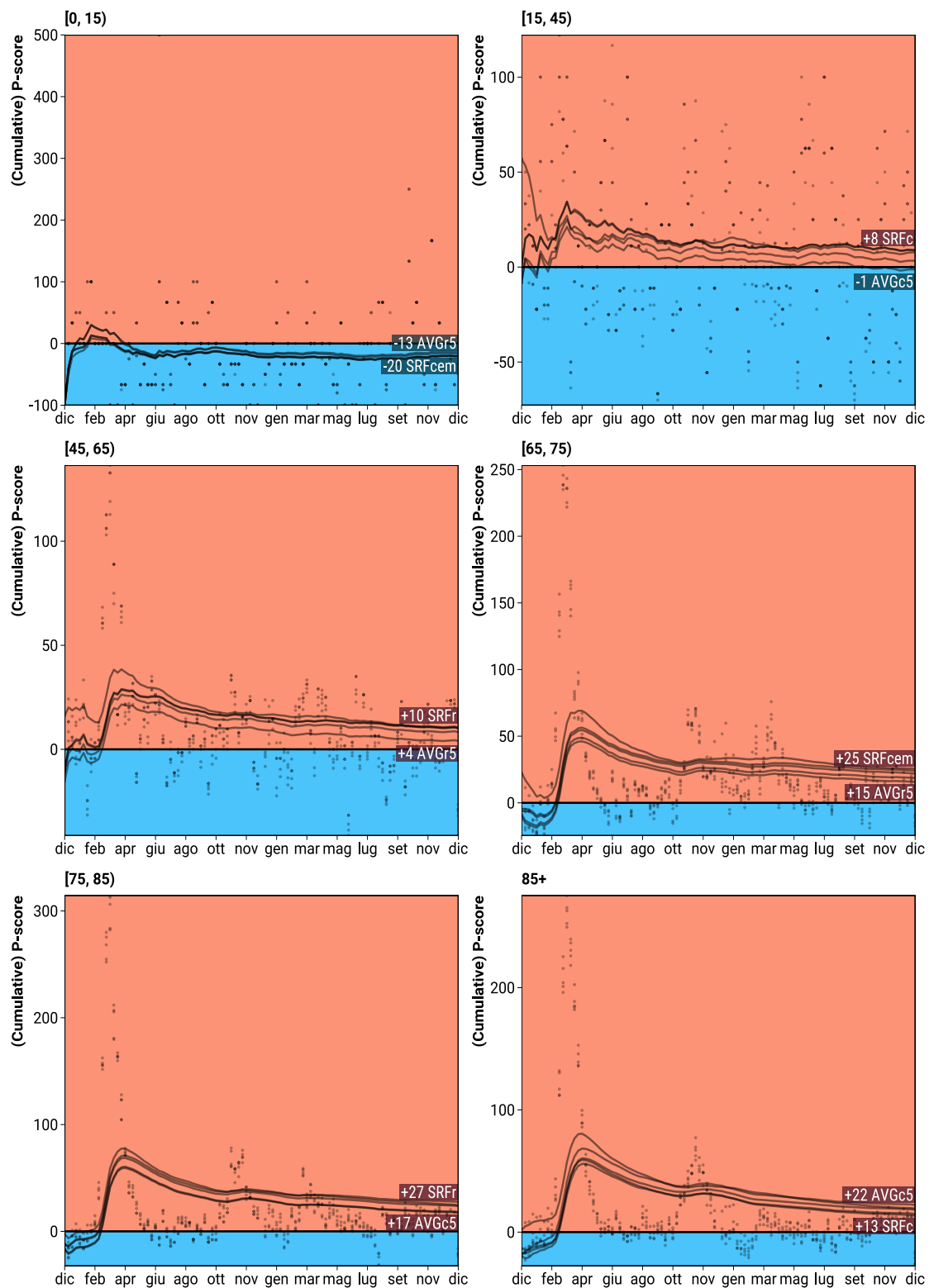


Figura B.20: Popolazione femminile della Lombardia.

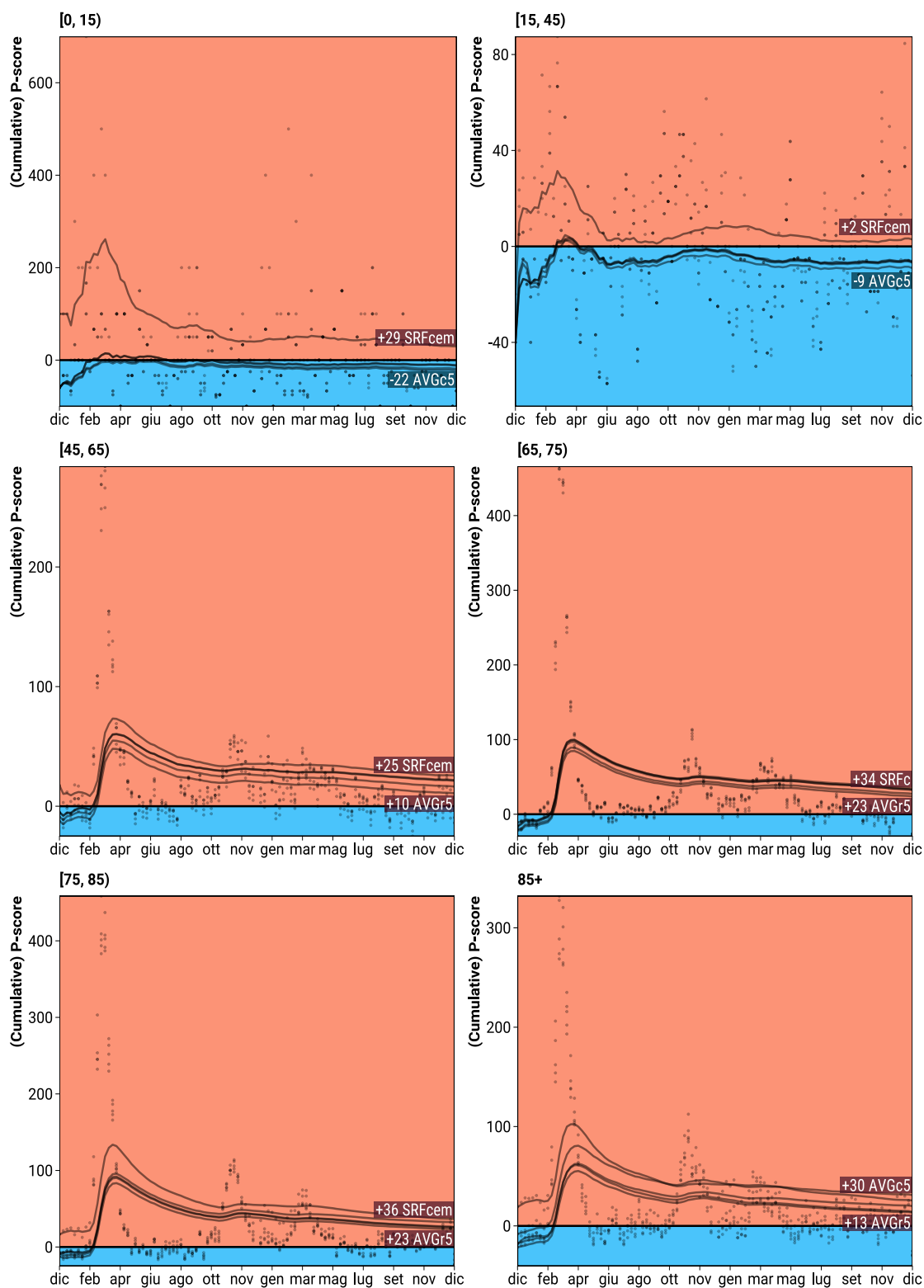


Figura B.21: Popolazione maschile della Lombardia.

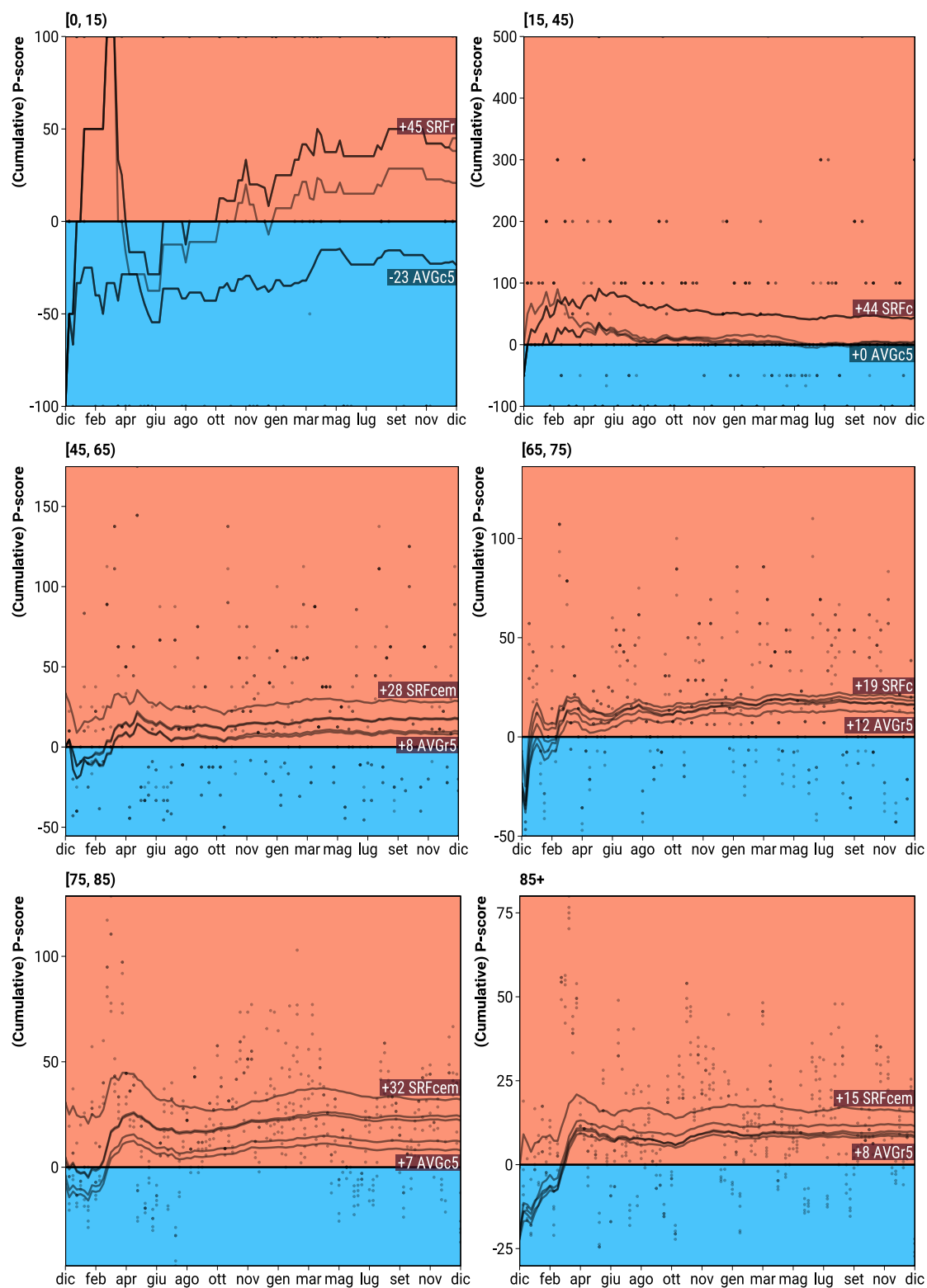


Figura B.22: Popolazione femminile delle Marche.

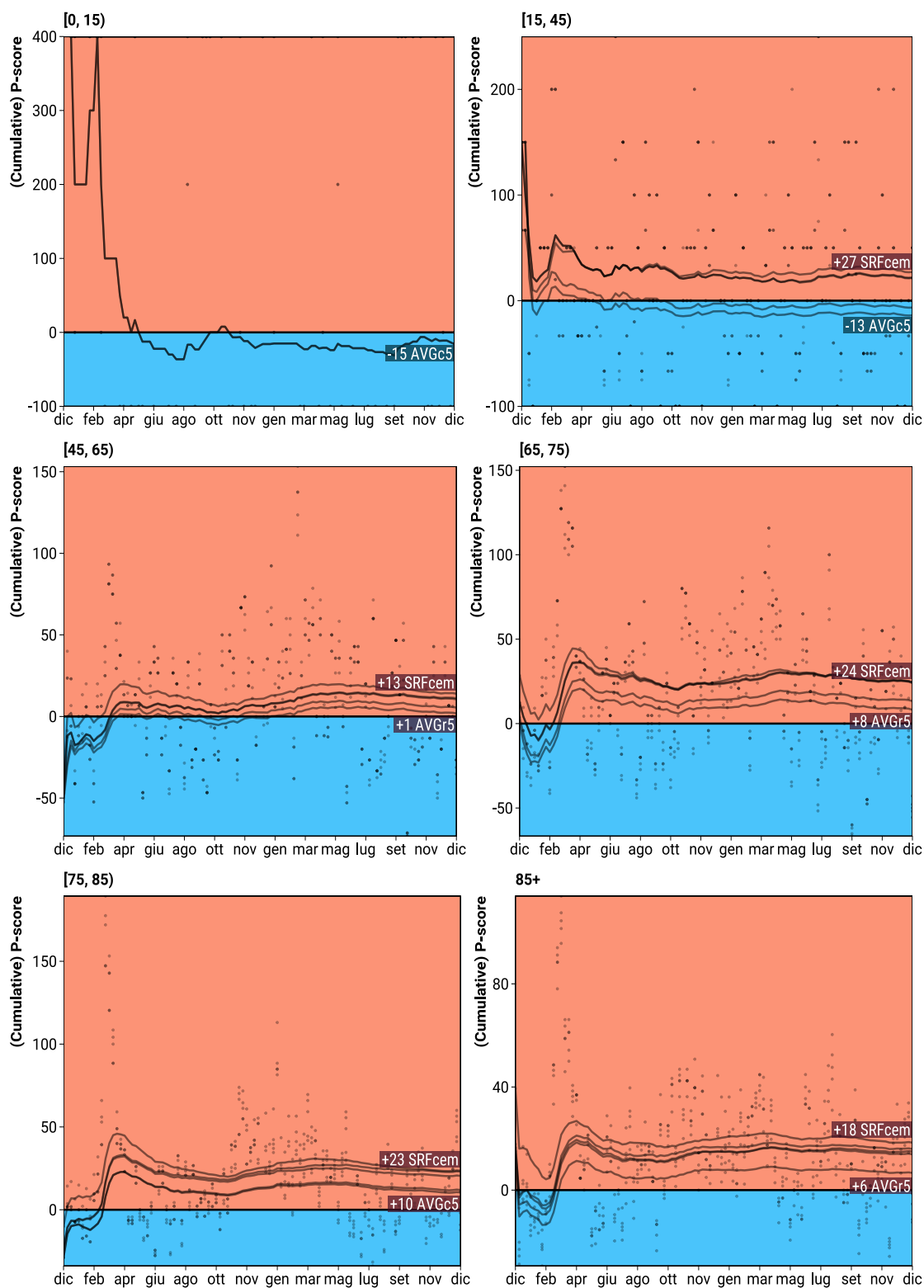


Figura B.23: Popolazione maschile delle Marche.

