

---

# Praktikumsbericht

---

Joshua Eid

401754

Bachelor Technomathematik

Betreuer / Betreuerin: Lennart Quante, Robin Middelanis

Institution / Unternehmen: Potsdam-Institut für Klimafolgenforschung (PIK) e.V.

06.08.2022

# Inhaltsverzeichnis

<b>1</b>	<b>Das Institut</b>	<b>1</b>
<b>2</b>	<b>Beschreibung meiner Tätigkeiten und Erläuterung der Resultate</b>	<b>2</b>
2.1	Das Modell <i>Acclimate</i> und die Ausgabe <i>AcclimateOutput</i> . . . . .	2
2.2	Die zeitliche Veränderung von Korrelationsstrukturen und die <i>rollende PCA</i>	4
2.3	Analyse und Interpretation der PCA . . . . .	6
<b>3</b>	<b>Fazit und Ausblick</b>	<b>9</b>

# 1 Das Institut

Das Potsdam-Institut für Klimafolgenforschung ist eine Forschungseinrichtung, welche sich hauptsächlich mit dem Thema globale Nachhaltigkeit auseinandersetzt. Zum einen wird in den Forschungsabteilungen *Erdsystemanalyse*, *Klimaresilienz*, *Transformationspfade* und *Komplexitätsforschung* Grundlagenforschung betrieben. Hierbei untersuchen Natur- und Sozialwissenschaftler\*innen in einem interdisziplinären Ansatz den globalen Klimawandel sowie die zugehörigen ökologischen, ökonomischen und sozialen Folgen. Zum anderen wird auf Grundlage dieser Forschung Beratung für Politiker\*innen angeboten und konkrete Handlungsmöglichkeiten für eine „sichere und gerechte Klimazukunft“ (PIK, 2022) aufgezeigt. Forschende des Instituts sind somit an der Arbeit des *Intergovernmental Panel on Climate Change (IPCC)*, des *Wissenschaftlichen Beirats der Bundesregierung Globale Umweltveränderungen (WBGU)*, des *Sachverständigenrats für Umweltfragen (SRU)*, der *Kohlekommission* oder der *Leopoldina* beteiligt. Das Institut ist zudem Mitglied der Leibniz-Gemeinschaft.

Ich habe das Praktikum in der Abteilung *Komplexitätsforschung* absolviert. Verschiedene Arbeitsgruppen entwickeln mit Hilfe von Maschinellem Lernen, Nichtlinearen Zeitreihenanalysen und Entscheidungsstrategien Methoden, um die Interaktion verschiedener, vom Klimawandel betroffener, Systeme zu verstehen. Hierbei wird zum Beispiel die Struktur- und Funktionsbildung in komplexen Netzwerken untersucht oder es werden Entscheidungen einzelner gesellschaftlicher Akteure und daraus resultierende Schadenskaskaden modelliert.

Über die vier Wochen war ich in der Arbeitsgruppe „Numerische Analyse globaler ökonomischer Folgen“ aktiv. Diese hat das agentenbasierte Modell *Acclimate* entwickelt, um die Ausbreitung von Produktionsverlusten aufgrund von Extremwetterereignissen entlang globaler Lieferketten zu analysieren. Hierbei wird der lokale Produktionsverlust infolge von Hitzewellen, Überflutungen in Flussgebieten und tropischen Wirbelstürmen simuliert, um anschließend die weltweiten Auswirkungen in einer Zeitreihenanalyse zu untersuchen. Der Anspruch ist „die Forschungslücke zwischen kurzfristigen wirtschaftlichen Schocks und langfristigen Wachstumseffekten zu schließen“ (PIK2, 2022).

Ich habe mich insbesondere mit dem sogenannten *Postprocessing* des Modells beschäftigt. Hierbei sollen automatisierte Methoden entwickelt werden, um die Ausgabe des Modells am Ende einer Zeitreihe systematisch nach Mustern zu untersuchen. Diese können anschließend genauer von den Forschenden betrachtet werden. Aufgrund der hohen Menge an erzeugter Information soll die Analyse somit vor allem zeiteffizienter gestaltet werden.

