

# 가뭄의 기상-수문-사회경제 영향경로 분석을 위한 인과추론(Causality) 방법론 검토

Causal Inference Approaches for Understanding  
Meteorological-Hydrological-Socioeconomic Drought Propagation

성경민

K O R E A  
E N V I R O N M E N T  
I N S T I T U T E

■ 저 자 성경민

■ 연구진

연구책임자 성경민 (한국환경연구원 부연구위원)

■ 연구자문위원 (가나다 순)

김현규 (한국환경연구원 부연구위원)

신주영 (국민대학교 교수)

주경원 (한국과학기술기획평가원 부연구위원)

© 2024 한국환경연구원

발행인 이 창 훈  
발행처 한국환경연구원  
(30147) 세종특별자치시 시청대로 370  
세종국책연구단지 B동(과학·인프라동)  
전화 044-415-7777 팩스 044-415-7799  
www.kei.re.kr  
발 행 2024년 12월 09일  
등 록 제 2015-000009호 (1998년 1월 30일)  
ISBN 979-11-5980-961-3 95530

이 보고서를 인용 및 활용 시 아래와 같이 출처를 표시해 주십시오.  
성경민 (2024), 「가뭄의 기상-수문-사회경제 영향경로 분석을 위한 인과추론(Causality) 방법론 검토」,  
한국환경연구원.

# 요약

## ■ 주요 연구 내용

- 본 연구의 목적은 강수량 부족에서 기인하는 기상학적 가뭄이 수문학적, 사회경제적 가뭄 등 다양한 형태로 사회 및 환경에 전파되는 경로를 통합적으로 이해하는데 있음
  - 특히 전파경로 분석에 인과추론 모델링 방법론의 적용 가능성을 검토함으로써 향후 관련 모델 구축에 대한 기반을 마련하고 전파경로에 대한 정량화에 주목함
- 인과관계 분석을 인과추론과 인과발견 두 가지로 분류하여 설명하고 이에 해당하는 다양한 방법론에 대한 특징 및 장단점을 분석함
  - 기존 기후 관련 연구는 주로 인과발견 분석 방법을 사용하였으나 본 연구는 탐욕적 동등성 탐색(GES) 등 새로운 인과추론 방법의 적용 가능성을 제공했다는 데 의의가 있음
- 기존 인과관계 분석 기법을 이용하여 가뭄의 전파경로를 분석한 연구들을 검토하고 이를 기반으로 새로운 연구 방향을 제시함
  - 전 세계적으로 위성 관찰 데이터, 미래 시나리오 데이터 등 자료량의 증가 및 통계, 머신러닝 개발 등 기술 개발에 맞춘 고도화된 가뭄 분석 연구의 가능성을 제시함

## ■ 정책 제언

- 인과추론 방식을 기반으로 기상-수문-사회경제로의 가뭄 전파 과정에서 취약한 경로를 파악함으로써 이를 단계 맞춤형 대응 정책 수립에 기초 근거로 활용할 수 있음
- 물 부족 원인의 핵심 요인을 식별하고 중요성을 부여함으로써 제한된 자원을 효과적으로 배분하기 위한 정책의 우선순위를 결정하는 데 기초자료로 활용 할 수 있음
- 가뭄의 복합적인 영향에 대한 이해 증진을 통해 농업, 수자원, 환경, 경제 등 여러 분야를 아우르는 통합적인 가뭄 대응 체계를 구축할 수 있는 정책을 제안할 수 있음

주제어: 기상학적 가뭄, 수문학적 가뭄, 사회경제적 가뭄, 인과관계 추론, 가뭄 전파경로



# 차 례

I. 가뭄 전파경로 연구의 배경 및 필요성 .....	1
1. 가뭄 전파경로 연구의 배경 .....	1
2. 인과관계 적용의 필요성 .....	5
3. 연구 목적 .....	6
II. 통계 및 머신러닝 인과분석 방법 조사 .....	7
1. 인과발견(causal discovery) .....	7
2. 인과추론(causal inference) .....	10
3. 인과분석 방법론 조사 .....	12
III. 가뭄 전파경로 분석 방법론 조사 .....	20
1. 가뭄 전파 과정 분석 .....	20
2. 인과추론을 적용한 가뭄 전파 연구 .....	24
IV. 향후 연구 방안 .....	30
1. 인과추론 모델 개발 시 고려 사항 .....	30
2. 향후 모델 개발 방안 .....	31
참고문헌 .....	33



# I 가뭄 전파경로 연구의 배경 및 필요성

## 1. 가뭄 전파경로 연구의 배경

- 가뭄은 한 계절 또는 그 이상에 걸친 기간 동안 지역의 정상적인 상태에 비해 수자원이 부족해지는 현상을 말함
  - 오랜 시간에 걸쳐 광범위한 지역에서 발생하여 시작 및 종료 시점과 발생 지역을 예측하기 어렵다는 특성이 있음
  - 지역의 기후, 생태계, 경제 특성 등에 따라 다양한 양상으로 나타나며 식량 생산 피해, 경제손실 등 다양한 피해를 초래
  - 제3차 기후변화 적응을 위한 국가 대책에서 가뭄은 건강, 생태, 국토, 농수산 등 다양한 부문에서 주요 위해성 요소로 나타남<sup>1)</sup>



자료: 국가가뭄정보포털, “[쉽게보는가뭄] 가뭄이란/특징/영향”, 검색일: 2024.11.1.

