

Die Wissenschaft der Attribution – Ein philosophischer Kommentar

Alfred Dandyk

Grundlage dieses Aufsatzes ist das folgende Video:

<https://www.youtube.com/watch?v=WjxFeh5SUt0>

Es handelt sich um einen Vortrag des kanadischen Ökonomen Professor Ross McKittrick. Das Thema ist die sogenannte Attributions-Forschung, die es sich zur Aufgabe gemacht hat, bei bestimmten klimatischen Ereignissen den natürlichen Anteil und den anthropogenen Anteil unterscheiden zu können.

Genauer gesagt, beschäftigt sich McKittrick in dem Video mit einer bestimmten Arbeit der Klimaforscher Myles Allen und Simon Tett, die den folgenden Titel trägt: ‚Checking for model consistency in optimal fingerprinting‘. Gemäß den Verlautbarungen der Autoren soll diese Arbeit eine

unvoreingenommene und "optimale" (präzise) Messung des Ausmaßes der Auswirkungen von Treibhausgasen auf Klimamuster (der "Fingerabdruck")

liefern. McKittrick berichtet, dass die in dieser Arbeit propagierte Methode sofort vom IPCC gefördert und seitdem intensiv genutzt worden ist:

The image shows a video player interface. On the left, a dark slide titled "Optimal Fingerprinting" contains a bulleted list:

- In 1999 Myles Allen and Simon Tett published a paper ("AT99") setting out a method they claimed did the following:
 - Provided unbiased and 'optimal' (precise) measurements of the size of the effect of GHG's on climate patterns (the 'Fingerprint')
 - Allowed researchers to check that the statistical model is properly specified (the 'Residual Consistency' or RC Test)
- The method was immediately promoted by the IPCC and heavily used ever since

On the right, a document titled "Checking for model consistency in optimal fingerprinting" by M. R. Allen & S. F. B. Tett is displayed. The document includes an abstract, an introduction, and contact information for the authors.

Abstract Current approaches to the detection and attribution of an anthropogenic influence on climate involve quantifying the level of agreement between model-predicted patterns of externally forced change and observed changes in the recent climate record. Analyses of uncertainty rely on simulated variability from a climate model. Any numerical representation of the climate is likely to display too little variance on small spatial scales, leading to a risk of spurious detection results. The risk is particularly severe if the detection strategy involves optimisation of signal-to-noise because statistical aspects of model variability may automatically be given high weight through the optimisation. The solution is to confine attention to aspects of the model and of the real climate system in which the model simulation of internal climate variability is adequate, or more accurately, cannot be shown to be deficient. We propose a simple consistency check based on standard linear regression which can be applied to both the space-time and frequency domain approaches to optimal detection and demonstrate the application of this check to the problem of detection and attribution of anthropogenic signals in the radiosonde-based record of recent trends in atmospheric vertical temperature structure. The influence of anthropogenic greenhouse gases can be detected at a high confidence level in this diagnostic, while the combined influence of anthropogenic sulphates and stratospheric ozone depletion is less clearly evident. Assuming the time-scales of the model response are correct, and neglecting the possibility of non-linear feedbacks, the amplitude of the observed signal suggests a climate sensitivity range of 1.2–3.4 K, although the upper end of this range may be underestimated by up to 25% due to uncertainty in model-predicted response patterns.

1 Introduction

A common overall approach has emerged to the detection of anthropogenic climate change. A detection statistic is defined and evaluated in an observational dataset. This might be a global mean quantity (e.g. Solomon et al. 1996), a model versus observation pattern correlation (Mitchell et al. 1995a; Tett et al. 1996), the observed trend in pattern correlation (Santer et al. 1996), or some form of "optimised fingerprint" (Hasselmann 1979; Hasselmann and Frankignoul 1985; Bell 1986; Hasselmann 1993; Santer et al. 1994a; North et al. 1995; Hegerl et al. 1996; North and Stevens, 1999). The same detection statistic is then evaluated treating sections of a control run of a climate model (in which there is no secular change in forcing) as "pseudo-observations" to provide an estimate of the distribution of that statistic under the null-hypothesis of no anthropogenic change. If the observed value of the chosen statistic lies in the uppermost 100 × P% percentile of the distribution estimated from the control, then detection is claimed with a 100 × P% risk of a type-1 error (so P = probability of a false positive). Clearly, this approach to quantifying the risk of error requires complete confidence in the realism of the model simulation of internal climate variability.