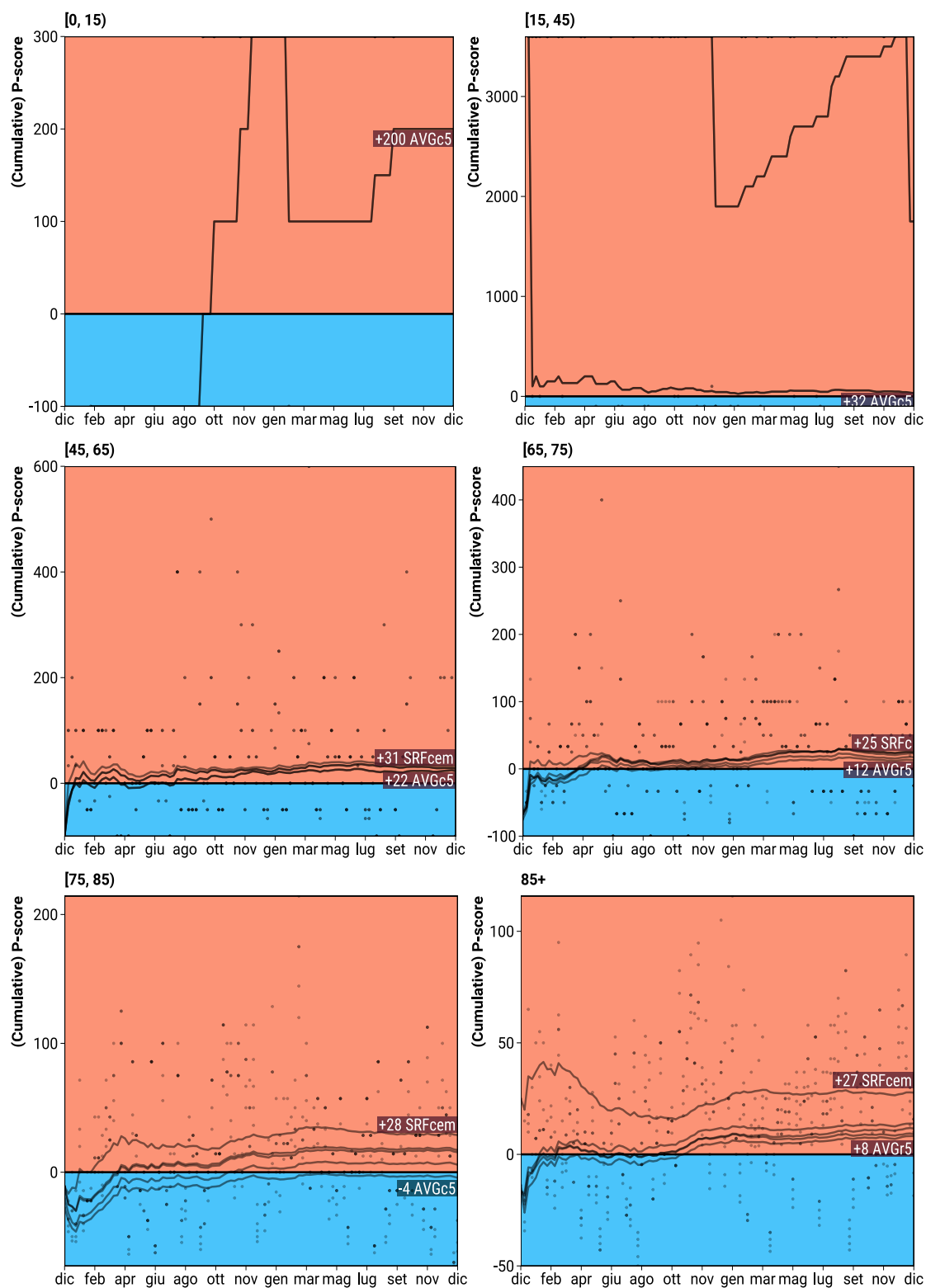


Figura B.24: Popolazione femminile del Molise.

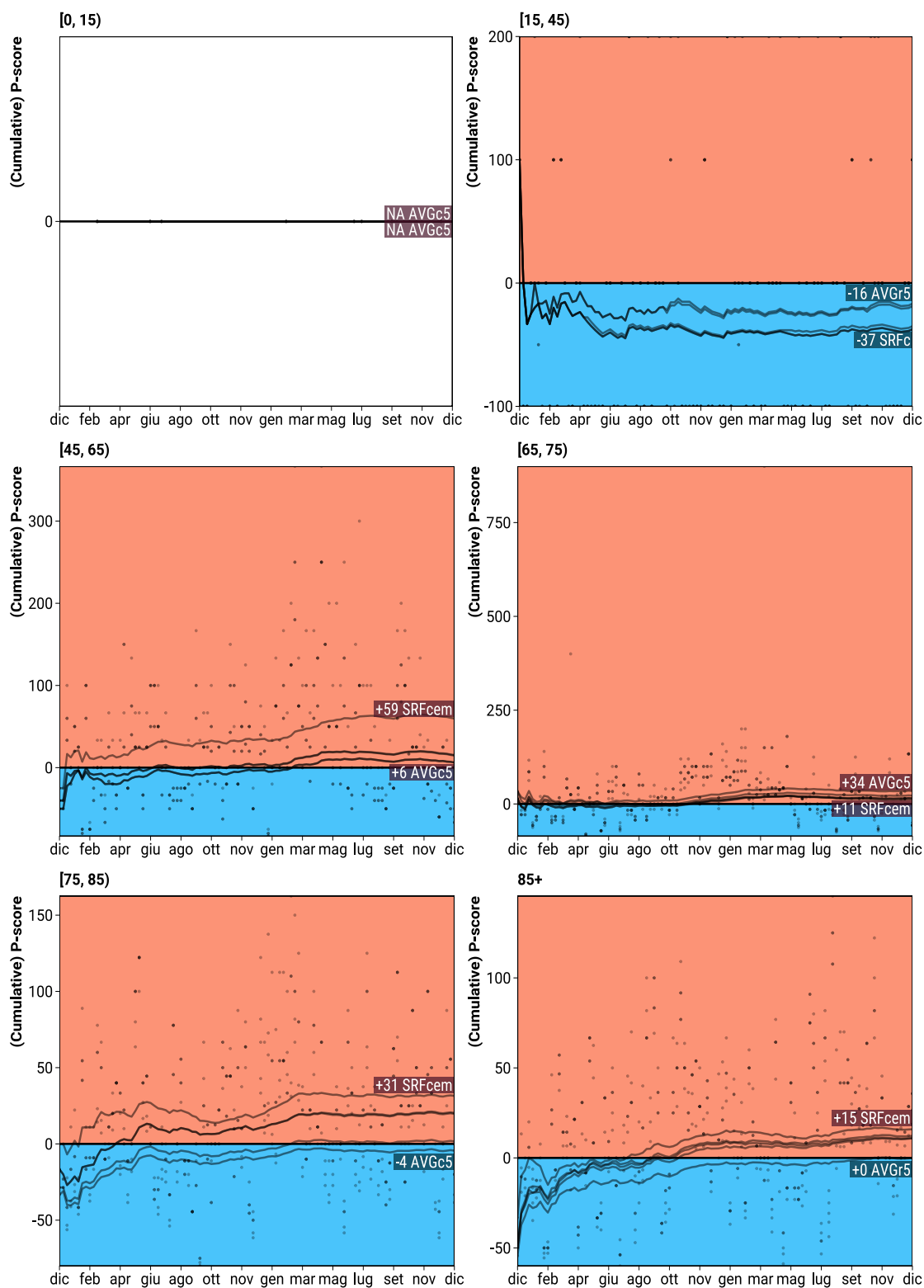


Figura B.25: Popolazione maschile del Molise.

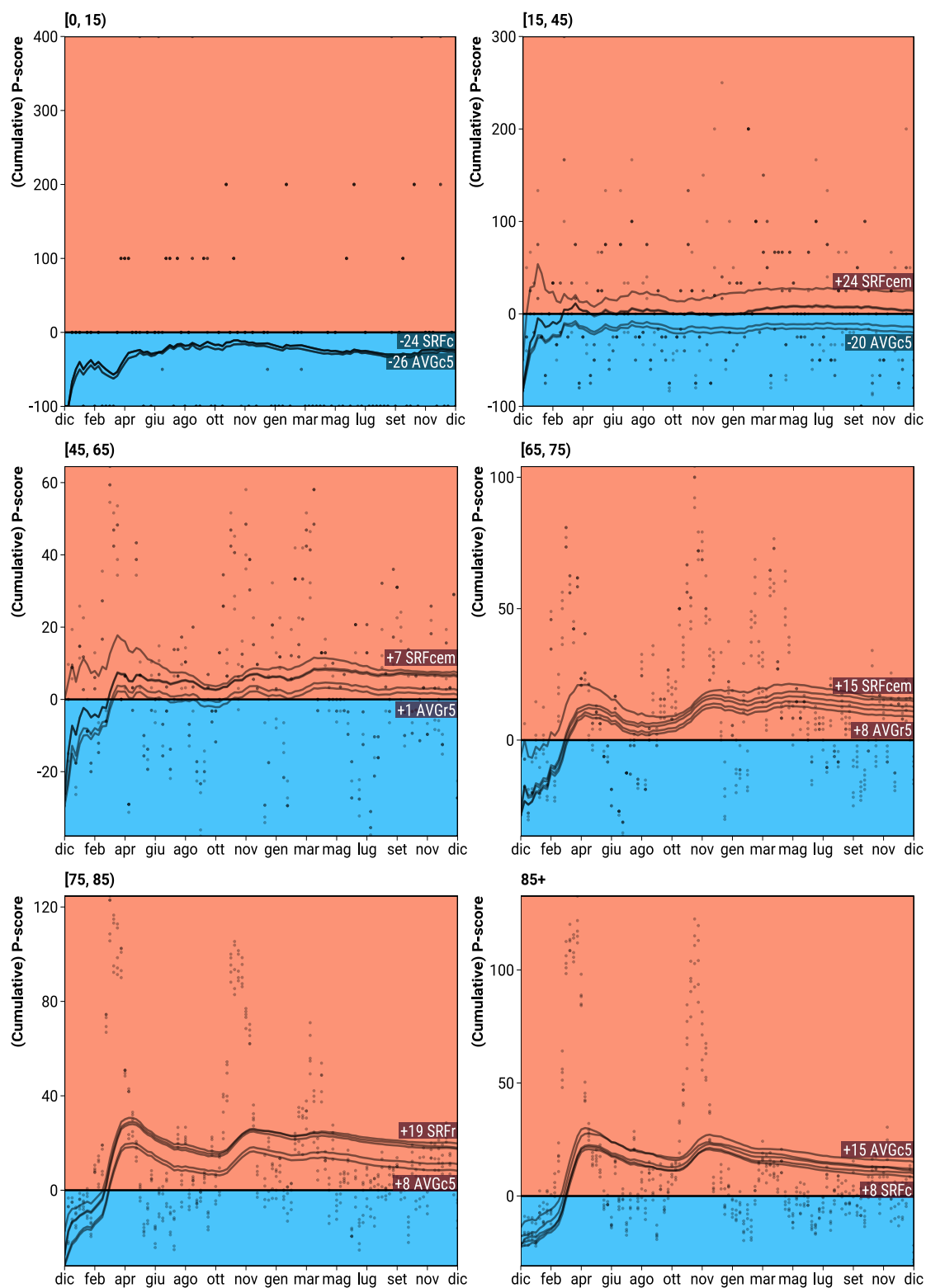


Figura B.26: Popolazione femminile del Piemonte.