〈그림 1-1〉 가뭄의 종류와 정의

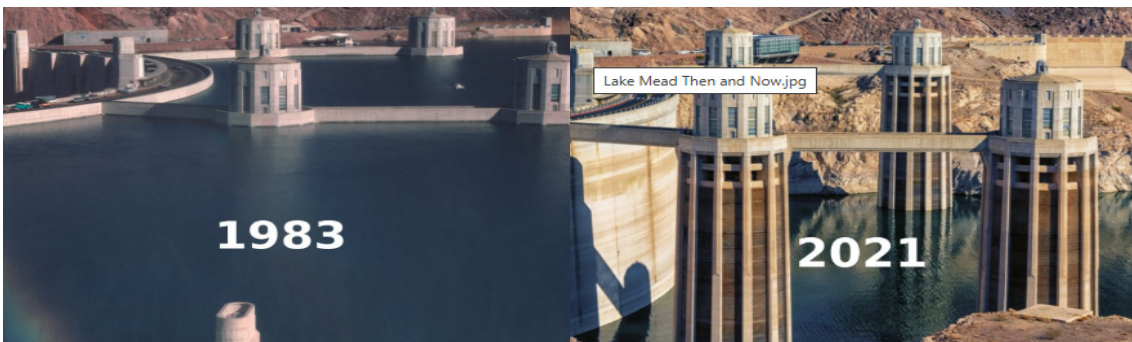
- 가뭄은 일반적으로 강수량 부족에서 기인하나 물 부족이 발생하는 원인과 피해 요소에 따라 더욱 다양한 유형으로 분류됨
  - (기상학적 가뭄) 지역의 강수량 또는 강수량과 증발량의 차이가 정상 상태의 강수량보다 적은 기간이 오랫동안 지속되는 현상

1) 관계부처합동(2023).

- (수문학적 가뭄) 하천, 호수, 지하수 등 수자원 유출량이 평상시보다 감소한 현상을 의미하며 주로 기상학적 가뭄의 장기화로 야기됨
  - (사회·경제적 가뭄) 물 부족이 인구와 경제활동에 영향을 미치는 현상을 의미하며 물 수요가 공급을 초과할 때 나타남
  - (농업적 가뭄) 토양수분 부족으로 인해 농작물의 정상적인 생장이 저해되는 현상을 의미하며 강수량 부족 외에 토양 상태, 작물 종류 등이 복합적으로 작용함
- 기후변화로 인해 지역의 기후 특성이 변화하고 가뭄의 원인 및 영향의 복잡성이 증가함에 따라 대응이 더욱 어려워지고 있음
- 기존 연구는 기후변화가 가뭄의 강도 및 빈도 특성 변화에 영향을 미쳤으며 그 변화 양상이 지역 및 계절별로 상이함을 밝힘
    - 전 세계적으로 건조한 지역 또는 기간이 더욱 건조해지는 가뭄의 강화 현상이 나타남<sup>2)</sup>
    - 대한민국에서도 남부지방에 역대 최장기 가뭄(2022년 봄철- 2023년 봄철) 발생이 발생하는 등 기후변화로 인한 전례 없는 가뭄을 경험함<sup>3)</sup>



자료: KBS(2022.12.31) “[굿바이2022] 최악의 가뭄..섬 식수난·광주는 제한급수 걱정”, 검색일: 2024.11.24



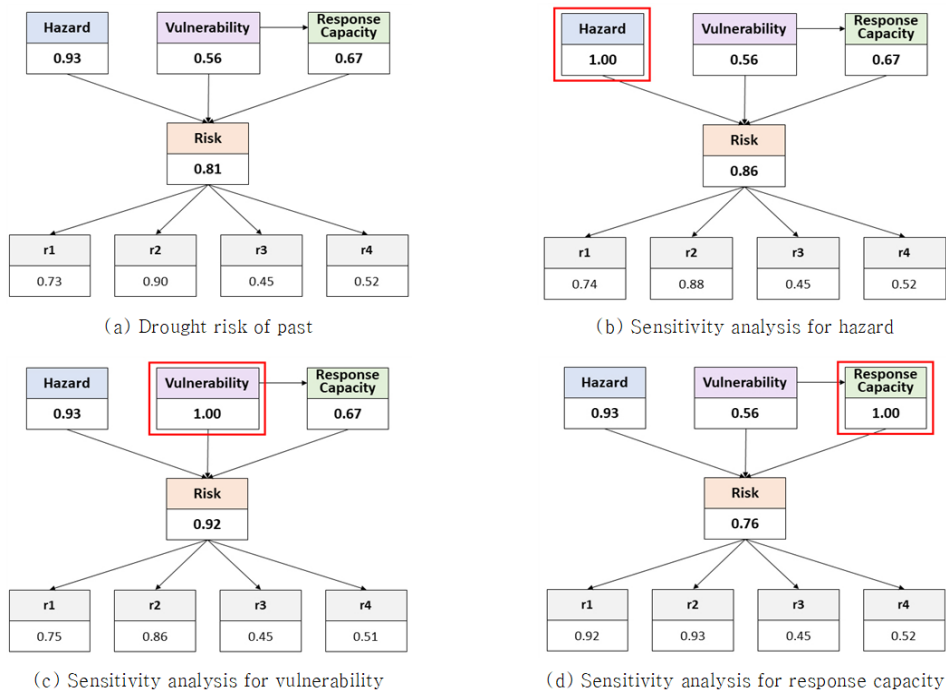
자료: City of Las Vegas(2022.8.16), 검색일: 2024.11.24.