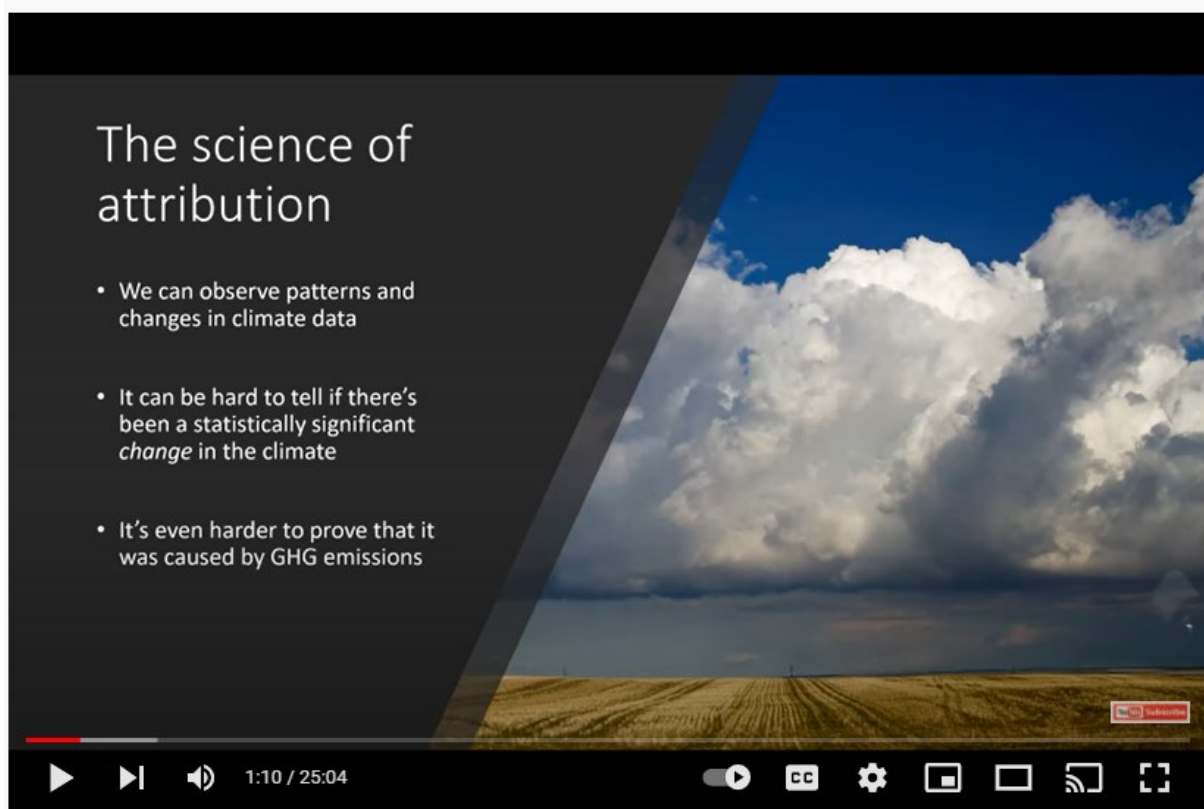
Hasselmann (1977) distinguishes between "detection" of anthropogenic climate change (ruling out, at a certain level of confidence, the null-hypothesis that the climate system is in a steady state).

M. R. Allen (2012)
Space Science Department, Earthland Applied Laboratory,
Chilton, Didcot, OX11 0QJ.
Also at: Department of Physics,
University of Oxford, UK.
E-mail: m.r.a@epla.ac.uk

5:41 / 25:04

Man kann ohne Übertreibung sagen, dass diese *Fingerprint-Methode* eine entscheidende Grundlage der Attributions-Forschung geworden ist, vor allem, wenn es darum geht, den Einfluss der CO₂-Konzentration in der Atmosphäre auf den Klimawandel zu erforschen.