## 2 Beschreibung meiner Tätigkeiten und Erläuterung der Resultate

Das Praktikum war in drei Teile gegliedert. In der ersten Woche habe ich mich mit dem Modell *Acclimate* vertraut gemacht sowie die dort verwendete Datenstruktur der Ausgabe in Python kennengelernt und erste Visualisierungen ausprobiert. In der zweiten und dritten Woche habe ich einen Fachaufsatz zur zeitlichen Veränderung von Korrelationsstrukturen im Finanzmarkt gelesen, die dort verwendeten Techniken verstanden und diese dann auf die Ausgabe von *Acclimate* übertragen. In der dritten und vierten Woche habe ich anschließend die aus den angewandten Techniken erhaltenen neuen Informationen analysiert, diese Erkenntnisse der Arbeitsgruppe im wöchentlichen Kolloquium vorgestellt, besprochen und weitere mögliche Vorgehensweisen erarbeitet.

### 2.1 Das Modell *Acclimate* und die Ausgabe *AcclimateOutput*

*Acclimate* ist ein dynamisches agentenbasiertes Modell, welches die Ausbreitung von Produktionsverlusten aufgrund von Extremwetterereignissen entlang globaler Lieferketten beschreibt. Hierbei wird ein globales Liefernetzwerk unter der Annahme einer bestehenden nachfrageorientierten Wirtschaft nachgebaut. Einzelne Länder dienen hier als wirtschaftliche Regionen, welche durch Handelsströme untereinander vernetzt sind. Jedes Land wird in 27 Sektoren unterteilt, um es wirtschaftlich zu beschreiben. Einzelne Sektoren sind zum Beispiel die Landwirtschaft, finanzielle Dienstleister, Holz, Öl oder Konsument\*innen. Diese Sektoren dienen als Agenten wodurch bei 268 wirtschaftlichen Regionen (die Vereinigten Staaten von Amerika und China werden werden abermals in Bundesstaaten beziehungsweise Provinzen unterteilt) über 7000 Agenten entstehen. Alle Agenten versuchen nun ihren Profit zu maximieren. Zur Simulation des Liefernetzwerkes werden bis zu 44 Variablen auf regionaler und sektoraler Ebene berechnet und über die Zeitreihe hinweg verfolgt. Beispiele hierfür sind das Bruttoinlandsprodukt und der Import/Export auf regionaler oder die Produktionskapazität und der Produktionspreis auf sektoraler Ebene. Um den Profit zu maximieren, können Agenten nun beispielsweise die Produktionskapazität ändern oder ihre Lieferketten regional umstrukturieren. Hierbei können sie insbesondere gewisse Lagerbestände vor Ort und während des Transports anpassen.

Um nun zunächst lokale Auswirkungen einer Klimakatastrophe zu simulieren wird das anfängliche Equilibrium des Liefernetzwerkes durch die Reduktion der Produktionskapazität einzelner Agenten, das sogenannte *Forcing*, gestört. Hierfür werden Daten bekannter Klimakatastrophen, zum Beispiel des Hurricanes Harvey, verwendet.

Das Modell soll zum einen die bereits erwähnte Forschungslücke schließen. Zum anderen sollen die weniger erforschten *indirekten Produktionsverluste*, also die Verluste, die nicht lokal entstehen, sondern durch die Lieferketten global verbreitet werden, analysiert werden.

Diese Verluste können einerseits stromabwärts einer Lieferkette durch Lieferengpässe entstehen, wenn Produkte eines ursprünglich betroffenen Agenten für die Produktion weiterer Güter benötigt werden und sich dies folglich auf andere Produkte ausweitet. Andererseits können Verluste stromaufwärts einer Lieferkette entstehen, wenn eingeschränkte Produktionskapazitäten ursprünglich betroffener Agenten sich auf die Nachfrage von Gütern jener Agenten auswirkt und somit der Absatzmarkt anderer Agenten eingeschränkt wird und diese wiederum ihre Produktion zurückfahren. Die beiden erwähnten Lagerungen von Gütern stellen sich in diesem Modell als zentral heraus, um diese indirekten Produktionsverluste zu mildern.

Nachdem ich in den ersten Tagen verschiedene erklärende und erste analysierende Arbeiten zu *Acclimate* durchgelesen hatte, habe ich mich dem *Postprocessing* zugewendet. Im *Postprocessing* werden die für alle Agenten oder Regionen vorhandenen, über die Zeit berechneten 44 Variablen analysiert.