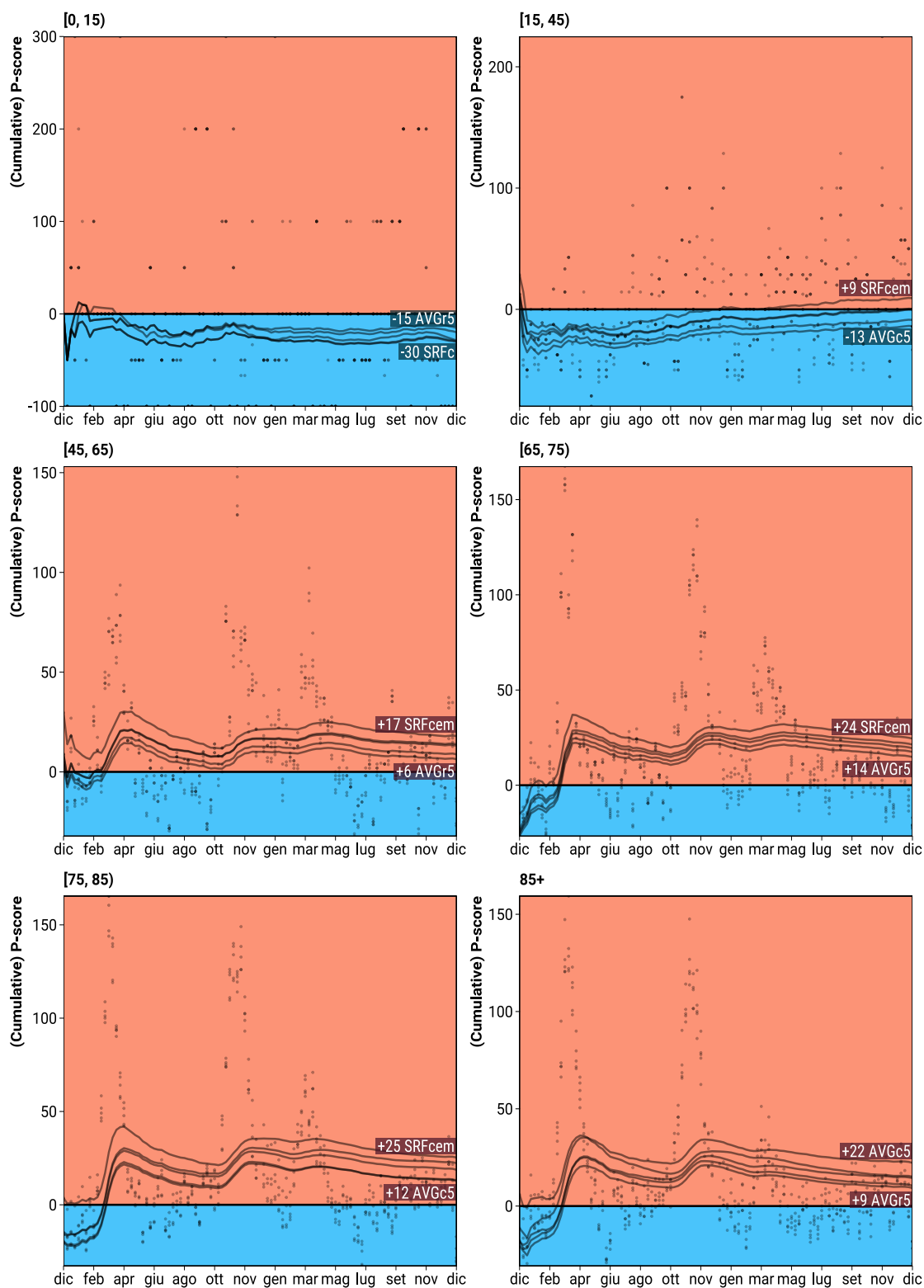


Figura B.27: Popolazione maschile del Piemonte.

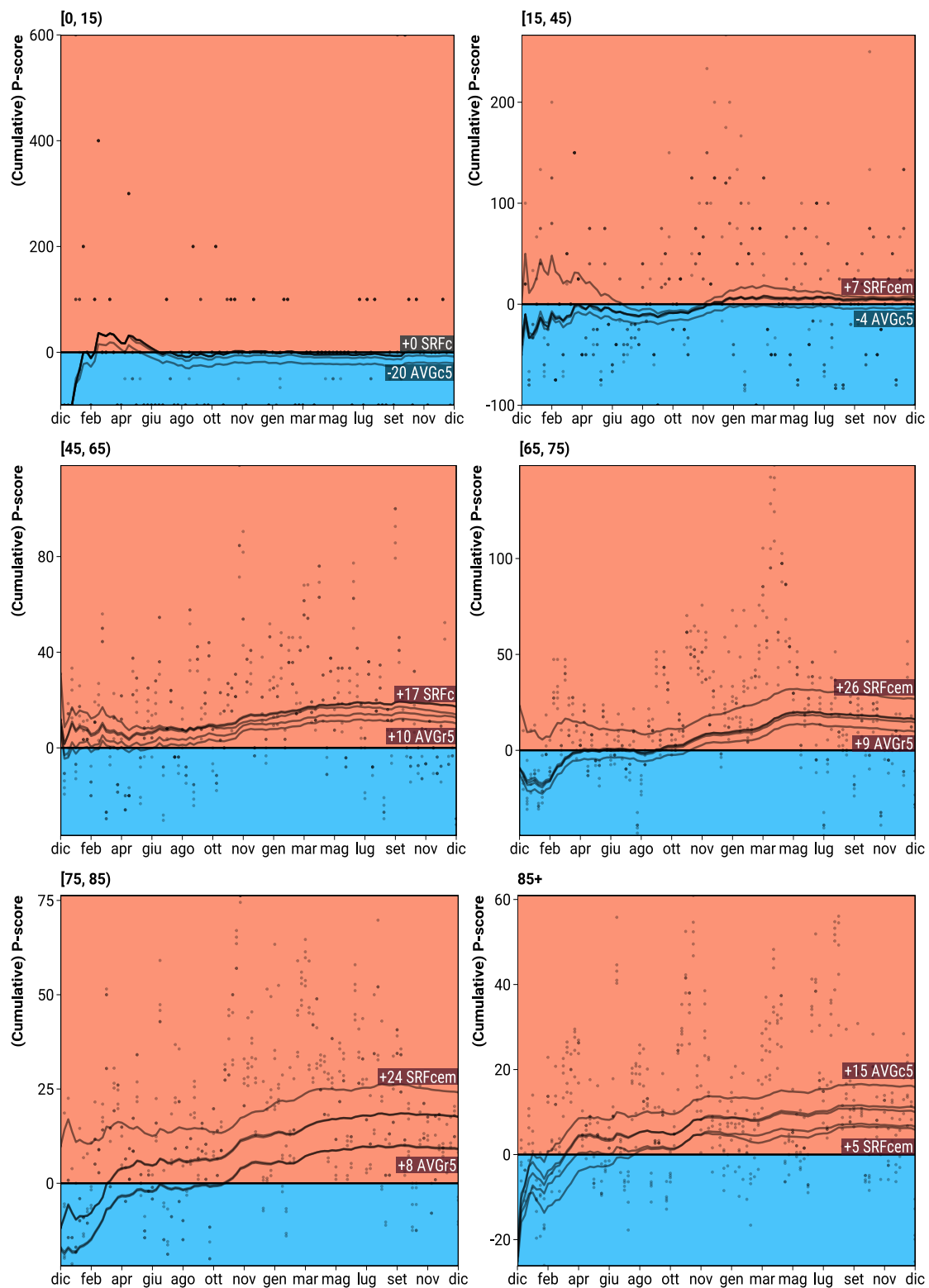


Figura B.28: Popolazione femminile della Puglia.

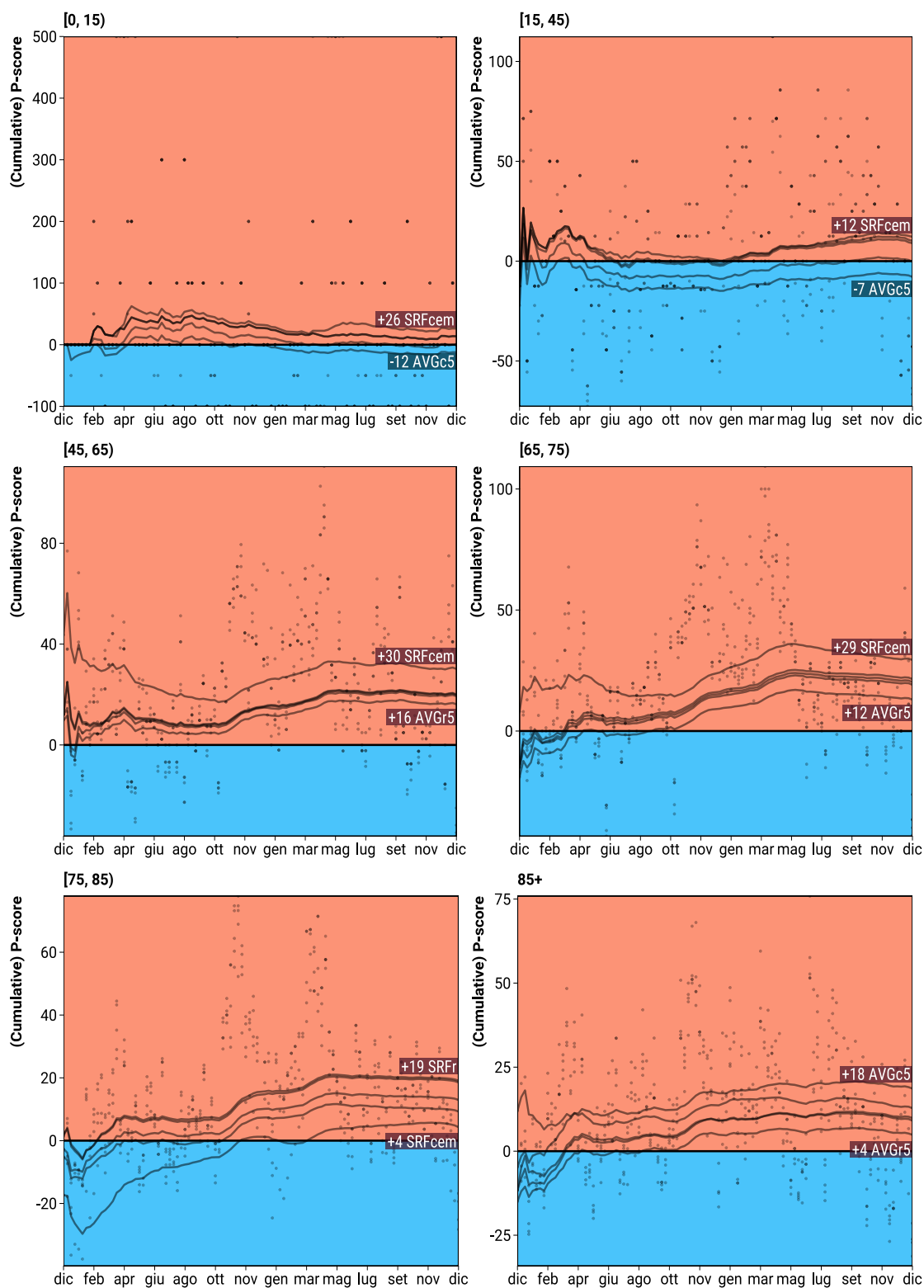


Figura B.29: Popolazione maschile della Puglia.

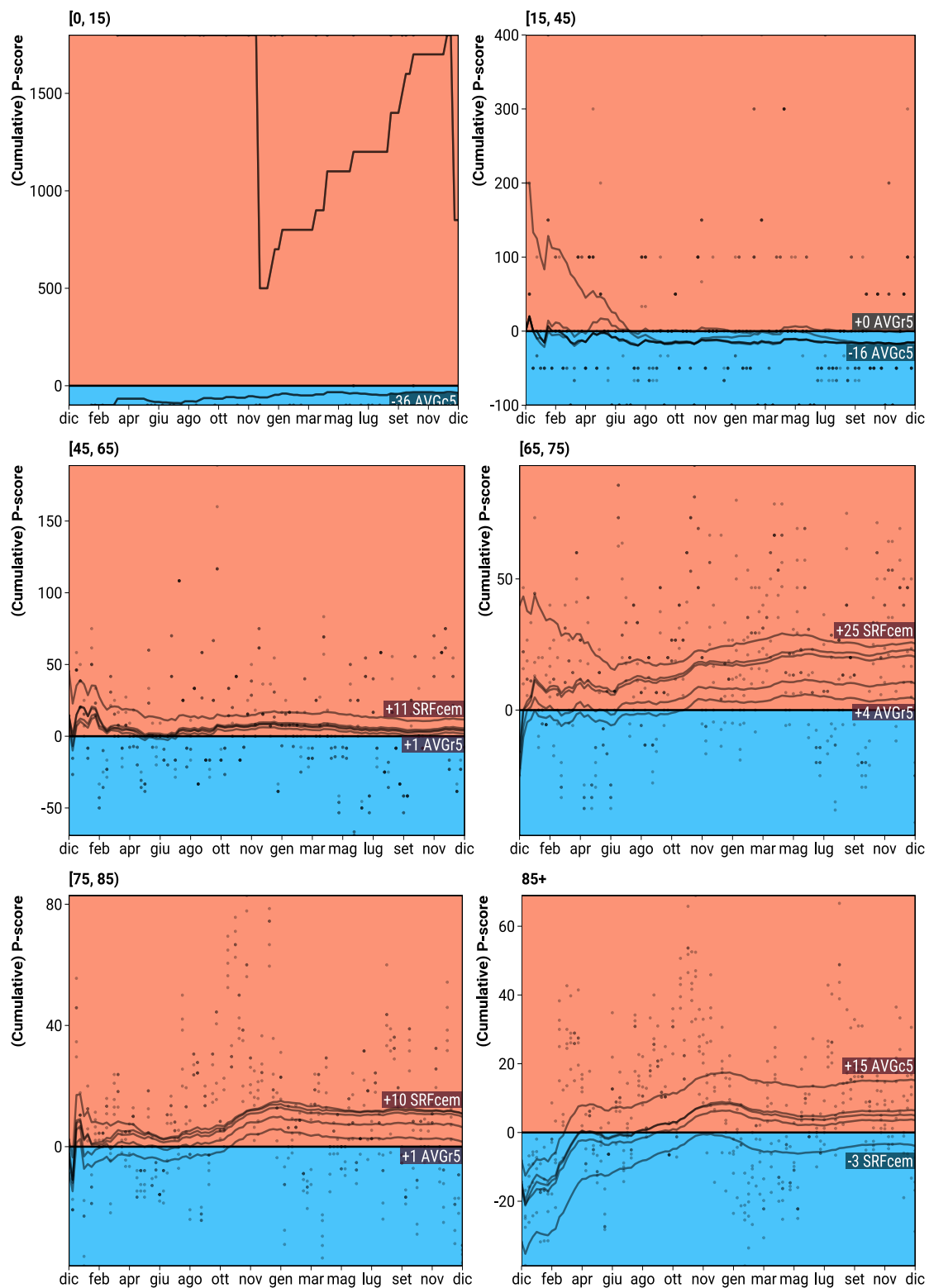


Figura B.30: Popolazione femminile della Sardegna.