Die folgende Abbildung zeigt drei wesentliche Aspekte der Attributionsforschung im Sinne McKitricks:

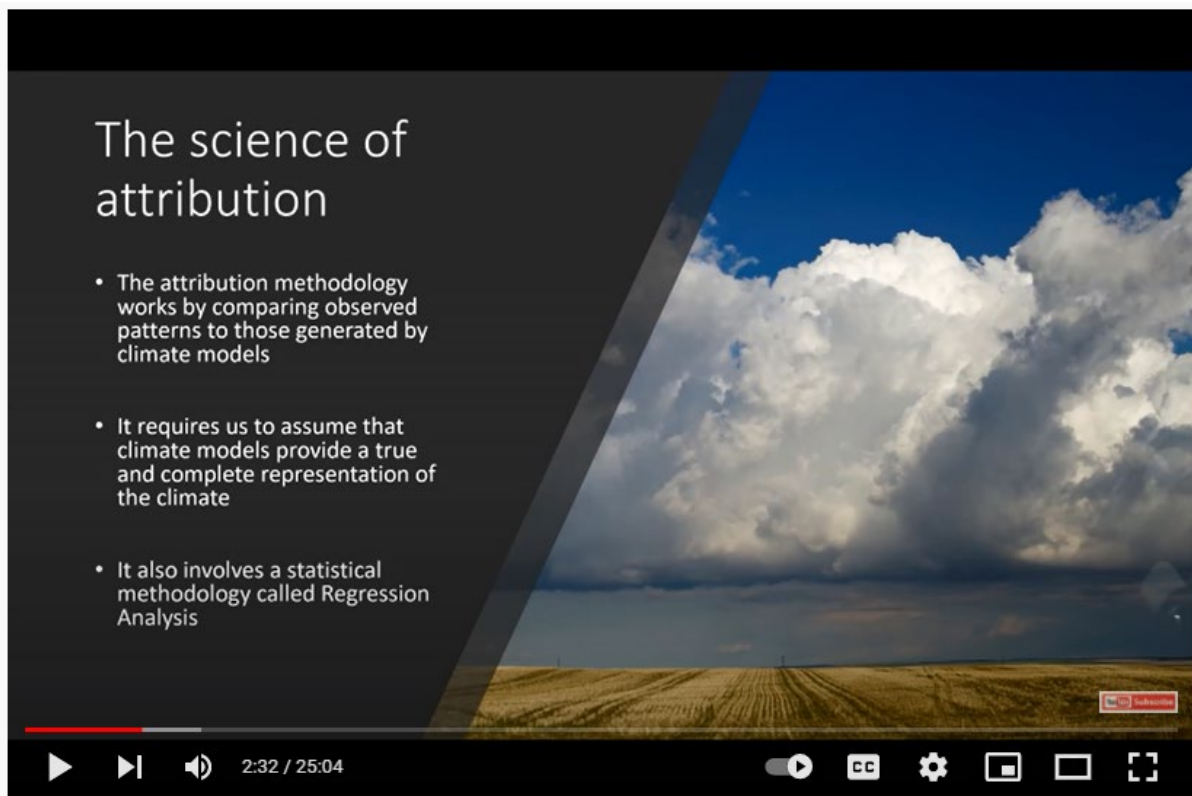


- Beobachtung von Mustern und Veränderungen in den Klima-Daten
- Beurteilung, ob bestimmte Veränderungen des Klimas statistisch signifikant sind
- Beurteilung, ob eine bestimmte Veränderung durch Emission von Treibhaus-Gasen verursacht worden ist.

Nach McKitricks Einschätzung sind Antworten auf Punkt 2 und Punkt 3 schwierig zu finden. Angenommen, es wurden drei besonders heiße und trockene Jahre beobachtet. Ist dieses Muster nun Ausdruck eines Klima-Wandels oder handelt es sich eher um eine Erscheinung im Rahmen der natürlichen Variabilität des Klimas? Noch schwieriger ist die Beantwortung der Frage, ob das genannte Phänomen irgendetwas mit der CO₂-Konzentration in der Atmosphäre zu tun hat oder nicht.

Die genannte Problematik kann noch weiter differenziert werden: Es gibt erstens die Frage nach statistischen Korrelationen und zweitens das Problem des Ursache-Wirkungs-Verhältnisses. Die erste Frage lautet: Kann irgendeine statistische Korrelation zwischen dem Ereignis und der CO₂-Konzentration in der Atmosphäre aufgedeckt werden? Die zweite Frage lautet: Wenn es eine Korrelation gibt, kann diese Korrelation als Ursache-Wirkungs-Verhältnis gedeutet werden?