Hierfür wurde von den Betreuern eine neue Python-Klasse *AcclimatOutput* erstellt, welche auf dem Konzept multidimensionaler Arrays beruht. Um das Selektieren und Aggregieren von Daten zu vereinfachen, werden in sogenannten *Datasets* die 44 Variablen über verschiedene Dimensionen, wie den Agenten, den Regionen und der Zeit, gespeichert. Ich habe mich zunächst mit dieser neuen Python-Klasse vertraut gemacht, wobei im Laufe der Zeit immer wieder Fehler in der Konstruktion auftraten, welche dann in Zusammenarbeit mit den Betreuern behoben wurden.

Nach dieser Einarbeitung konnte ich die erzeugten Daten nun beliebig selektieren oder aggregieren und mit Hilfe von Pipelines verschiedene Berechnungen durchführen. So habe ich beispielsweise für verschiedene Agenten und verschiedene Variablen in einem festen Zeitfenster die Standardabweichung berechnet und dieses Zeitfenster über die gesamte Zeitreihe laufen lassen. In Abbildung 1 ist beispielhaft die sich zeitlich verändernde Standardabweichung des Produktionswertes des Agenten Landwirtschaft in Afghanistan während der Simulation von sommerlichen Hitzewellen in der Nordhemisphäre zu sehen. Es ist zu erkennen, dass das Ansteigen und Abfallen dieser Standardabweichung zeitlich mit den Hitzewellen übereinstimmt.

Darauffolgend habe ich mich in mir bereits bekannte oder neue Visualisierungspackages eingearbeitet, wie zum Beispiel Matplotlib, Plotly, Holoviews oder Seaborn. So konnte ich beispielsweise mehr Interaktivität für die Benutzenden schaffen: in Plots konnte nun hineingezoomt werden oder über Benutzendenoberflächen Regionen oder Agenten ausgewählt werden. Zudem wurden Korrelationsplots wie in Abbildung 2 möglich, bei der der Produktionswert in Relation zur Produktionskapazität des Agenten Landwirtschaft in Deutschland abgebildet wird.

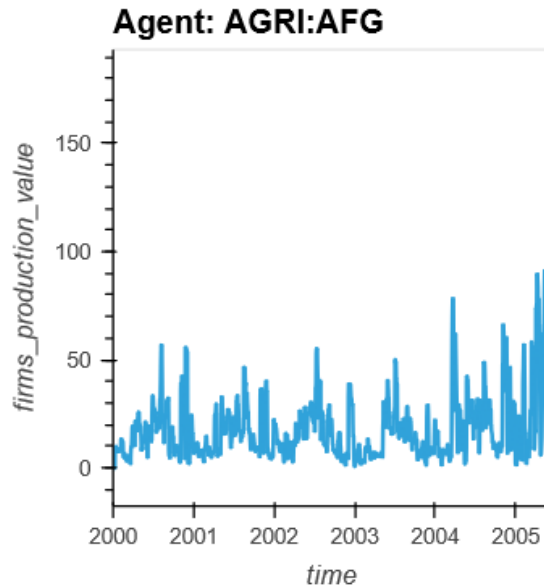


Abbildung 1: Rollende Standardabweichung des Produktionswertes des Agenten Landwirtschaft in Afghanistan.

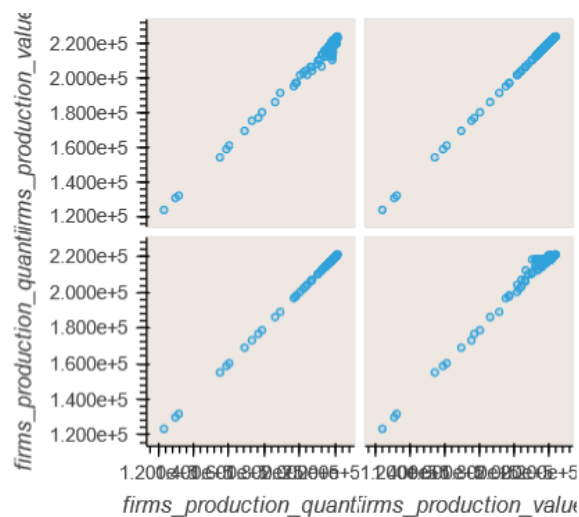


Abbildung 2: Korrelationsplot bezüglich Produktionswert und Produktionskapazität des Agenten Landwirtschaft in Deutschland.