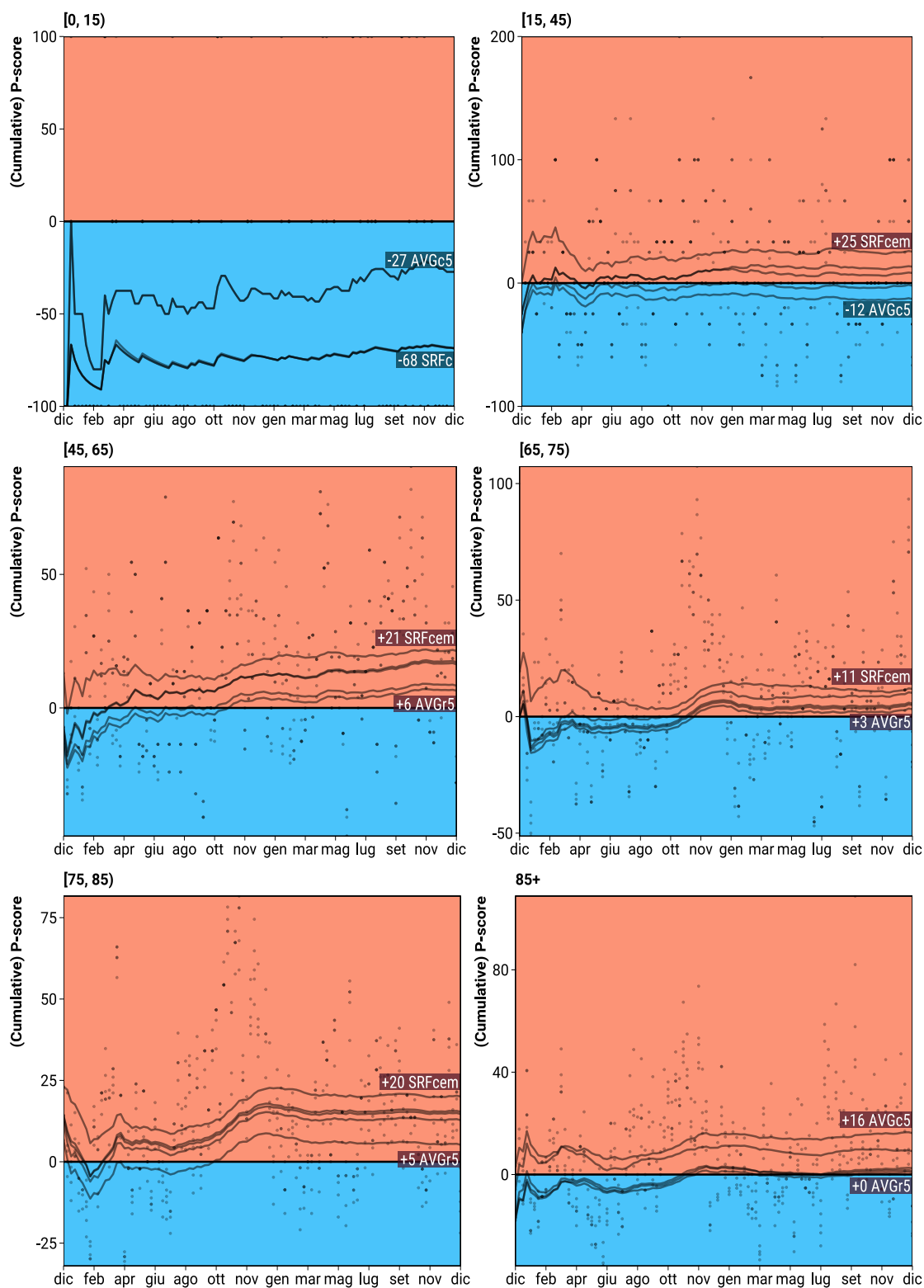


Figura B.31: Popolazione maschile della Sardegna.

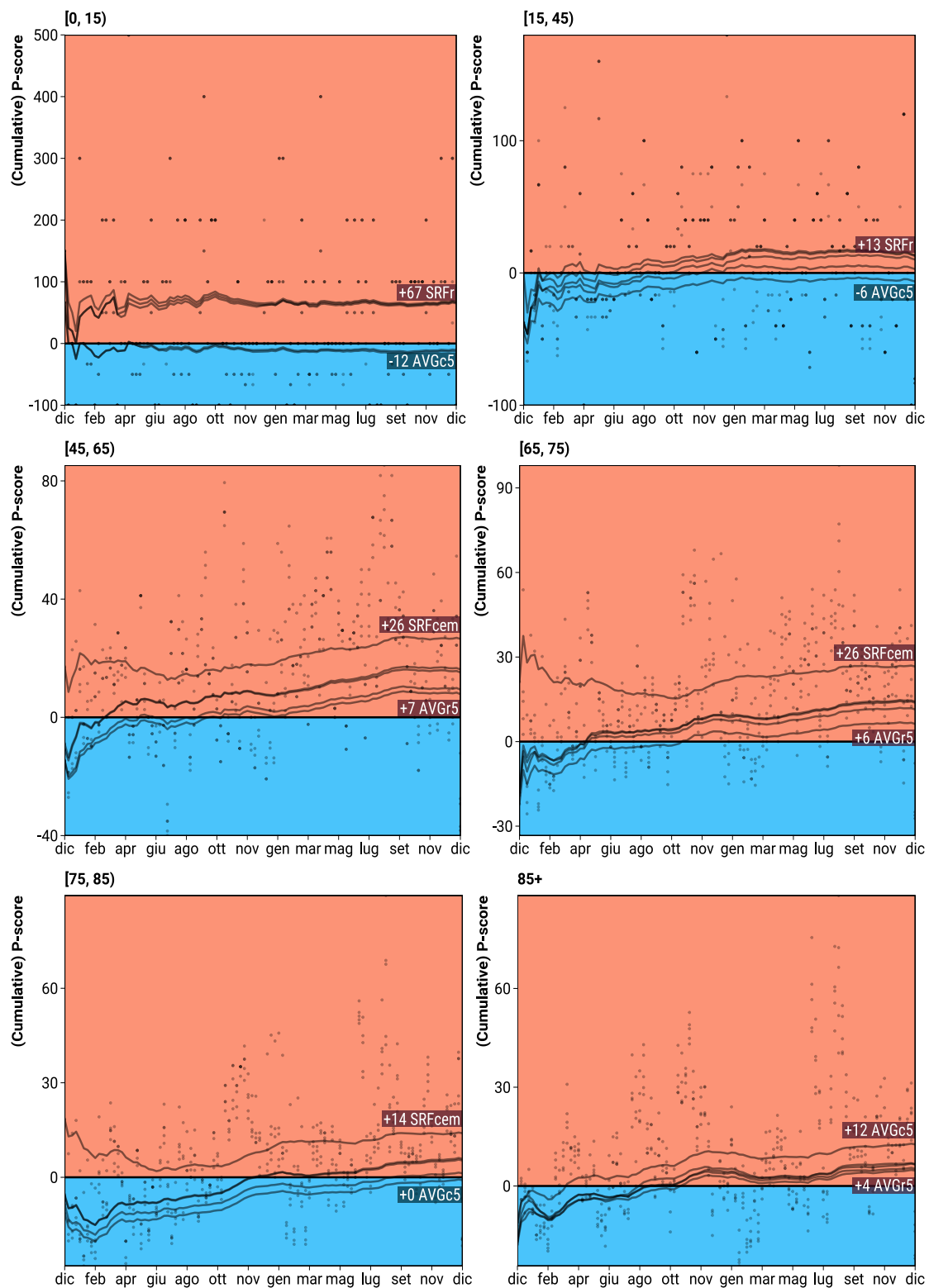


Figura B.32: Popolazione femminile della Sicilia.

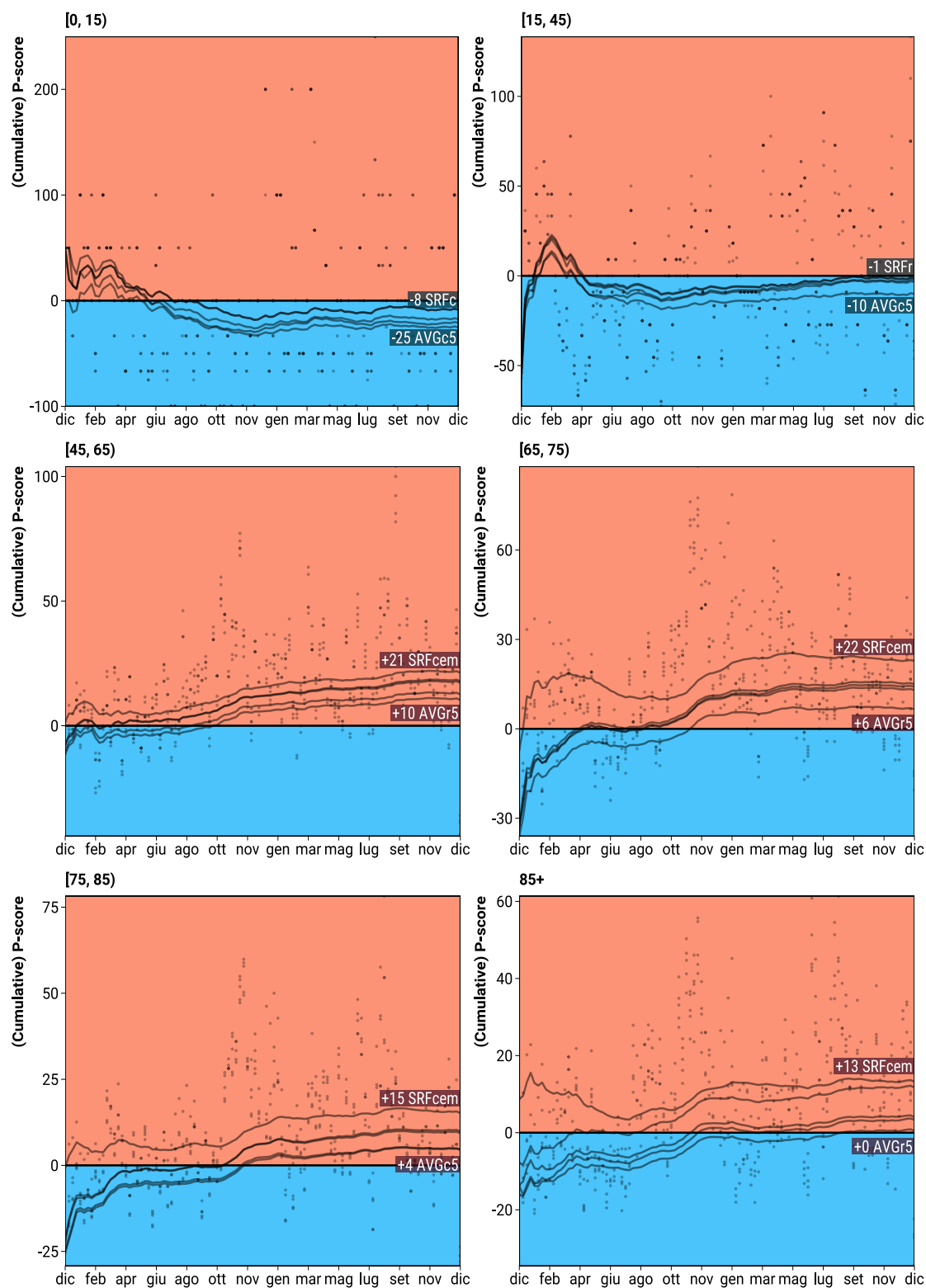


Figura B.33: Popolazione maschile della Sicilia.

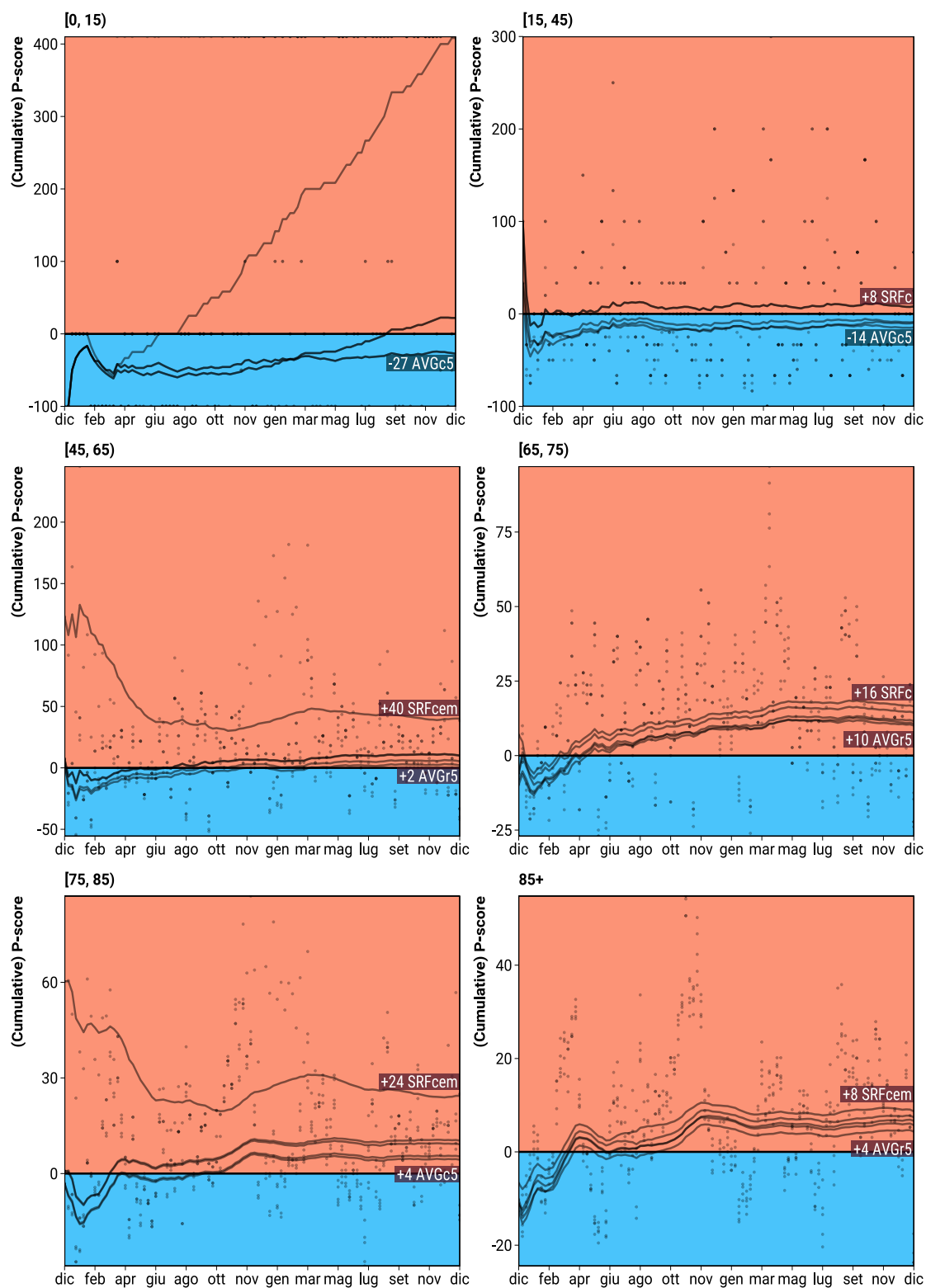


Figura B.34: Popolazione femminile della Toscana.