Die Methode zur Erforschung solcher Fragen liegt in folgendem: Man vergleicht beobachtete Muster und Veränderungen dieser Muster mit den Resultaten von Computer-Simulationen. Bei diesen Computer-Simulationen dient die CO₂-Konzentration in der Atmosphäre als veränderlicher Parameter. Man sucht dann nach der Simulation, die am besten mit den beobachteten Mustern zusammenpasst. Wenn eine Simulation mit einem bestimmten Wert der CO₂-Konzentration am besten mit den beobachteten Daten zusammenpasst, dann soll dieses als Beweis für ein Ursache-Verhältnis zwischen der CO₂-Konzentration und dem beobachteten Muster gelten:



Ich habe versucht, mir dieses Verfahren an einem Beispiel klarzumachen: Angenommen, man kennt ein bestimmtes Muster innerhalb der Mittelalterlichen Warmzeit. Angenommen weiterhin, es gelingt dieses Muster mit Hilfe eines bestimmten Computer-Modells und der Anpassung der Parameter, insbesondere mit einem CO₂-Wert von 280 ppm, zu reproduzieren. Dann ist das ein starker Hinweis dafür, dass dieses Klima-Modell eine gute Simulation der Realität darstellt.

Darüber hinaus sei angenommen, man habe ein bestimmtes Ereignis innerhalb der modernen Erwärmung, zum Beispiel den Starkregen im Ahrtal, und sei in der Lage, dieses Ereignis mit der genannten Simulation zu reproduzieren, allerdings nur unter der Voraussetzung, dass man die CO₂-Konzentration auf den modernen Wert von 400 ppm erhöht. Dann kann das als ein starker Hinweis darauf gedeutet werden, dass dieses Ereignis durch die erhöhte CO₂-Konzentration verursacht worden ist. So oder so ähnlich stelle ich mir die logische Struktur der Argumentation im Rahmen der Attributions-Forschung vor.

McKittrick weist darauf hin, dass dieses Verfahren die statistische Methode der Regressions-Analyse benutzt. Seine Kritik richtet sich gegen gewisse Behauptungen der Autoren, die

technische Aspekte dieser Regressions-Analyse betreffen. Eine Behauptung lautet, dass das Verfahren unvoreingenommene und präzise Messungen von Regressions-Koeffizienten liefert, eine weitere Behauptung lautet, das Verfahren liefere einen Beweis dafür, dass das Modell gültig sei, das heißt, dass es sowohl die natürliche Variabilität als auch die anthropogenen Einflüsse korrekt darstelle.

Die letzte Behauptung wird mittels eines Verfahrens begründet, das ‚residual consistency‘ genannt wird. Es sind diese technischen Aspekte der Regressions-Analyse, welche die Behauptung begründen sollen, dass die moderne Erwärmung hauptsächlich oder sogar ausschließlich anthropogenen Ursprungs ist.

McKitrick geht nun dazu über, eine eher technische Sprache zu benutzen. Demnach verwenden die Autoren das sogenannte *Gauß-Markov-Theorem*, dessen Anwendung einen sogenannten ‚best linear unbiased estimator‘ liefert, einen ‚besten linearen unvoreingenommenen Schätzer‘, kurz: BLUE. Die Behauptung der Autoren lautet demnach technisch gesprochen: Sie können das Gauß-Markov-Theorem so anwenden, dass sie einen Schätzer erhalten, der ‚blue‘ ist, so dass man sagen kann, die erhaltenen Regressions-Koeffizienten seien so unvoreingenommen und so präzise wie möglich.

Die Anwendung des Gauß-Markov-Theorems ist allerdings an Voraussetzungen gebunden. Wenn diese Voraussetzungen nicht gegeben sind, dann sind auch die Regressions-Koeffizienten nicht unvoreingenommen und auch nicht präzise. Es kann sogar sein, dass sie vollkommen sinnlos sind.