## 2.2 Die zeitliche Veränderung von Korrelationsstrukturen und die *rollende PCA*

Um im Postprocessing die Ausgabe des Modells weitergehend zu analysieren, war für den Rest des Praktikums der Fachaufsatz *Temporal evolution of financial-market correlations* [F11] leitgebend.

In diesem Fachaufsatz wird der Zusammenhang zwischen den Preisen von 98 Finanzprodukten wie Dividenden- und Rentenpapiere, Währungen oder Rohstoffen über einen Zeitraum von elf Jahren von 1999 bis 2010 analysiert. Da die Preise der Finanzprodukte

auf dieselben wirtschaftlichen Ankündigungen und Marktnachrichten reagieren, liegt die Vermutung nahe, dass bezüglich der verschiedenen Finanzprodukte ähnliche Preiszeitreihen entstehen und diese möglicherweise korrelieren. Weitergehend stellt sich die Frage wie genau diese Korrelationsstrukturen aussehen und ob sich diese möglicherweise über die Zeit verändern. Daraus ableitend reicht es aus bestimmte Finanzprodukte zu beobachten, da sich der Rest „ähnlich“ verhält. Zudem sollten Anleger\*innen ihr Portfolio ändern wenn die darin enthaltenen Vermögenswerte stark korreliert sind.

Die Forschenden können tatsächlich zeigen, dass gewisse Korrelationsstrukturen vorhanden sind und diese sich zum Beispiel nach der Kreditkrise 2007/2008 verändern (die verschiedenen Finanzprodukte korrelieren zum Beispiel stärker untereinander und es entstehen teilweise neue Korrelationsstrukturen). Doch was ist eigentlich mit einer Korrelationsstruktur und einer Änderung dieser gemeint? Dies erklärt sich durch die Vorgehensweise der Autor\*innen: das Anwenden einer „rollenden“ Hauptkomponentenanalyse (oder PCA - Principal Component Analysis).

Die Zeitreihen der Preise der 98 Finanzprodukte werden als Matrix gesehen, mit Finanzprodukten als Zeilen und Zeiteinheiten der Zeitreihe als Spalten, und betrachten zunächst einen Zeitausschnitt  $Z_1$  (siehe auch Abbildung 3).

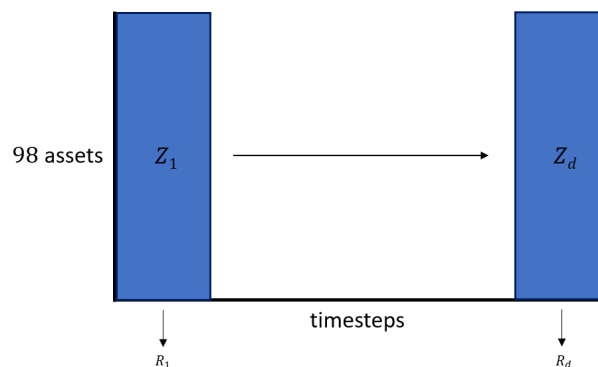


Abbildung 3: Rollende PCA.

Dieser multivariate Datensatz wird standardisiert. Dann wird die *Korrelationsmatrix* beziehungsweise *Kovarianzmatrix* berechnet bei der die Varianz der Variablen (Finanzprodukte) auf der Hauptdiagonalen und die Kovarianzen auf der Nebendiagonale enthalten sind. Es ist möglich diese Matrix spektral zu zerlegen, wobei die Eigenwerte der Größe

nach sortiert sind und die Matrix  $E$  orthogonal mit zugehörigen Eigenvektoren ist:

$$\begin{aligned}
 R &= \frac{1}{T} Z Z^T \\
 &= \begin{bmatrix} \text{Var}(Z_1) & \text{Cov}(Z_1, Z_2) & \cdots & \text{Cov}(Z_1, Z_N) \\ \text{Cov}(Z_2, Z_1) & \text{Var}(Z_2) & \cdots & \text{Cov}(Z_2, Z_N) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(Z_N, Z_1) & \text{Cov}(Z_N, Z_2) & \cdots & \text{Var}(Z_N) \end{bmatrix} \\
 &= \begin{bmatrix} E_1 & \cdots & E_N \end{bmatrix} \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_N \end{bmatrix} \begin{bmatrix} E_1 & \cdots & E_N \end{bmatrix}^T \\
 &= E D E^T
 \end{aligned}$$

Da die Matrix  $E$  orthogonal ist, kann die Spektralzerlegung als Rotation des Variablenraums gesehen werden. Die Matrix  $E$  kodiert die Koordinaten der neuen Variablen (auch Hauptkomponenten genannt), welche Linearkombinationen der ursprünglichen Variablen sind. Diese Koordinaten werden *Loadings* genannt.