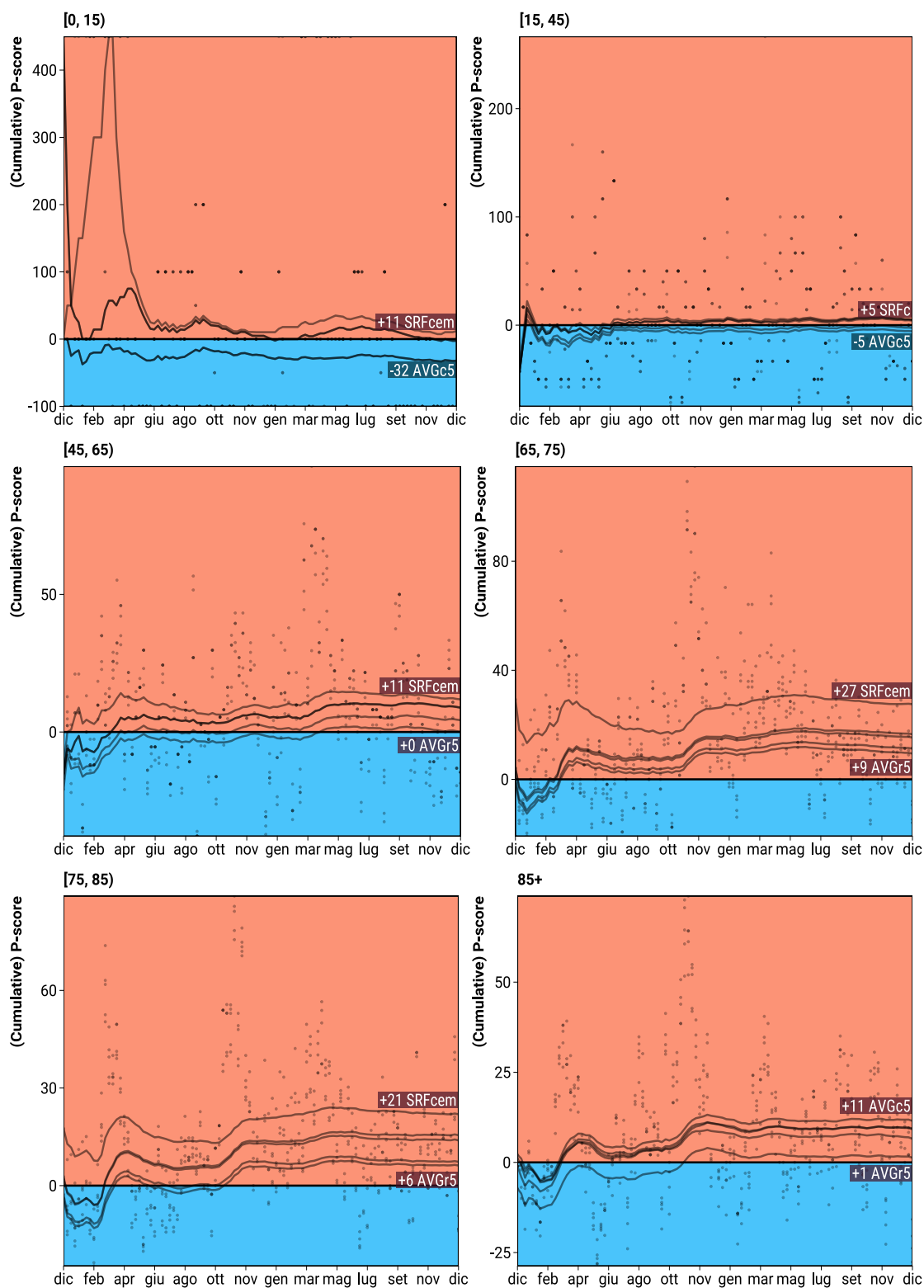


Figura B.35: Popolazione maschile della Toscana.

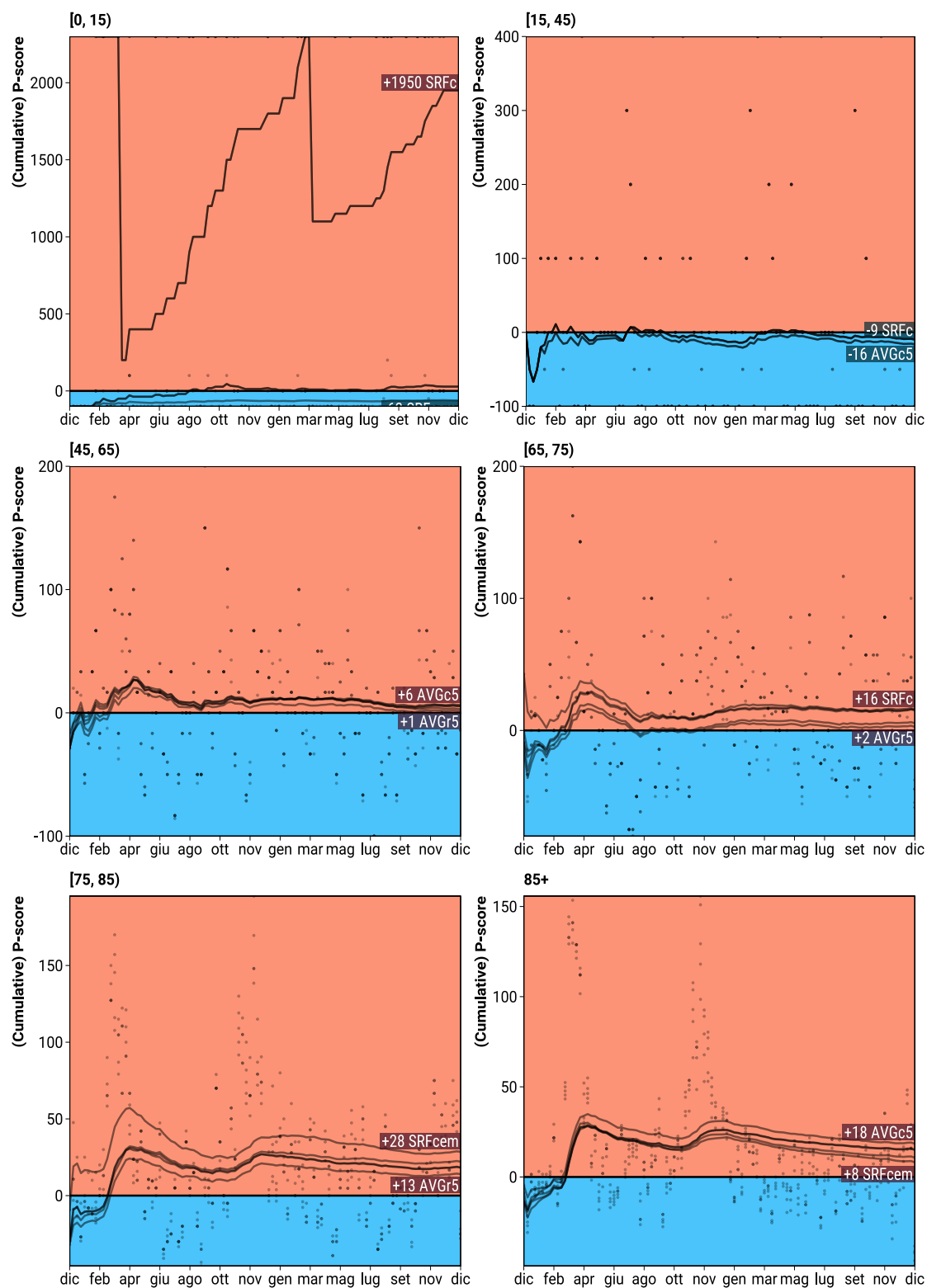


Figura B.36: Popolazione femminile del Trentino-Alto Adige.

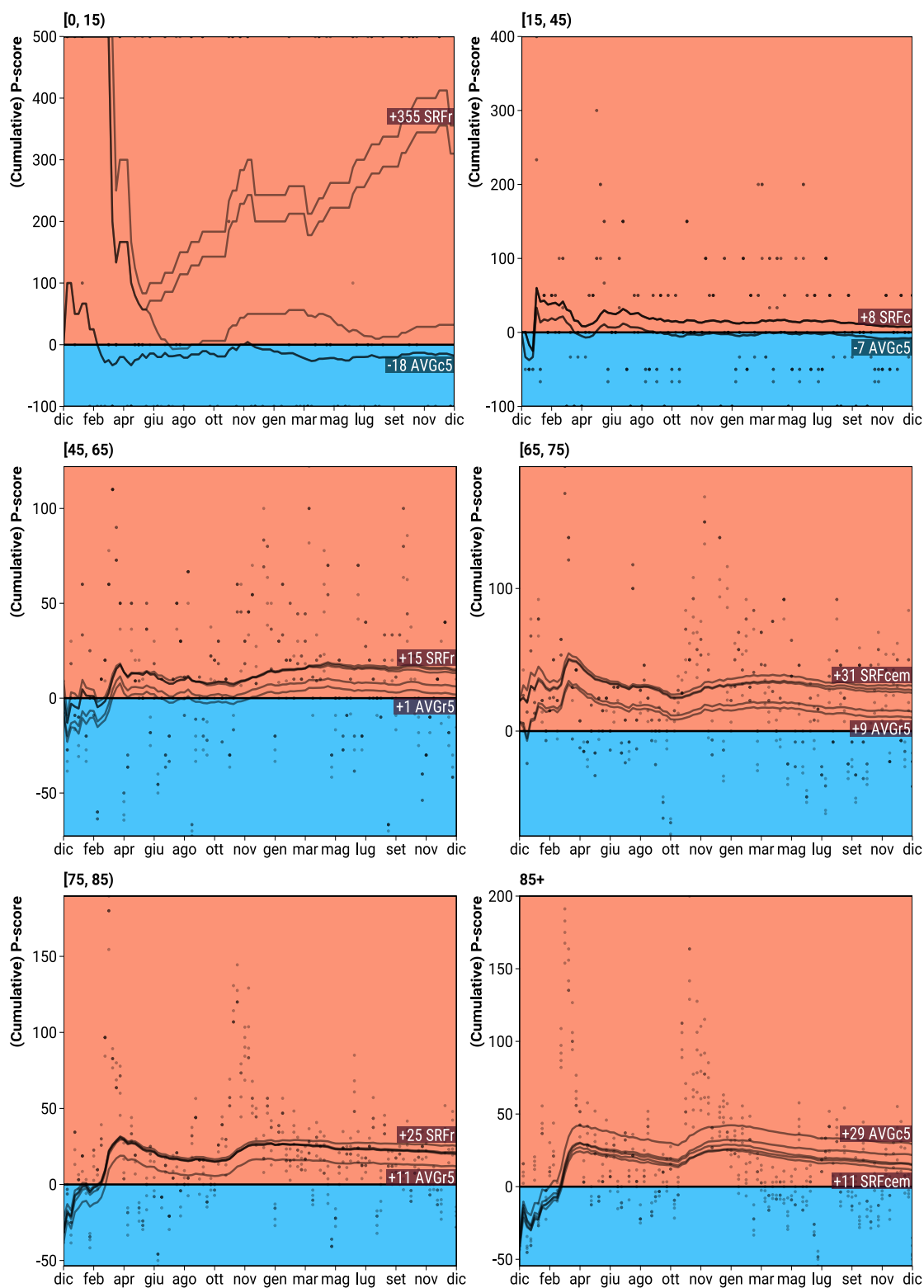


Figura B.37: Popolazione maschile del Trentino-Alto Adige.