〈그림 1-2〉 2022년 전남지역(상)과 미국 Mead 호수의 강수량 부족으로 인한 용수 고갈

2) Sung et al.(2024).

3) 환경부 보도자료(2024.4.29), “기상청, 정부합동 2023년 이상기후보고서 발간”.





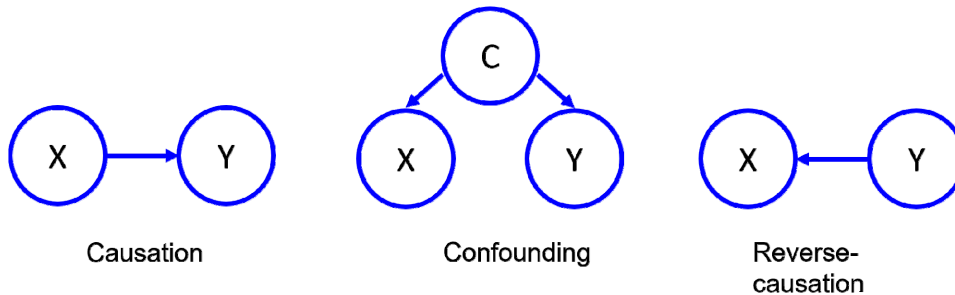
자료: 김지은 외(2021).

〈그림 1-4〉 청주시의 가뭄 위험도 평가 및 위해성, 취약성, 대응능력 변화에 따른 가뭄 영향 인자들 사이의 민감도 분석

- 가뭄의 확산은 주로 강수량의 부족에서 시작하여 수문 시스템, 토양, 식생, 사회 전반의 물 부족 현상으로 전파되는 매우 복잡한 수문순환의 물리학적 과정을 의미함
  - 가뭄의 전파 형태는 유역 특성, 강수 형태, 대기 습도, 온도 등 수많은 변수에 의해 좌우되어 그 이해와 예측이 매우 어려움
  - 기후변화로 인하여 가뭄의 시작인 강수량의 변동성이 심해짐에 따라 수문시스템으로 전이되는 경로에 대한 불확실성이 더욱 증가함
- 현재까지 가뭄의 영향에 대한 연구가 다양하게 이루어졌으나 여전히 전파경로에 대한 연구는 초기 단계에 머물러 있음
  - 수문학적 가뭄에서 사회경제적 가뭄으로 전파되는 현상에 대한 연구가 매우 부족함
- 기술의 발달에 따라 위성 관측을 비롯한 기록 데이터가 증가하면서 데이터 분석 기반의 가뭄 특성 연구의 가능성이 넓어짐
  - 선행 연구에 대한 폭넓은 이해를 기반으로 신기술을 접목하는 연구 방향에 대한 추후 논의가 강조됨

## 2. 인과관계 적용의 필요성

- 단순 연관성 분석은 가뭄의 자연적, 사회적, 경제적 요인의 복잡한 전파경로를 이해하는데 한계가 있음
  - 요인 간의 연관성(association) 분석은 두 요인의 상호작용을 밝힐 수 있으나 두 변수 간의 인과관계를 추론(inference)할 수는 없음
  - 즉 연관성은 인과성을 보장하지 않음. 두 변수가 연관되어 있더라도 독립변수와 종속변수는 가짜 상관관계(spurious correlation)로 연결되어 있을 수 있으며, 이는 독립변수에 개입하더라도 종속변수에 영향을 미치지 않을 가능성 시사함(그림 2-1 참조)<sup>6)</sup>
    - 예) 아이스크림 판매량과 익사사고 발생 건수는 모두 여름철에 증가하므로 두 변수 사이에 상관관계가 존재하나 두 요인 사이에 인과관계는 존재하지 않음
  - 또한 독립변수가 종속변수에 실제 인과적 영향을 미치는 경우에도 다른 변수로부터 공변량(covariate)의 영향을 받으면 개입 변수의 인과적 효과가 왜곡될 수 있음 (confounding, 그림 2-1 참조).
  - 연관성 분석을 인과추론에 적용하는 경우, 원인-결과에 대한 잘못된 판단, 즉 가짜 상관관계(spurious correlation)로 인하여 추후 잘못된 의사결정을 초래할 수 있음



자료: 저자 작성.

〈그림 1-5〉 변수 간 상관관계의 구조

- 따라서 정확한 의사결정을 위한 가뭄의 원인-결과 관련 변수 식별 및 관계 추론에 인과적 분석(causal reasoning)이 필수적임
  - 예를 들어 특정 지역에서 농업활동이 물 부족에 미치는 영향, 그 반대로 물 부족이 농업

6) 가짜 상관관계(Haig, 2003) - 독립변수와 종속변수 간에 상관관계가 있지만 실제로는 인과관계가 없는 경우를 의미함. 교란변수(confounding variable)와 역인과(reverse causation)는 이러한 가짜 상관관계의 원인이 될 수 있음.