Entscheidend ist nun, dass die Totale Varianz (die Verallgemeinerung der empirischen Varianz eines univariaten Datensatzes auf einen multivariaten Datensatz) der Spur der Kovarianzmatrix entspricht und invariant bezüglich Rotationen ist. Die Hauptkomponenten sind somit nach der Größe des Anteils an der Totalen Varianz geordnet.

Dieses Verfahren wird nun wiederholend auf ein Zeitfenster angewendet, welches entlang der Zeitreihe vom Anfang bis zum Ende verschoben wird. Dadurch entstehen sich zeitlich verändernde Loadings, Hauptkomponenten und entsprechende Anteile an der Totalen Varianz (siehe auch Abbildung 3).

Dieses Vorgehen habe ich auf die Ausgabe des Modells angewendet, wobei ich die 268 Regionen als ursprüngliche Variablen verwendet und über mehr als anderthalb Jahre die Entwicklung des Bruttoinlandprodukts nach dem Hurricane Harvey im Jahr 2001 betrachtet habe. Die PCA habe ich sowohl selbst in Python programmiert als auch mit dem Package *scikitlearn* realisiert.

## 2.3 Analyse und Interpretation der PCA

Unter einer Korrelationsstruktur können nun, um zur Frage aus dem vorherigen Abschnitt zurückzukommen, mehrere Aspekte verstanden werden:

- wie viel Totale Varianz durch eine festgelegte Anzahl an Hauptkomponenten erklärt werden kann (z.B. durch die ersten fünf, es werden immer „die ersten ...“ gewählt, da diese per Sortierung am meisten Totale Varianz erklären),
- wie viele Variablen einen entscheidenden Beitrag („Größe“ der Loadings) zu einer



Hauptkomponente leisten,

- welche Variablen einen entscheidenden Beitrag zu einer Hauptkomponente leisten,
- aus den zwei vorherigen Aspekten: ob einer Hauptkomponente eine Bedeutung in der realen Welt gegeben werden kann.

Hierbei entstehen vor allem bei dem Versuch die zeitliche Veränderung zu interpretieren viele Fragen und Unsicherheiten, welche genauer untersucht werden müssen. Hat die erhöhte erklärte Totale Varianz einer Hauptkomponenten mit der erhöhten Korrelation zwischen (wenigen) Variablen zu tun oder ist der Grund ein marktübergreifender Effekt bei dem immer mehr Variablen entscheidend zu einer Hauptkomponente beitragen? Können den Hauptkomponenten sinnvolle Bedeutungen über einen langen Zeitraum gegeben werden und somit in der realen Welt vice versa Hauptkomponenten einen Einfluss auf die Variablen haben? Was bedeutet es, wenn die Beteiligungsquote für jede Hauptkomponente steigt und zeitgleich die erklärte Totale Varianz gleichbleibt? Bedeutet dies weniger Korrelation zwischen den Variablen, da mehr Variablen benötigt werden, um die gleiche Totale Varianz zu erklären?

Diese und weitere Fragen wurden im Anschluss an die Präsentation meiner Ergebnisse während des wöchentlichen Jour Fixe der Arbeitsgruppe in meiner letzten Praktikumswoche, übertragen auf das Acclimate Modell, besprochen. Hierfür wurden unter anderem die folgenden Graphen, die ich durch die PCA erstellen konnte, verwendet: der Anteil der Totalen Varianz der einzelnen ersten fünf Hauptkomponenten, der kumulierte Anteil der Totalen Varianz der ersten fünf Hauptkomponenten und die sogenannte Beteiligungsrate der einzelnen ersten fünf Hauptkomponenten. Die Abbildung 4 zeigt das erzeugte Forcing durch den Hurricane Harvey.

