

Ad alcune note non ho risposto perché mi trovo d'accordo.

Nelle aggiunte che hai fatto al testo ho controllato più i concetti che hai scritto piuttosto che il come li hai scritti, lascio quel compito alla prof.

## Nota 3

---

Sono d'accordo, se è una cosa complessa dal punto di vista del codice meglio concentrarsi su altro al momento.

## Nota 6

---

I risultati con i 95% credible intervals sono molto belli! Un possibile miglioramento (minimo, quindi fallo se è una cosa veloce): proverei a riordinare i credible intervals sull'asse x in base a un criterio, così da ottenere un grafico ancora più leggibile. Per esempio, potresti ordinarli in base alla distanza tra predicted mean e valore vero, oppure per larghezza del credible interval.

## Nota 6 (Kriging)

---

Per il titolo, l'unica cosa che mi viene in mente è di mettere il titolo della sezione come "comparing the two algorithms", poi partizionare il capitolo in due sottosezioni, la prima chiamata "Simulations" e la seconda chiamata "Real-world application", oppure "Application on air pollution data" o qualcosa del genere. Se ti ispira quest'idea prova a proporla alla prof., sicuramente lei è più esperta di me nella scrittura.

Per quanto riguarda i grafici, suggerirei di rimanere sul penultimo plot della sezione perché dà una buona visione di qual è il range dei dati e di quanto le time series selezionate siano variabili rispetto ai dati. Nell'ultimo grafico della sezione infatti tutte le time series sembrano estremamente variabili (quasi come fosse del rumore bianco) e non si riesce a capire che la time series blue è molto più stabile di quella rossa.

Piccola nota grafica: nei grafici, lascerei i true values in nero, per rappresentare il fatto che questi valori non sono conosciuti all'algoritmo e che quindi non subiscono un clustering (solitamente se coloriamo qualcosa è per mostrare in che cluster sta), ciò che l'algoritmo mette in un cluster sono i valori della prediction, che hai giustamente colorato.

Inoltre, aggiungerei due parole sulla time series rossa, mostrando per esempio in che cluster viene messa, dove si trova spazialmente e magari che covariate ha; tentando di trovare una giustificazione intuitiva del perché la predizione va abbastanza male. Per esempio, potrebbe essere un punto spazialmente molto lontano da tutti gli altri, oppure un punto che viene spesso messo in un singleton, e così via.

## Nota 8

---

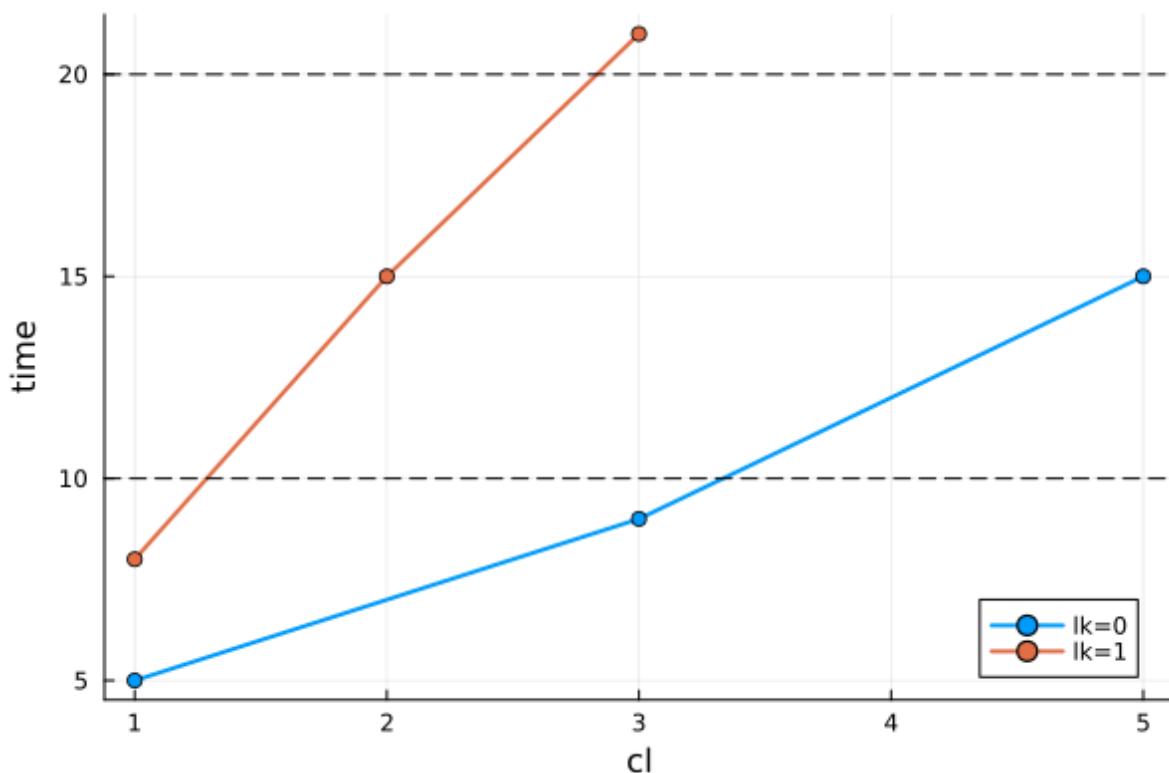
Molto belli i plot dei 95% credible intervals. Piccolo suggerimento grafico: potresti evidenziare in qualche modo i credible intervals che non contengono lo 0, per mostrare subito a colpo d'occhio i punti in cui sappiamo al 95% che una certa covariata ha effetto sull'inquinamento.