- 활동에 미치는 영향이 상호 연관되어 있는지 영향의 방향성을 식별 가능
- 특정 요인의 변화 등 가정된 상황에 대한 결과 예측이 가능해짐으로써 배수시스템 개선 등의 정책이 가뭄 심각성 완화에 기여하는 효과를 예측할 수 있게 됨
  - 따라서 가뭄 근본 원인을 파악하고 이를 대상으로 대응 정책을 도출하여, 이에 대한 효과를 예측하는 데 기여
  - 앞서 언급한 교란변수(confounders) 통제를 통한 분석이 더욱 명확하게 가능해짐에 따라 잘못 적용하는 오적응 문제를 해결할 수 있음. 예를 들어 특정 지역의 도시화가 물 부족에 영향을 미치는지 면밀하게 관계를 식별하여 도시화에 대한 정확한 판단이 가능
  - 복잡성이 강한 수문-사회 시스템에 대하여 시간과 공간 데이터의 인과추론 분석을 실시하여 기존 가뭄 전파 추론 연구를 발전시키는 데 학문적으로 기여할 수 있음

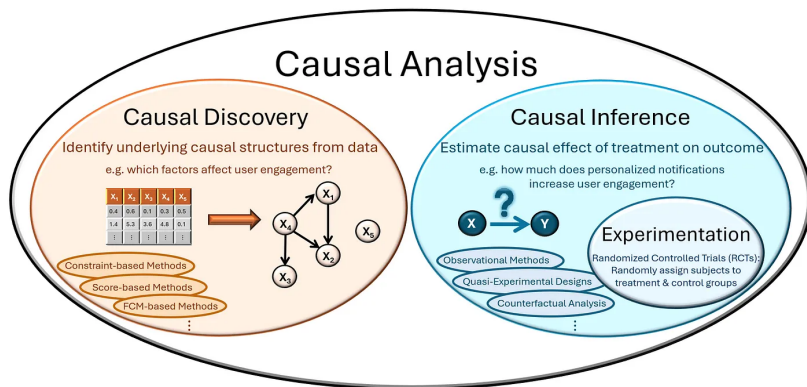
### 3. 연구 목적

- 기상-수문-사회경제적 가뭄의 전파경로에 대한 통합적 이해를 증진하고 정량화 방안을 논의
  - 국내외 가뭄 전파경로 및 전파 시간 분석과 관련된 학술 논문과 보고서 검토로 관련 연구에 대한 이해도를 향상
  - 기존에 주로 연구되어 있는 기상-수문학적 가뭄 관계를 기반으로 아직까지 연구가 많이 이루어지지 않은 사회경제적 가뭄을 통합한 분석에 대한 가능성 검토
- 가뭄 전파경로 분석에 인과추론 방법론을 적용할 수 있는지 검토하고 향후 모델 구축에 대한 기반을 마련
  - 다양한 데이터를 기반으로 인과추론 모델링 기법이론과 기능을 파악
  - 가뭄 간의 영향 인자 식별 및 관계 정량화에 인과추론 방법을 적용할 수 있는지 검토

# II 통계 및 머신러닝 인과분석 방법 조사

## 1. 인과발견(causal discovery)

- 인과발견의 목적은 관찰 데이터의 변수 간 인과관계 유무 및 그 구조 규명, 즉 인과 변수를 식별하는 데 있음(Zanga et al., 2022)
  - 처리 변수의 효과 추정보다 가짜 상관관계와 실제 인과 구조를 구별하는 데 초점을 맞추며, 밝혀진 인과 구조는 이후에 인과 모델 개발 등 다양한 목적으로 활용이 가능함
  - 특정 질병의 잠재적 인과 변수를 식별(Corander et al., 2022; Shen et al., 2020)하거나 인간-로봇 상호작용의 메커니즘 분석(Castri et al., 2022)에 적용
  - 데이터 기반 통계 및 머신러닝 접근법과 실험적 접근법이 존재하나 본 연구의 주제를 고려하여 데이터 기반 접근법만을 고려



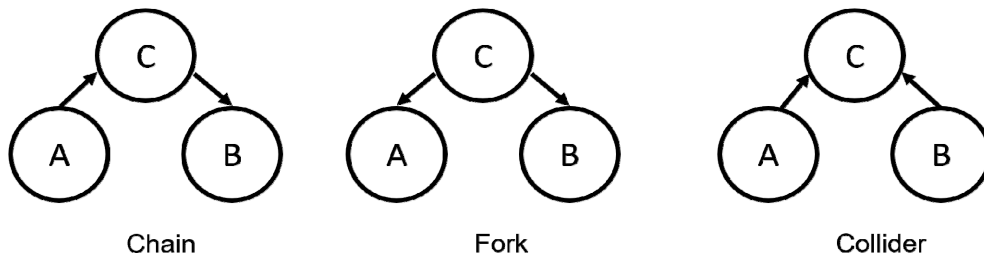
자료: Meghanath G.(2024, 11.15), Causal analysis overview: Causal inference versus experimentation versus causal discovery. 검색일: 2024.11.30.7)

〈그림 2-1〉 데이터 기반 인과분석 방법론: 인과발견과 인과추론 차이

- 인과관계는 〈그림 2-2〉와 같이 크게 세 가지 구조로 구분되며 그림상에서 각 점(node)은 변수를 나타내고 연결선(edge)은 인과관계를 의미함

7) Meghanath(2024)는 “Causal Analysis Overview”를 통해 인과관계 유무를 규정하는 인과발견과 인과관계의 구조를 살피는 인과추론과의 차이를 설명하고 실험 방식과 데이터 기반 방식의 차이를 설명함.

- 인과발견은 제약 기반, 점수 기반, 함수 기반의 세 가지 접근 방식으로 분류함(Glymour et al., 2019)



자료: 저자 작성.