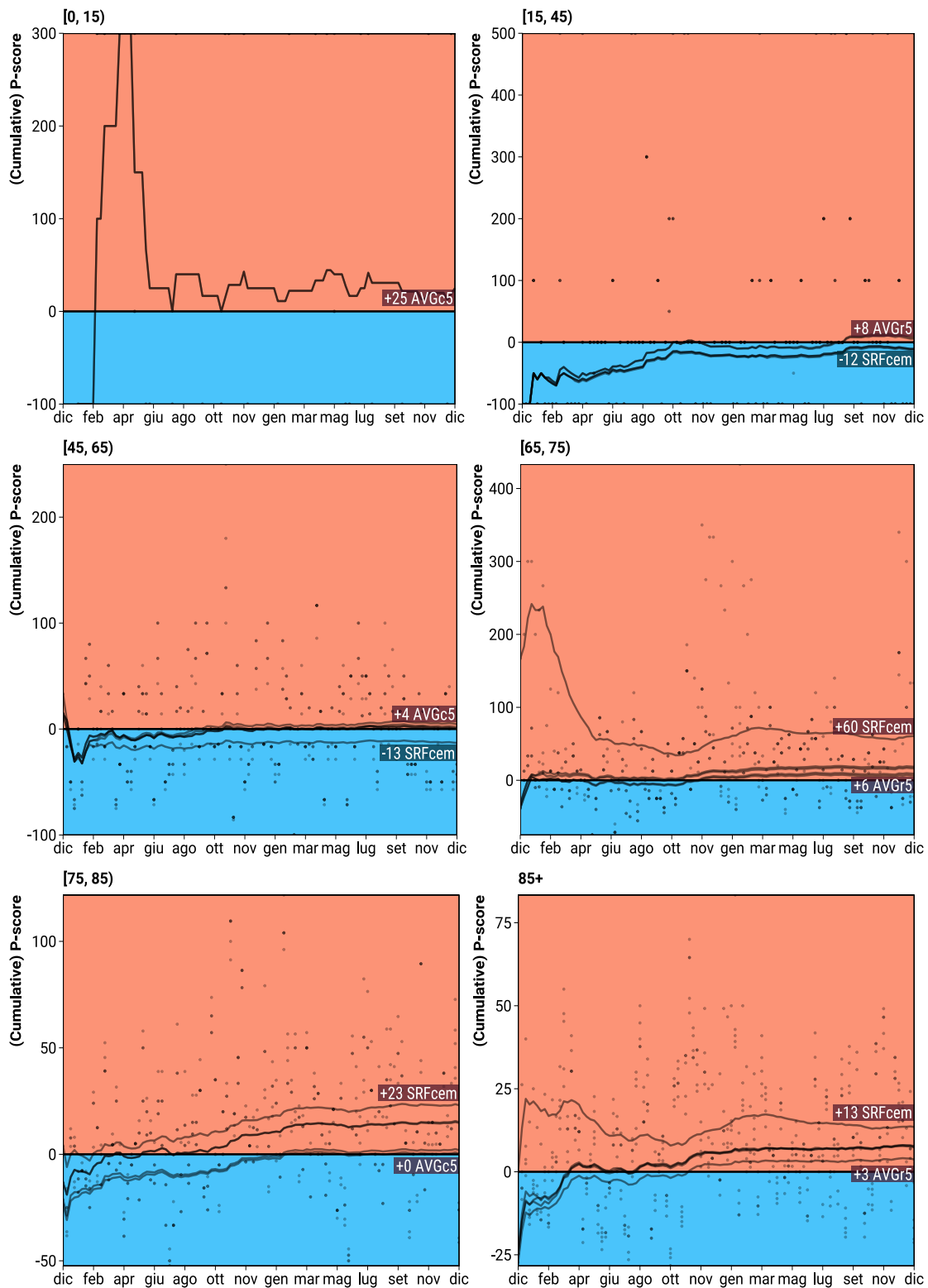


Figura B.38: Popolazione femminile dell'Umbria.

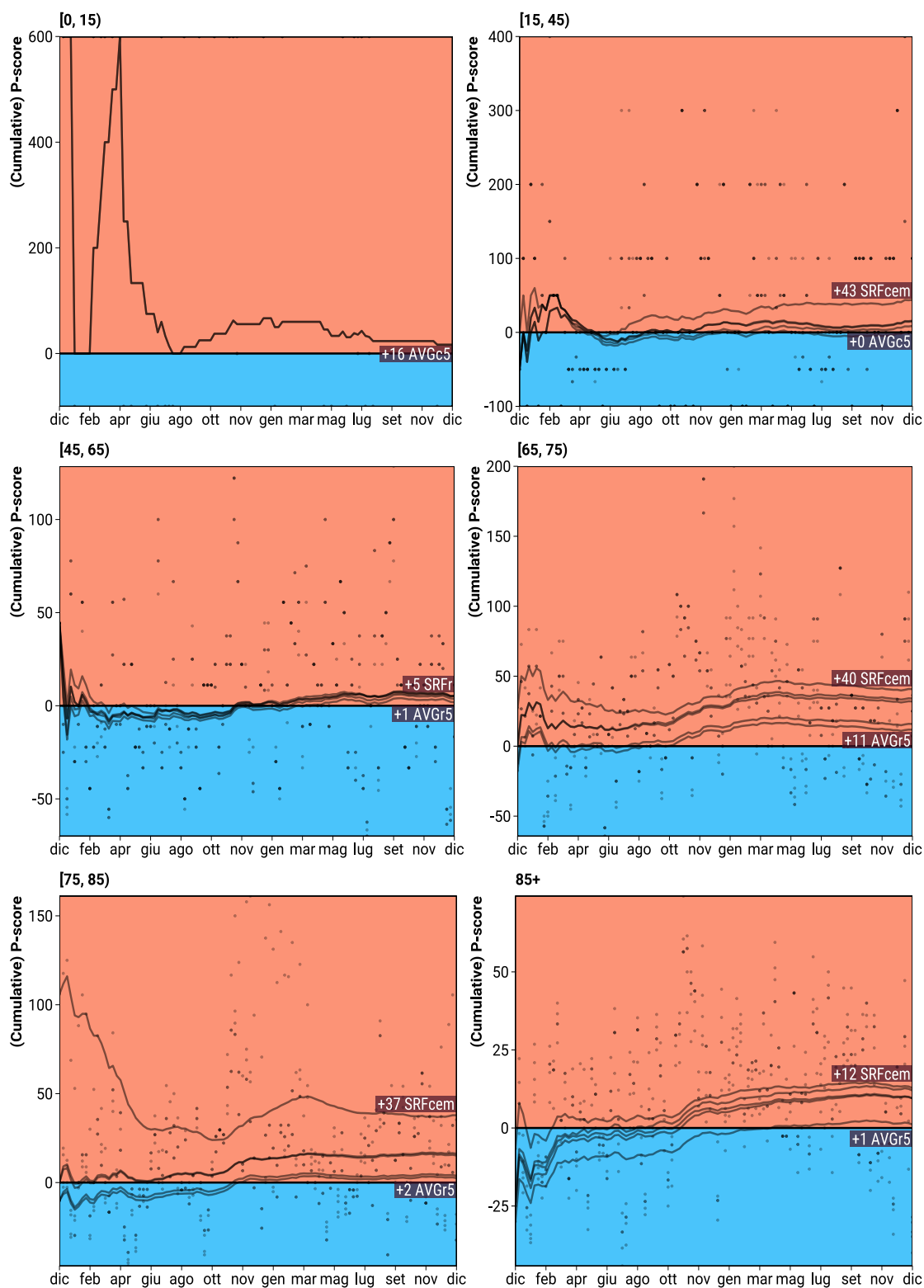


Figura B.39: Popolazione maschile dell'Umbria.

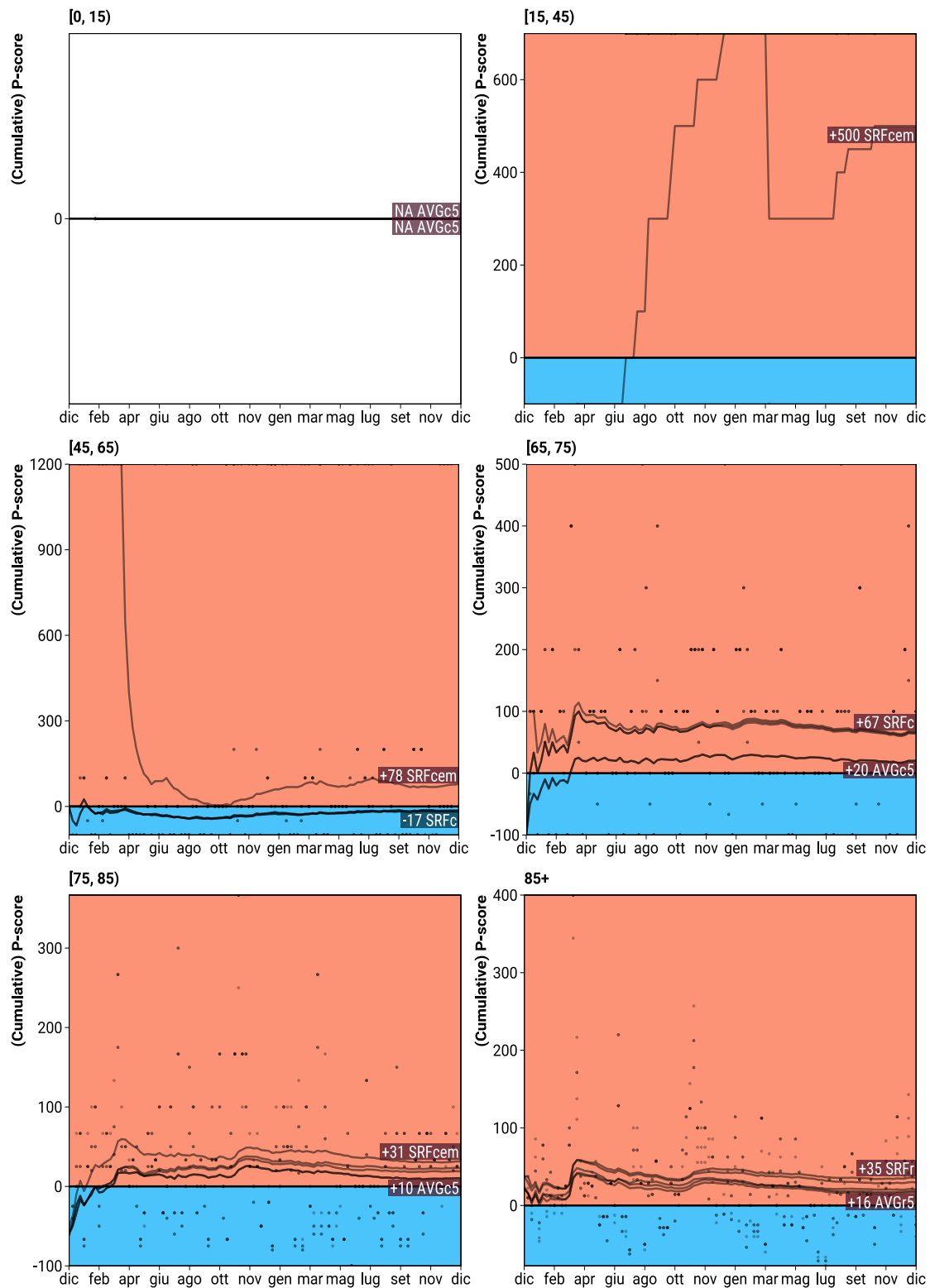


Figura B.40: Popolazione femminile della Valle d'Aosta.

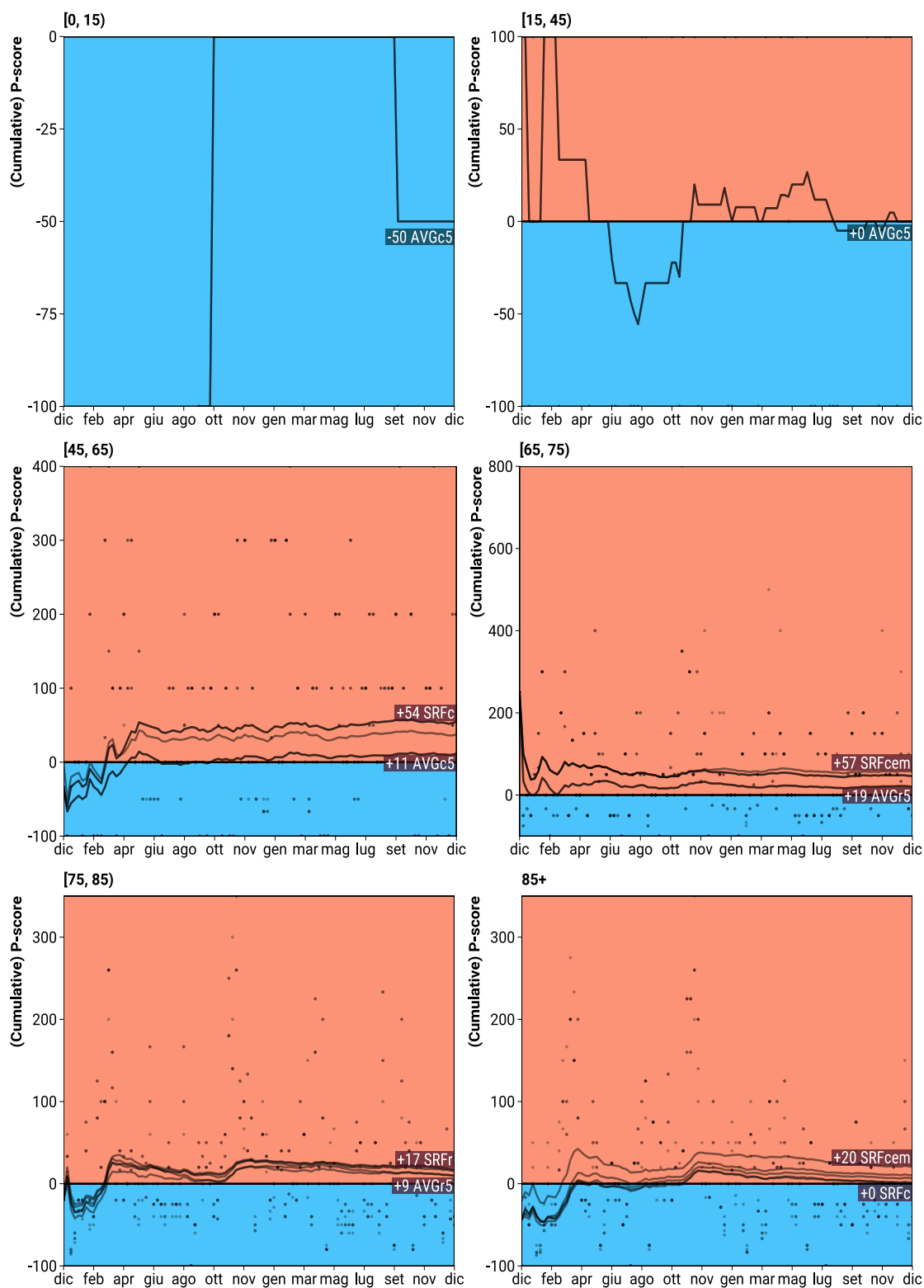


Figura B.41: Popolazione maschile della Valle d'Aosta.