McKitrick behauptet nun, dass die Autoren das Gauß-Markov-Theorem nicht korrekt formuliert haben und dass sie die notwendigen Bedingungen für die Anwendung dieses Theorems weder aufgelistet noch überprüft haben. McKitrick versichert sogar, dass die notwendigen Voraussetzungen für die Anwendung des Gauß-Markov-Theorems in diesem konkreten Fall nicht erfüllt sind, so dass auch die Basis für die Vertrauenswürdigkeit der Ergebnisse fehlt. McKitrick nennt diese Ergebnisse ‚potentiell voreingenommen, vollständig irreführend und zum großen Teil sinnlos‘.

Folgt man der Argumentation McKitricks, dann haben die Autoren die notwendige Vorarbeit für die Anwendung des Gauß-Markov-Theorems nicht geleistet. Stattdessen haben sie einen sogenannten ‚RC-Test‘ (residual-consistency) angewandt, dessen Bezug zum Gauß-Markov-Theorem unklar bleibt. McKitrick sagt, dass die Anwendung dieses Tests in diesem Zusammenhang vollkommen bedeutungslos (meaningless) ist.

McKitrick hat seine Bedenken in einer anerkannten wissenschaftlichen Zeitschrift veröffentlichen lassen. Seine Vorbehalte sind in diesem Screenshot noch einmal zusammengefasst:

McKitrick 2021 Comment on AT99 Fingerprinting Method

- Published in *Climate Dynamics*: the same journal that published AT99
- Shows that there were numerous mathematical in AT99 and untrue claims about the reliability of the results
- Also shows that the RC test yields meaningless results



13:41 / 25:04

Demnach zeigt diese wissenschaftlich begutachtete Veröffentlichung, dass in der beanstandeten Arbeit zahlreiche mathematische und unwahre Behauptungen hinsichtlich der Verlässlichkeit der Resultate zu finden sind. Darüber hinaus signalisiert die Veröffentlichung, dass der RC-Test sinnlose Resultate liefert.

Unter der Voraussetzung, dass die Behauptungen Mc-Kitricks korrekt sind, wären die Implikationen für die Arbeit des IPCC verheerend. Sie würden beweisen, dass ein sehr großer Anteil der Attributions-Literatur auf einer Methode basiert, die falsche und bedeutungslose Resultate liefert:

The image shows a video player interface. On the left, a slide titled "Implications" contains two bullet points:

- Despite IPCC claims, there is no basis for confidence that any applications of AT99 fingerprinting yield valid results
- A huge amount of the attribution literature (including studies attributing changes in atmospheric moisture, snow cover, forest fires, etc.) is based on a method that gives wrong and meaningless results

On the right, a tweet from Richard Tol (@RichardTol) is displayed. The tweet, dated August 21, says: "Allen & Tett introduced an FGLS estimator that is wrong and a test to show that they're right when they're not". Below it is a reply from the same user, dated August 20, saying: "another literature that is entirely wrong link.springer.com/article/10.100". The video player controls at the bottom show a progress bar at 15:45 / 25:04 and various icons for play, volume, and settings.

McKittrick verdeutlicht die dadurch entstandene Situation mit Hilfe einer anschaulichen Geschichte. Angenommen, das FBI hat auf der Basis einer neuen Methode zur Auswertung von Fingerabdrücken in den letzten 10 Jahre eine große Anzahl von Personen des Mordes überführt. Die Überführten wurden nach Beweislage verurteilt und befinden sich in Haft. Nun stellt sich nach 10 Jahren heraus, dass die Methode fehlerhaft und unzuverlässig ist. Wie ist die jetzt entstandene Situation zu beurteilen?

Es ist klar, dass die Beweise unzuverlässig sind. Daraus folgt aber nicht notwendigerweise die Unschuld der Verurteilten. Dennoch müssten die Urteile aufgehoben und das Verfahren neu eröffnet werden. Alles muss auf Null gestellt werden; die Verurteilten haben zunächst wieder als unschuldig zu gelten.

McKittrick hat die Autoren der beanstandeten Arbeit mit seiner Kritik konfrontiert. Zunächst erhielt er keine Antwort, später entwickelte sich dann doch ein Austausch, allerdings auf einer nicht-technischen Ebene. Mit anderen Worten: Die Autoren der beanstandeten Arbeit gingen nicht auf die technische Analyse McKittricks ein, insbesondere wiesen sie die Kritik McKittricks nicht in einem technischen Sinne zurück. Stattdessen stellten sie fest, dass diese Kritik mittlerweile irrelevant geworden sei, weil sie die damals angewandte Methode nicht mehr verwenden würden. Außerdem sei es die Aufgabe der Kritiker, nachzuweisen, dass die Ergebnisse der Attributionsforschung falsch seien.