〈그림 2-2〉 데이터 내 변수들 간의 인과관계를 시각적으로 표현한 인과 그래프(causal graph)

#### 가. 제약 기반 인과발견 접근법

- 변수 간 조건부 독립성 검정을 활용해 잠재적 인과 그래프 구조를 추론하는 접근법으로 인과관계가 조건부 독립성 검정 결과의 집합으로 설명될 수 있다는 가정에 기반
  - 체인(chain) 및 포크(fork) 구조에서 A와 B는 C가 매개 변수(mediator) 또는 교란 변수(confounder)로 작용하기 때문에 서로 종속관계(dependent)이나 C에는 조건부 독립(conditional independent)관계를 보임(그림 2-2참조)
  - 체인과 포크 구조는(i) A와 B는 무조건적으로 종속적이고(ii) C에 대해 조건부 독립적이 라면 A와 B 사이에 인과관계가 없다는 독립성 검정 결과로 표현됨
  - 이러한 조건부 독립성이 기각되면 A와 B 사이에 인과관계를 발견하는 방식으로 인과관계를 추론
- 인과발견의 결과가 조건부 독립성 검정의 정확도에 의존한다는 한계가 존재
  - 숨겨진(confounding) 변수에 대한 분석이 부정확한 경우 오류가 전파할 수 있다는 위험 존재
  - 대부분 변수 데이터의 선형성(linear)을 가정하므로 복잡한 비선형 관계(non-linear)를 다루기에는 한계가 있음
- 그래인저 인과관계(Granger Causality), PCMC(PC-based method for Momentary Conditional Independence)가 대표적인 제약 기반 인과관계 접근법에 속함
  - 그래인저 인과관계와 PCMC에 대해서는 본 보고서 제 3장에서 자세히 서술함

#### 나. 점수 기반 인과발견 접근법

- 주어진 인과 그래프 구조가 데이터를 얼마나 잘 설명하는지를 평가하기 위해 특정 점수를 계산하고 이를 최적화하는 방식으로 인과관계를 도출하는 방법
  - 점수 산정 방식은 변수 간의 함수 관계(예: 선형 가우시안)와 데이터 유형(예: 이산형, 연속형)에 따라 결정되며 BIC(Bayesian Information Criterion), BGe(Bayesian metric for Gaussian networks) 등이 사용됨
  - 가능한 모든 그래프 구조를 가정한 후 각 구조에 대한 점수를 계산
  - 인과 구조별로 점수를 계산하여 최대 점수를 얻는 인과 그래프를 최적의 인과구조로 간주
- 점수 기반 접근법 알고리즘으로는 탐욕적 동등성 탐색(GES:Greedy Equivalence Search) 방식이 있음
  - 점수를 끊임없이 관찰하면서 그래프 구조를 점진적으로 개선하는 방식으로, 빈그래프에서 시작하여 점수가 개선되는 방향으로 연결선을 계속 추가함
  - 점수 개선이 멈추면 불필요한 에지를 제거하는 방식을 통해 점수가 가장 높을 때 나타나는 최적의 그래프를 도출
- 탐색된 그래프 공간에 인과 그래프가 존재하지 않으면 최적의 그래프를 찾지 못할 수 있다는 것이 점수 기반 접근법의 대표적인 한계임
  - 최대 점수를 가진 그래프가 반드시 실제 인과 그래프를 나타낸다고 보장할 수 없으며 데이터 노이즈 또는 다른 외부 변수로 잘못된 구조가 도출될 가능성이 있음
  - 최적화 과정을 통해 인과 그래프를 추론하므로 추론된 결과가 매우 복잡한 경우 직관적인 해석이 어려움

#### 다. 함수 기반 인과발견 접근법

- 변수 간의 가정된 관계를 탐구해 인과관계를 추론하는 방식으로 데이터의 분포 형태를 가정하고 독립변수와 오류 항 간의 독립성을 검토하여 인과관계의 방향을 결정함
  - 대표적인 방법으로 선형 논가우시안 비주기형 모델이 있음(LiNGAM:Linear Non-Gaussian Acyclic Model)
- LiNGAM은 대표적인 함수 기반 인과발견 방법으로 두 변수 간의 관계는 아래와 같이 정의함

$$Y := f(X) + b \quad \text{식(2-1)}$$

- X와 Y는 비가우시안 분포를 따르는 변수로,  $f(X)$ 는 선형 함수,  $b$ 는 오류 항을 나타냄
  - X와  $b$ 가 독립적이라면, X에서 Y로의 잠재적 인과관계가 존재하며, 반대로 X와  $b$ 는 종속관계이면 Y가 X를 유발한다고 추론함
  - 이 추론은 X와 Y가 비가우시안 분포를 따를 때에만 적용이 가능하며 가우시안 분포를 따를 경우 이 방법으로 인과관계 식별이 불가능
- 함수 기반 접근법은 특정 함수 및 분포 가정에 의존하며 관측 데이터가 특정 함수를 따르지 않으면 적용이 어렵다는 한계가 존재