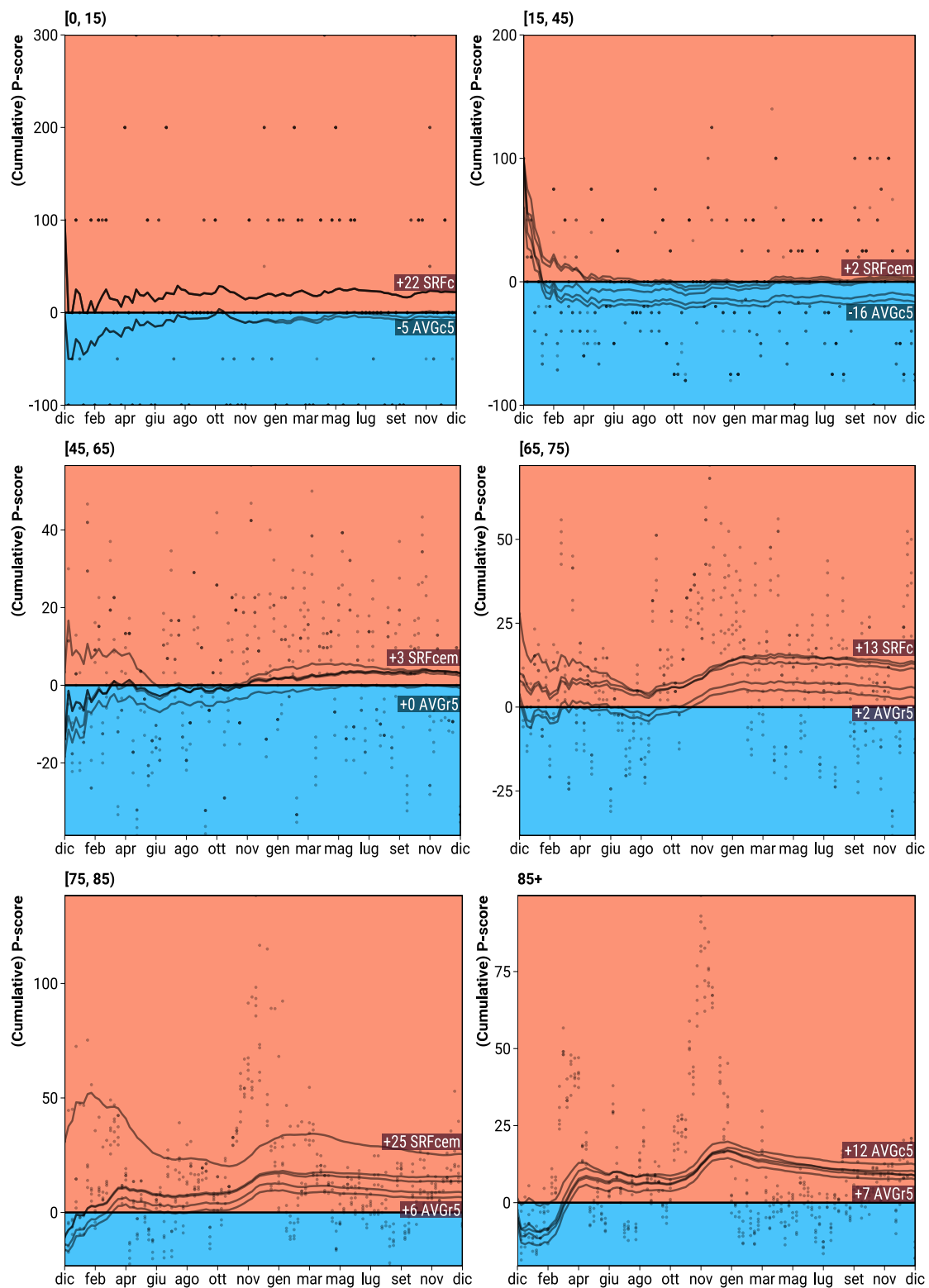


Figura B.42: Popolazione femminile del Veneto.

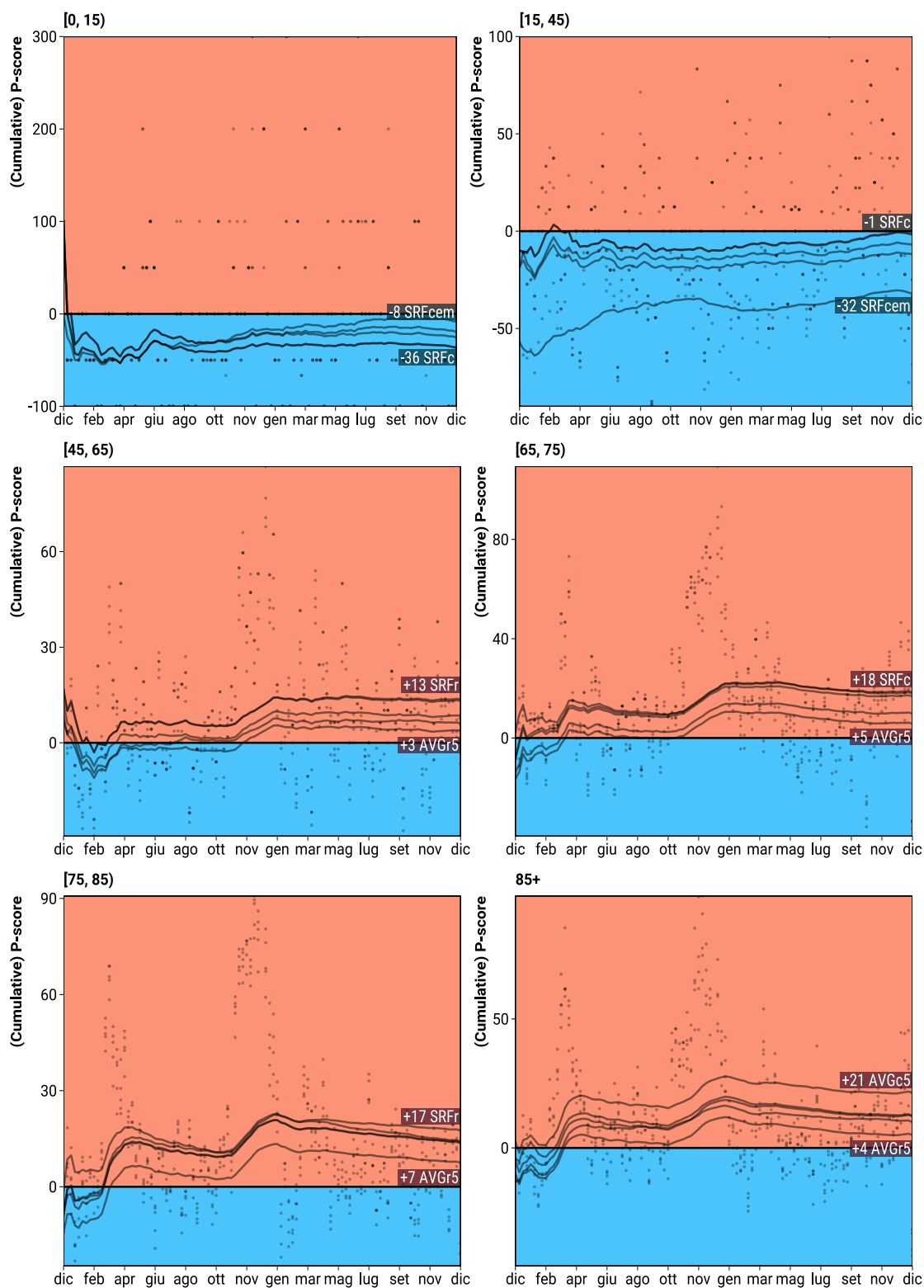


Figura B.43: Popolazione maschile del Veneto.

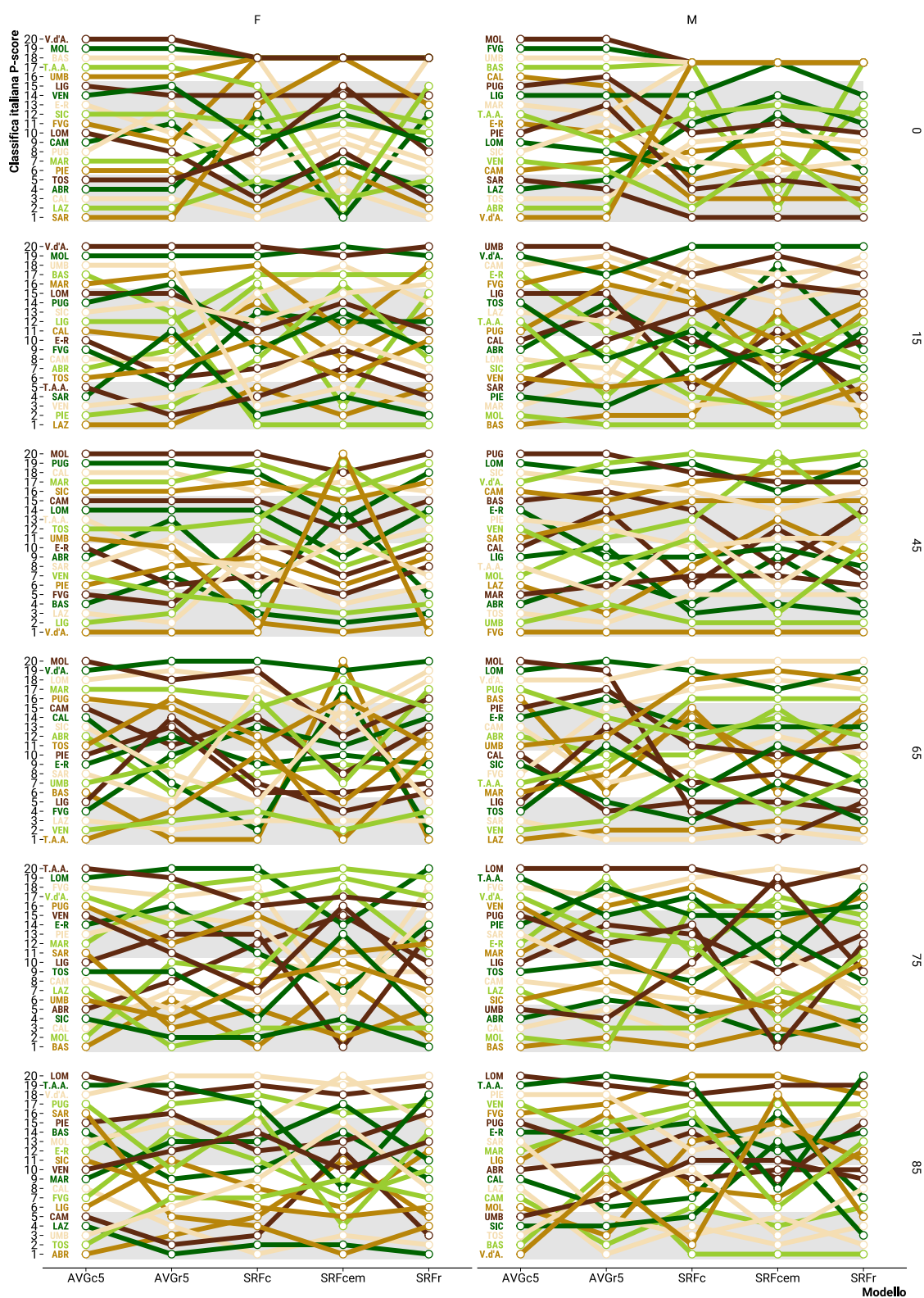


Figura B.44: Stratificazione per sesso ed età della classifica delle regioni italiane fatta rispetto all'eccesso di mortalità percentuale per il 2020-2021 sotto cinque modelli diversi.

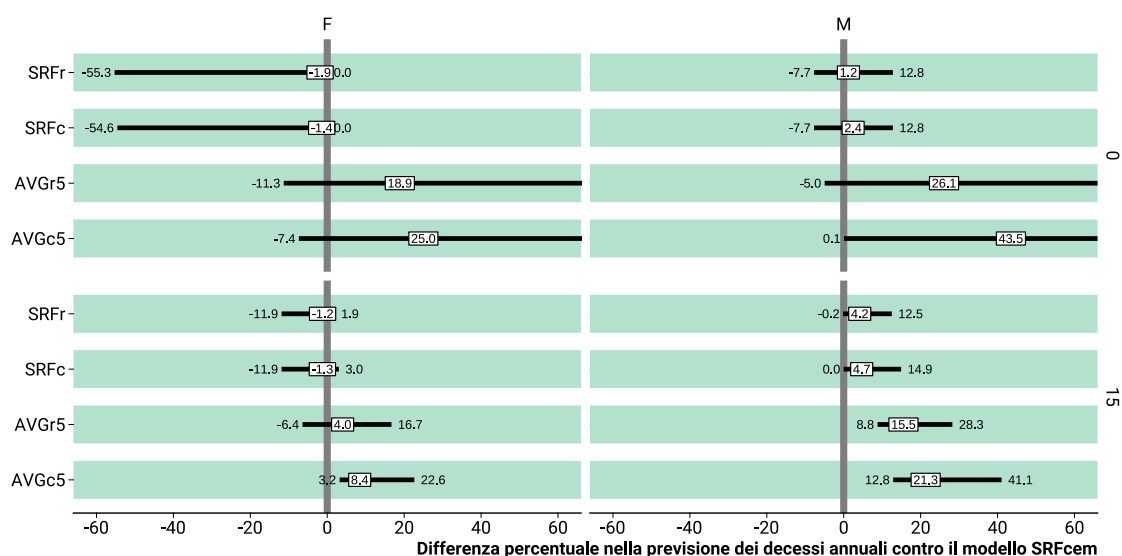


Figura B.45: Differenza percentuale delle morti predette per il 2020-2021 da diversi modelli contro le previsioni del modello Serfling Euromomo. Classi di età  $[0, 15)$  e  $[15, 45)$ . Le distribuzioni italiane della differenza degli errori sono riassunte dai quantili 0.25, 0.5 e 0.75. I valori del quantili 0.75 della prima classe di età per i modelli *5 year average* non compaiono nel grafico perché hanno un valore maggiore di 364 o pari a Inf.