Zur Bewertung dieser Reaktion ist das obige anschauliche Beispiel nützlich: Die Verteidiger der Verurteilten weisen darauf hin, dass die Beweise ungültig sind, weil die zugrundeliegende Methode fehlerhaft und unzuverlässig ist. Das FBI antwortet, die Kritik sei irrelevant, weil man mittlerweile eine andere Methode benutze. Außerdem müssten die

Anwälte die Unschuld der Verurteilten nachweisen. Ich denke, es wird jedem einleuchten, dass eine solche Reaktion des FBI abwegig wäre.

Ich kann mir nicht anmaßen zu entscheiden, ob McKitricks Kritik korrekt ist oder nicht. Allerdings kann auch ein Laie erkennen, dass die Reaktion der Kritisierten inadäquat zu nennen ist. Sie haben eine bestimmte Methode benutzt; McKitrick kritisiert diese Methode und behauptet, sie sei fehlerhaft und unzuverlässig. Es wäre die wissenschaftliche Pflicht der Autoren, diese Beanstandung begründet zurückzuweisen oder sie anzuerkennen. Der Hinweis, mittlerweile verwende man eine andere Methode ist lächerlich. Denn es geht ja um diese eine Frage, ob die verwandte Methode korrekt angewandt worden ist oder nicht. Diese Frage ist in sich berechtigt und sie ist wert, beantwortet zu werden. Die Reaktion der Autoren ist als eine Immunsierungs-Strategie hinsichtlich der eigenen Arbeiten zu betrachten. Denn sie behaupten ja, dass ihre Arbeit immun gegen Kritik sei, weil sie mittlerweile durch neuere Arbeiten obsolet geworden sei.

Diese Art der Immunsierung ist im Rahmen der Klima-Forschung nicht neu. Sie tauchte auch im Zusammenhang mit der sogenannten Hockeyschläger-Kurve auf. Die Kritik an der Datenbasis dieser Kurve wurde neutralisiert, indem man auf neue Verfahren und neue Arbeiten verwies. Man tat so, als wenn die Verwendung neuer Verfahren die Kritik an den alten Verfahren obsolet machen würde. Das Gegenteil ist jedoch richtig. Denn die Kritik an den alten Verfahren ist eine wertvolle Grundlage für die Bewertung der neuen Verfahren.

Es ist für mich vollkommen klar, dass es sich hier nicht mehr um Wissenschaft, sondern um Pseudowissenschaft handelt. Kritik ist das Wesen der Wissenschaft, Immunsierung gegen Kritik ist das Wesen der Pseudowissenschaft.

Es scheint auch so zu sein, dass man in diesem Fall ein Beispiel für den Untergang einer einstmals hochstehenden Wissenschaftskultur sehen kann. Denn für den Fall, dass McKitrick recht hat, haben die Autoren offensichtlich die von ihnen angewandte Methode selbst nicht verstanden oder – was noch schlimmer wäre – sie haben diese Methode bewusst falsch angewandt.

Der Gipfel der Dekadenz ist allerdings darin zu sehen, dass diese Fehler beim IPCC nicht aufgefallen sind. Das IPCC hat diese Methode sogar gefördert und sie ist auf dieser Basis auch vielfach von anderen Wissenschaftlern angewandt worden, ohne dass diesen die Unkorrektheit des Verfahrens aufgefallen wäre.

McKitrick vermutet sogar, dass die jetzt mit dem Nobelpreis gewürdigten Arbeiten Hasselmanns auf demselben Mangel beruhen. Das wäre dann allerdings der Tiefpunkt der wissenschaftlichen Kultur. Denn man muss ja bedenken, dass die von McKitrick beanstandeten Unrichtigkeiten elementarer Art sind, deren Vermeidung in grundlegenden Statistik-Kursen gelehrt wird. Es handelt sich also nicht um einen nachvollziehbaren und verzeihlichen Fehlgriff, der jedem mal passieren kann, sondern um einen dicken Hund, den man eher bei einem Studenten vermuten würde als bei einem examinierten Wissenschaftler.