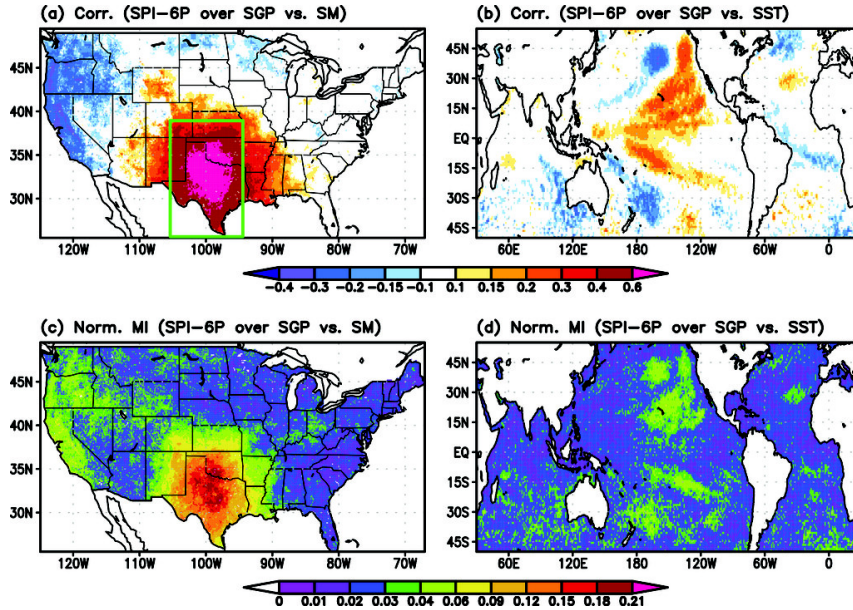
## 2. 인과추론(causal inference)

- 인과추론은 처리변수(treatment variable)가 결과변수(outcome variable)에 미치는 인과적 효과 추정에 초점을 맞춤
- 다른 변수로부터의 공변량이 효과 추정에 편향을 초래할 수 있는 상황에서도 독립변수의 독립적인 인과 효과(direct causal effect)를 보다 정확하게 측정하는 것을 목표로 함
  - 예를 들어 특정 약물이 질병에 미치는 효과 추정(Yazdani & Boerwinkle, 2014), 보건 정책이 인구에 미치는 장기적 영향 평가(Zigler et al., 2016) 등에 활용됨

### 가. 상호 정보량(mutual information)

- 정보 이론(information theory)은 어떤 사건의 발생에 대한 불확실성을 '정보량'이라는 개념으로 정량화하며, 더 불확실한 사건일수록 더 큰 정보량(단위: bit)을 갖는 것으로 정의함
- 한 시계열에서 이산확률변수로 표현되는 여러 사건이 발생하는데, 모든 사건의 정보량에 대한 기댓값은 섀넌 엔트로피(Shannon entropy)로 계산됨(Shannon, 1948)

• 섀넌 엔트로피:  $H(X) = - \sum_x p(x) \log_2 p(x)$



자료: Shin, Dirmeyerm and Huang(2023)<sup>8)</sup>

〈그림 2-3〉 상호 정보량의 가뭄 연구 적용 사례

○ 두 시계열  $X = \{x_1, x_2, \dots, x_n\}$ 와  $Y = \{y_1, y_2, \dots, y_n\}$ 의 인과관계 분석 방법에 새넨 엔트로피를 적용하여 시계열간의 상호 정보량을 분석

- 시계열  $X$ 의 불확실성에서  $X$ 에서만 발생하는 불확실성을 제거했을 때의 정보량, 즉 한 시계열에서 다른 시계열에 대해 얻을 수 있는 정보의 양을 의미하므로  $I[X, Y]$ 는 결국 두 시계열 간의 상관관계를 나타냄

•상호 정보량(mutual information):  $I[X, Y] \equiv H(X) - H(X|Y)$

- 새넨 엔트로피를 적용한 상호 정보량은 아래와 같이 표현됨

$$I[X; Y] = \sum p(x, y) \log_2 \frac{p(x, y)}{p(x)p(y)}$$

- 상호 정보량은 두 변수 사이의 직접적인 인과관계보다 상관관계를 나타내므로 관계의 방향성을 설명할 수 없다는 한계가 있으나 다른 인과추론 결과 검증에 널리 활용됨
- 가뭄을 비롯한 여러 기후 인자 간의 상호작용 연구에 다양하게 적용됨(그림 2-3 참조)

8) Shin, Dirmeyer, and Huang(2023)은 가뭄과 육지와 해상 요인(driver) 간의 인과관계를 Correlation과 Mutual Information를 이용하여 비교 분석하였으며, 〈그림 2-3〉은 가뭄지수와 토양수분 이상치(좌) 전 세계 해수면온도(SST, 우)와의 상관관계(a, b)와 상호 정보량(c, d) 간의 비교 결과를 나타냄.