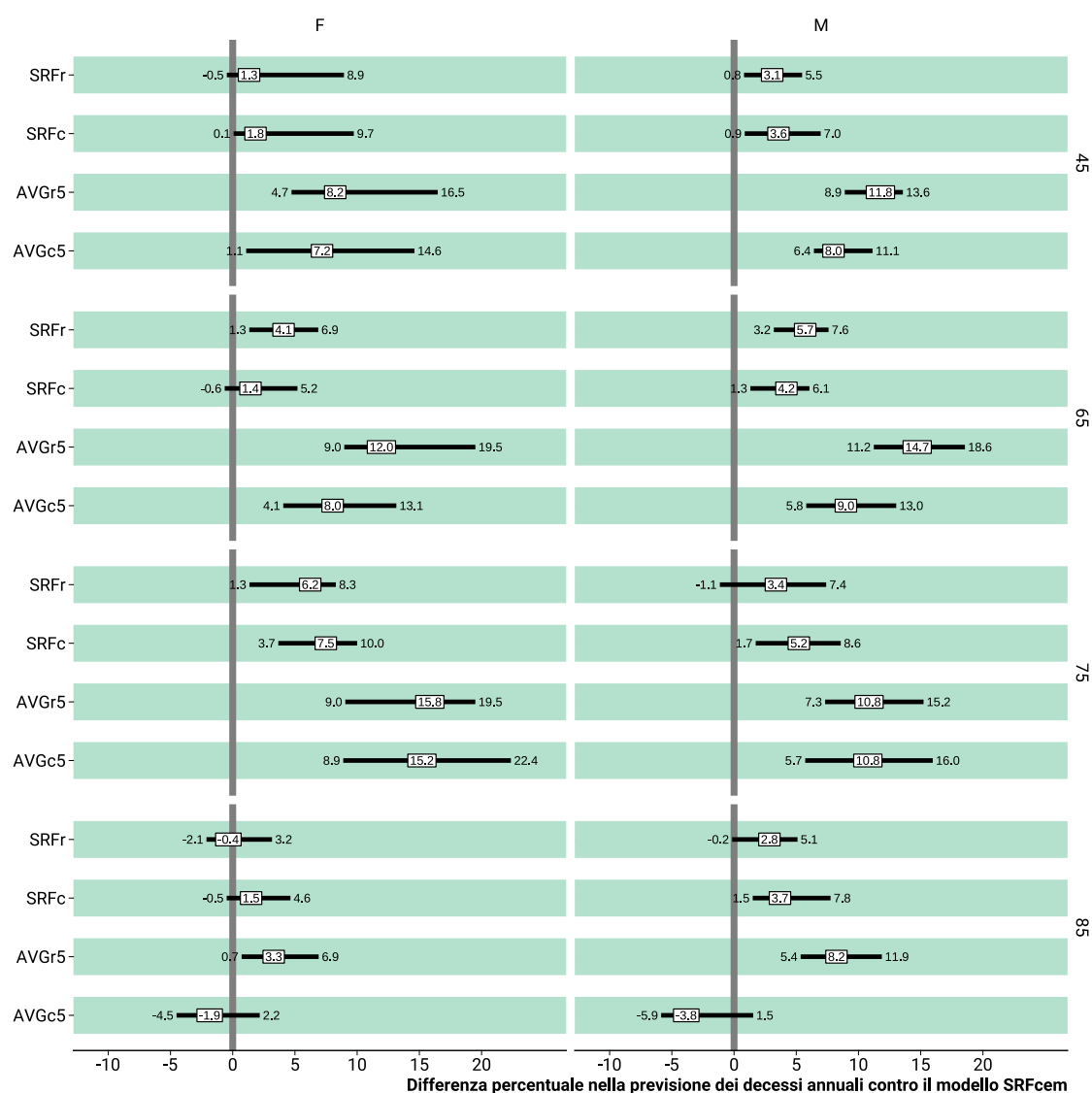


Figura B.46: Differenza percentuale delle morti predette per il 2020-2021 da diversi modelli contro le previsioni del modello Serfling Euromomo. Classi di età [45, 65) e [65, 75) e 85+. *Le distribuzioni italiane della differenza degli errori sono riassunte dai quantili 0.25, 0.5 e 0.75.*

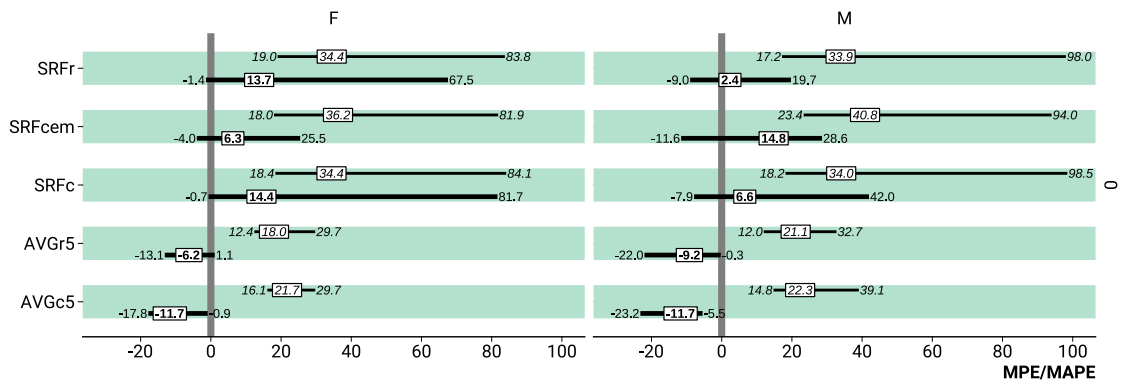


Figura B.47: Errore medio percentuale (MPE, in grassetto e sotto) ed errore assoluto medio percentuale (MAPE, in corsivo e sopra) commesso dai modelli, negli strati sesso-età, prevedendo le morti annuali sui dati test. Popolazione femminile e maschile di età compresa tra 0 anni e 15 anni esclusi. *La distribuzione degli errori a livello italiano è riassunta dai quantili 0.25, 0.5 e 0.75.*

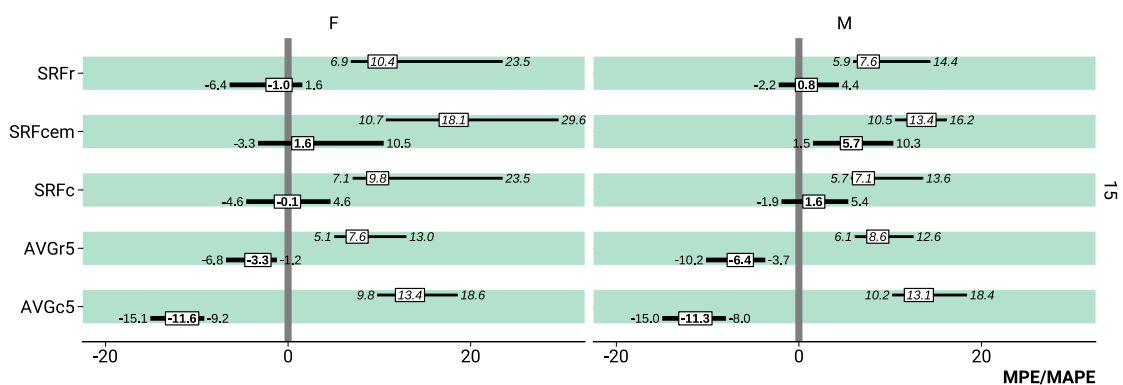


Figura B.48: Errore medio percentuale (MPE, in grassetto in basso) ed errore assoluto medio percentuale (MAPE, in corsivo in alto) commesso dai modelli, negli strati sesso-età, prevedendo le morti annuali sui dati test. Popolazione femminile e maschile di età compresa tra 15 anni e 45 anni esclusi. *La distribuzione degli errori a livello italiano è riassunta dai quantili 0.25, 0.5 e 0.75.*

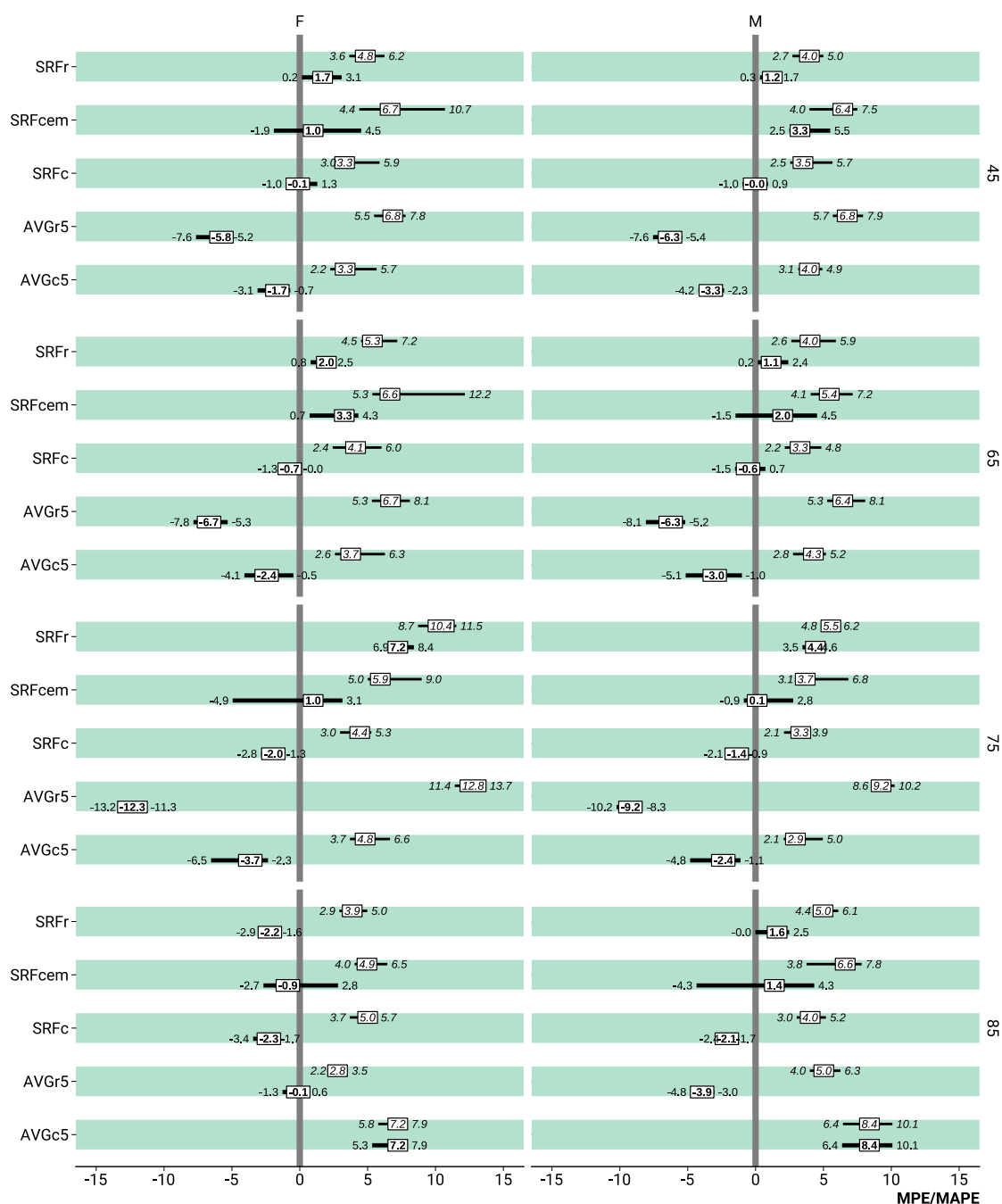


Figura B.49: Errore medio percentuale (MPE, in grassetto in basso) ed errore assoluto medio percentuale (MAPE, in corsivo in alto) commesso dai modelli, negli strati sesso-età, prevedendo le morti annuali sui dati test. Popolazione femminile e maschile con un'età maggiore di 45 anni, inclusi, suddivisa in classi di età. *Le distribuzioni degli errori a livello italiano sono riassunte dai quantili 0.25, 0.5 e 0.75.*

# Bibliografia

- [1] J. M. Aburto, R. Kashyap, J. Schöley, C. Angus, J. Ermisch, M. C. Mills, and J. B. Dowd. Estimating the burden of the covid-19 pandemic on mortality, life expectancy and lifespan inequality in england and wales: a population-level analysis. *Journal of Epidemiology and Community Health*, jan 2021.
- [2] A. Campbell and S. Ward. Comparisons of all-cause mortality between european countries and regions: 2020. *Office for National Statistics*, Mar. 2021.
- [3] Demo-Geodemo. Popolazione residente al 1 gennaio 2020, 2021.
- [4] Demo-Geodemo. Popolazione residente al 1 gennaio 2021, 2021.
- [5] Tino Hager. nager.date - worldwide public holiday, 2022.
- [6] ISTAT. Decessi e cause di morte, 2022. Ultimo accesso 7 maggio 2022.
- [7] ISTAT. Popolazione intercensuaria - ricostruzione della popolazione residente 2002-2019, 2022.
- [8] Corrado Magnani, Danila Azzolina, Elisa Gallo, Daniela Ferrante, and Dario Gregori. How large was the mortality increase directly and indirectly caused by the covid-19 epidemic? an analysis on all-causes mortality data in italy. *International Journal of Environmental Research and Public Health*, 17(10), 2020.
- [9] J. Schöley. Robustness and bias of european excess death estimates in 2020 under varyng model specifications. *medRxiv*, 2021.
- [10] Robert E. Serfling. Methods for current statistical analysis of excess pneumonia-influenza deaths. *Public Health Reports (1896-1970)*, 78(6):494–506, 1963.

- [11] Andreas Stanga, Fabian Standl, Bernd Kowall, Bastian Brunec, Juliane Böttcher, Marcus Brinkmann, and Ulf Dittmer, and Karl-Heinz Jöckel. Excess mortality due to covid-19 in germany. *Journal of Infection*, Sept. 2020.