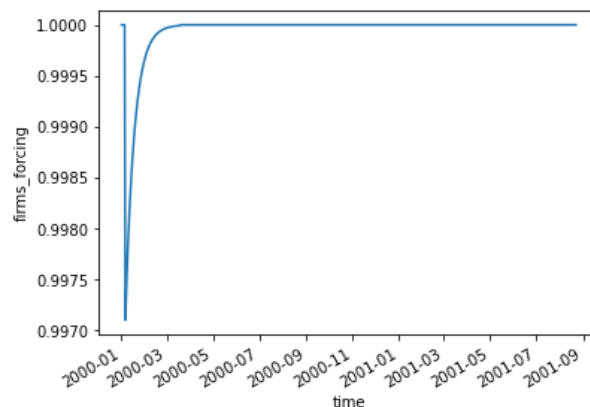


Abbildung 4: Das sogenannte Forcing durch den Hurricane Harvey.

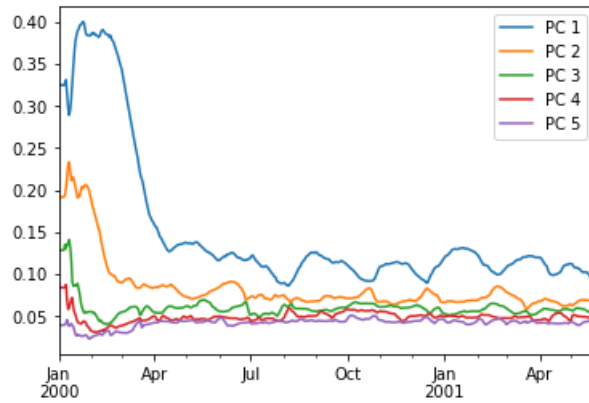


Abbildung 5: Der Anteil der Totalen Varianz.

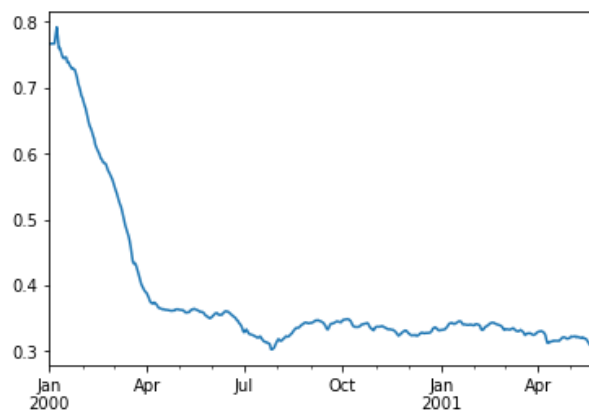


Abbildung 6: Der Anteil der Totalen Varianz kumuliert für die ersten fünf Hauptkomponenten.

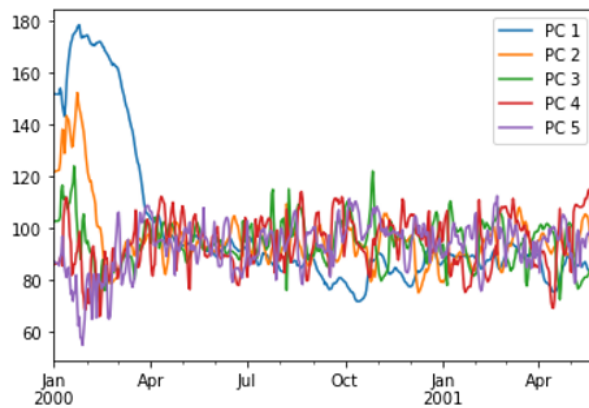


Abbildung 7: Die sogenannte Beteiligungsrate.

Außerdem wurden erste Versuche angestellt den Hauptkomponenten Bedeutungen zuzuschreiben oder ein Veränderung über die Zeit festzuhalten.

Wie in Abbildung 8 zu sehen ist, hat zum Beispiel ein Großteil der Regionen einen positiven Anteil an der ersten Hauptkomponente (hier spielt das Vorzeichen keine Rolle, ausschließlich welche Regionen das gleiche Vorzeichen haben). Mit der Zeit ändern jedoch

die Regionen das Vorzeichen (siehe Abbildung 9, wobei der Impuls des Wechsels eindeutig von Texas und Louisiana ausgeht – die Regionen, die unmittelbar durch das Forcing vom Hurricane Harvey betroffen sind).