### 3. 인과분석 방법론 조사

#### 가. 전송 엔트로피(transfer entropy)

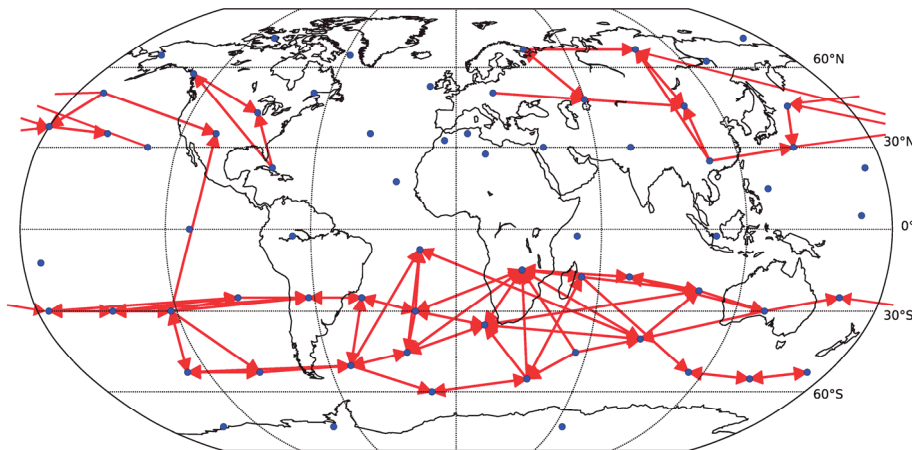
○ 앞서 설명한 새넨 엔트로피를 기반으로 변수  $X$ 의 과거 정보가 변수  $Y$ 의 미래에 대해 변수  $Y$ 의 과거 정보 이상의 추가적인 정보를 제공할 때, 변수  $X$ 가 변수  $Y$ 에 인과관계를 갖는다고 정의함

- 전송 엔트로피 모델:  $T_{X \rightarrow Y} = \sum_{x_t, y_t, y_{t+1}} P(x_t, y_t, y_{t+1}) \log_2 \frac{P(y_{t+1} | y_t, x_t)}{P(y_{t+1} | y_t)}$

- 위 모델은 정의된 시간 단계(time step)에서 한 단위의 지연(lag)만이 있는 경우를 나타내지만 필요에 따라 그 이상의 시간 지연에 대한 인과관계를 확인할 수 있음

○ 국가 간 분쟁 문제(Kushwaha and Lee, 2023), 뇌과학(Vicente et al., 2011), COVID-19(Delussu, 2023), 기후 과학(Hlinka et al., 2013; Bhaskar et al., 2017) 등 여러 주제에서 활발히 활용됨

- 인과관계 분석의 정확도를 향상하기 위해 연구마다 전송 엔트로피 계산 모델을 상황에 맞게 수정 및 개선하는 등 다양한 시도를 하고 있음



자료: Hlinka et al.(2013).

〈그림 2-4〉 전송 엔트로피를 이용하여 전 세계 기후 네트워크 사이의 연관관계 분석<sup>9)</sup>

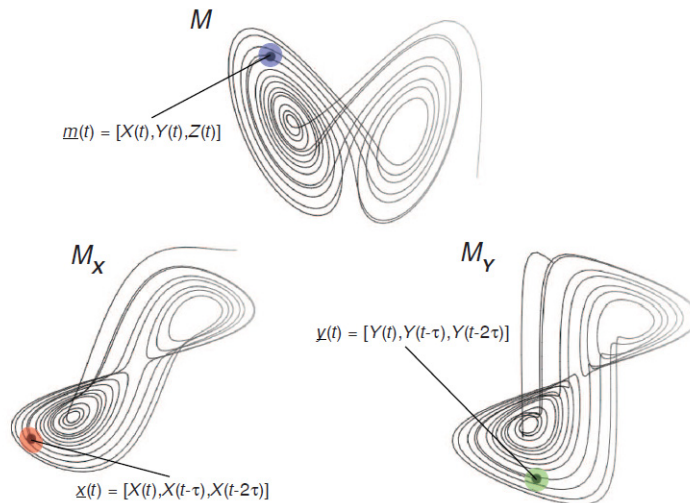
9) Hlinka et al.(2013)은 수십 년에 걸친 기후 네트워크의 변동성 변수들 간의 인과성 중에서 가장 강한 100개의 연결을 식별하였으며, 대부분의 네트워크는 양방향으로 상호작용을 주고받는 것으로 밝힘.

- 전송 엔트로피는 단순히 인과관계의 유무 파악뿐 아니라 전달되는 정보량의 정량화를 통해 인과 크기를 도출할 수 있으므로 인과관계 추론에 속하며 다양한 장점을 지님
  - 인과관계에서 시간 지연(temporal lag)과 방향성(directionality)을 고려할 수 있어 더욱 상세하고 정교한 인과추론이 가능
  - 모델에 의존하지 않기 때문에(model-free) 비선형적이고 일반화된 시계열 분석도 가능하다는 강점을 지니며 다른 정보 이론 방법과 결합하여 다양한 연구로 확장가능
  - 두 시계열을 무작위로 여러 번 섞은 후 전송 엔트로피를 반복 계산하는(몬테 카를로, Monte-Carlo simulation) 방식으로 p-value를 계산하여 인과관계의 통계적 유의성을 확인할 수 있음
- 데이터 크기에 의존적이기 때문에 짧은 시계열 데이터에 적합하지 않고 두 변수 간의 인과관계를 정보량으로만 판단하므로 관계가 양(positive)인지 음(negative)인지 판단할 수 없는 한계가 존재