Continuano a non convincermi i plot della prima colonna, perché secondo me guardandoli si ricavano le stesse informazioni dei plot della seconda colonna (cioè, come vanno i valori di beta), ma in maniera un po' più confusionaria. Dall'altra parte, capisco che non sia fattibile mostrare 72 traceplot e che selezionarne solo alcuni potrebbe sembrare cherry picking. Una soluzione che potrei proporti è quella di selezionare alcuni traceplot da mostrare nel capitolo, e poi di rimandare all'appendice in cui presenti tutti i traceplots. In questo modo fai vedere che l'MCMC converge (cioè che durante l'MCMC i valori di beta vanno su e giù per bene, esplorando lo spazio della

posterior) ma ti proteggi da qualsiasi accusa, perché chi vuole può andare in appendice e controllare che tutti i traceplot vanno abbastanza bene.

## Nota 12

Ottimo! Un ulteriore grafico (che non aggiunge alcuna informazione, ma che magari dà un maggiore colpo d'occhio) potrebbe essere il seguente: sull'asse delle x il numero di covariate nel clustering (quello che tu hai chiamato  $cl$ ), sull'asse delle y il tempo. Metti una linea verticale per CDRPM con spazio e un'altra per CDRPM senza spazio. Poi, plotti una time series per i risultati di JDRPM + space +  $Xlk + Xcl$  con  $lk=0$ , un'altra con  $lk=1$ , e così via. Così, a colpo d'occhio vedi subito quando la tua implementazione supera CDRPM. Potrebbe essere un piccolo abbellimento, quindi dagli una priorità minima rispetto al resto delle modifiche. Metto qui sotto un plot per mostrarti l'idea:



## Domande

Per quanto riguarda la domanda sul codice, non ho abbastanza esperienza riguardo le tesi per sapere se è una pratica comune. La mia opinione è questa: se qualcuno vuole leggere il codice allora andrà su GitHub (o clonerà la cartella) che è sicuramente più comodo che leggere su un PDF. Secondo me questa appendice potrebbe essere interessante per riportare qualche punto interessante del codice, un po' come hai fatto nel secondo capitolo della tesi. Quindi, potresti mostrare che l'interfaccia è leggibile e ben documentata, o alcune scelte fatte per aumentare la modularità e la velocità del codice, e così via.

Comunque, stiamo parlando di un'appendice molto tecnica, quindi ti suggerirei di dare priorità bassa a questa cosa.

magari non mettere il codice completo, ma mostrare come ogni parte dell'MCM algoritmo viene trasposta in Julia, per mostarne l'efficacia

magari non mettere il codice completo, ma mostrare come ogni parte dell'MCM algoritmo viene trasposta in Julia, per mostarne l'efficacia