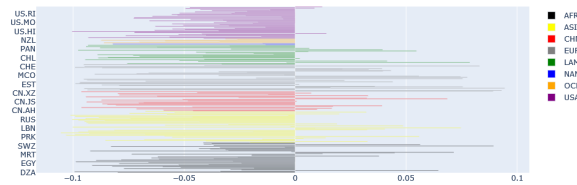


Abbildung 8: Die Loadings am Tag 0 der ersten Hauptkomponente.

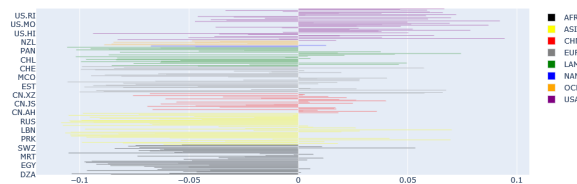


Abbildung 9: Die Die Loadings am Tag 10 der ersten Hauptkomponente.

ALs ein zweites Beispiel lässt sich bezüglich der Loadings der zweiten Hauptkomponente erkennen, dass die Vereinigten Staaten sich auf einer Seite befinden und der Rest der Welt auf der anderen Seite, jene der dritten Hauptkomponente lassen auf ein Cluster der Vereinigten Staaten und China auf der einen Seite und Europa auf der anderen Seite schließen.

### 3 Fazit und Ausblick

Leider hat die Zeit im Praktikum nicht mehr ausgereicht, um weitere Antworten auf die oben gestellten Fragen zu finden und die durch das Postprocessing erhaltene Information weitergehend zu analysieren. Die Arbeitsgruppe wird die Arbeit jedoch weiterführen und hat Ideen entwickelt, um die Ergebnisse der rollenden PCA weiter zu verwenden. Eventuell wird dadurch ein wissenschaftliches Paper entstehen.

Abschließend bin ich mit dem Praktikum sehr zufrieden. Ich konnte zum einen viel Wissen, Programmierfertigkeiten und Arbeitsweisen aus dem Studium anwenden. Zum anderen konnte ich die mir neue Technik der rollenden PCA aneignen und an einem aktuellem Forschungsprojekt mitarbeiten. Die vier Wochen haben mein Interesse an der Klimaforschung geweckt und mich dazu bewegt im kommenden Master Module zu belegen, die ihre

Anwendung in der Klimaforschung finden: Statistik und Machine Learning, Graphentheorie und Differentialgleichungen. Außerdem will ich mich im freien Wahlbereich noch mehr auf klimarelevante Themen fokussieren.

Ich fand es spannend zu sehen, aus welchen unterschiedlichen Fachrichtungen die Promovierenden kommen und an was für unterschiedlichen Themen diese mit ähnlichen Methoden arbeiten. Die Arbeitsatmosphäre war sehr kollegial und unterstützend, es wurde viel Wert darauf gelegt, dass ich Interesse und Spaß an meinen Aufgaben habe. Ich kann dementsprechend ein Praktikum am PIK nur wärmstens weiterempfehlen!

## Abbildungsverzeichnis

1	Rollende Standardabweichung des Produktionswertes des Agenten Landwirtschaft in Afghanistan. . . . .	4
2	Korrelationsplot bezüglich Produktionswert und Produktionskapazität des Agenten Landwirtschaft in Deutschland. . . . .	4
3	Rollende PCA. . . . .	5
4	Das sogenannte Forcing durch den Hurricane Harvey. . . . .	7
5	Der Anteil der Totalen Varianz. . . . .	8
6	Der Anteil der Totalen Varianz kumuliert für die ersten fünf Hauptkomponenten. . . . .	8
7	Die sogenannte Beteiligungsrate. . . . .	8
8	Die Loadings am Tag 0 der ersten Hauptkomponente. . . . .	9
9	Die Die Loadings am Tag 10 der ersten Hauptkomponente. . . . .	9

## Bibliographie

[PIK] <https://www.pik-potsdam.de/de>

[PIK2] <https://www.pik-potsdam.de/de/institut/abteilungen/komplexitaetsforschung/forschung/numerische-analyse-globaler-oekonomischer-folgen/numerische-analyse-globaler-oekonomischer-folgen-1>

[F11] Fenn, D. J., [...], *Temporal evolution of financial-market correlations*, Physical Review E 84, 026109, College Park 2011