#### 나. 수렴 교차 매핑(CCM: Convergent Cross Mapping)

- 시계열이 비선형 다이내믹 시스템에서 움직일 때 두 변수 시계열 간의 인과를 추론하는 방법론으로, 복잡한 생태계를 이해하기 위해 Sugihara et al.(2012)가 개발함
  - 비선형적이고 복잡한 상호작용에도 적용이 가능하여 생태계, 기후시스템 등 현실에 흔히 존재하는 비선형 복잡계(nonlinear complex system) 분석에 적합
  - 인과관계 추정 및 관계의 방향성과 발생 시점의 지연(time lag) 분석이 가능하여 관계를 다각도로 이해할 수 있음
  - 단순히 과거 정보로 추론하는 것이 아니라 어트랙터 공간을 바탕으로 인과관계를 추론하기 때문에 복합적인 인과관계도 추론이 가능함
  - 적절한 사전 데이터 처리 시 노이즈와 무작위 확률이 큰 데이터에도 적용이 가능하며 짧은 시계열도 처리가 가능함(Clark et al., 2015)
- CCM은 분석 실행 전에 사전 데이터 처리부터 실제 분석까지 다음과 같은 과정으로 이루어짐(Medina et al., 2019)
  - 사전 데이터 처리를 통해 노이즈(noise), 확률적 시그널(stochastic signal) 및 비정상성(nonstationary)요인을 제거하여 복잡계의 구조적 특성 및 상태 재구성의 정확도 향상을 추구

- CCM은 두 시계열 변수  $X$ 와  $Y$ 가 존재할 때, 변수  $X$ 가  $Y$ 에 인과적 영향을 미치기 위해  $Y$ 의 상태공간(state space)에  $X$ 에 대한 정보가 구조적으로 관련성을 가져야 한다고 정의
  - 구조: 시간에 따른 변수의 상태(state)를 담는 공간으로 어트랙터(attractor)라고 불림
- 올바른 CCM 분석을 위해 각 변수의 어트랙터는 먼저 시간 지연 임베딩(time delay embedding)을 통해 재구성(reconstruct)한 후 Taken's Theorem을 이용해 일대일로 매핑함(1-to-1 mapping)<sup>10)</sup>
- 변수  $X$ 가 변수  $Y$ 에 임베딩(embedding)이 될 경우, CCM은  $X$ 가  $Y$ 에 인과적 영향을 준다고 정의됨



자료: Sugihara et al.(2012).

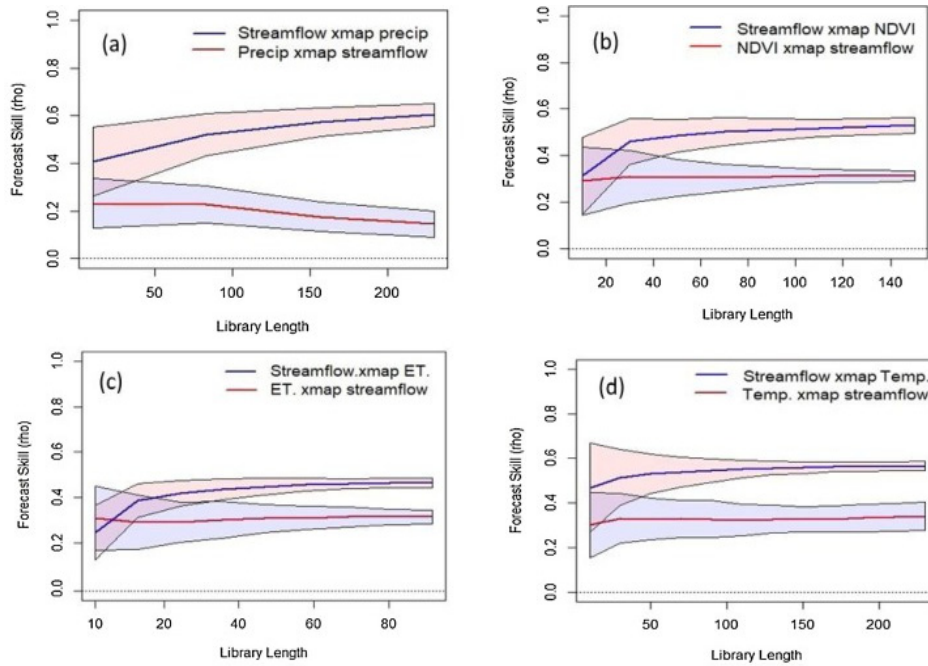
〈그림 2-5〉 CCM의 이론적 바탕에 대한 설명<sup>11)</sup>

○ CCM은 앞서 언급한 여러 강점으로 인해 최근 다양한 분야에서 주요 인과추론 방법으로 활발히 사용되고 있음

- 생태 시스템(Sugihara et al., 2012; Medina et al., 2019), 기후 과학(Zhang et al., 2021; Zhao et al., 2021), 수자원 공학(Wang et al., 2018; Kadir et al., 2020), 경제학(Huffaker and Fearne, 2019) 등의 분야에서 활발하게 사용

10) Takens(1981)가 고안한 것으로, 시간에 따라 동적으로 움직이는 시스템에서 변수들의 상태공간을 정의함

11) Sugihara et al.(2012)은 논문을 통해 CCM의 이론적 바탕에 대해 설명함. 그림에서 M은 X, Y, Z 변수가 모두 결합되어 있는 어트랙터이고, Mx, My는 각각 X와 Y에 대해서만 재구성된 어트랙터를 의미함. 색이 있는 점들은 일대일로 매핑된 지점을 나타냄.



자료: Kadir et al.(2020).

〈그림 2-6〉 CCM을 이용한 강수량, 하천, 식생, 하천유량, 온도 간 인과관계 확인

#### 다. PCMCI(Peter and Clark Momentary Conditional Independence) 검정

○ PCMCI는 Jacob Runge 등이 개발한 방법으로 비선형이거나 고차원 데이터의 인과관계를 효과적으로 분석 할 수 있음

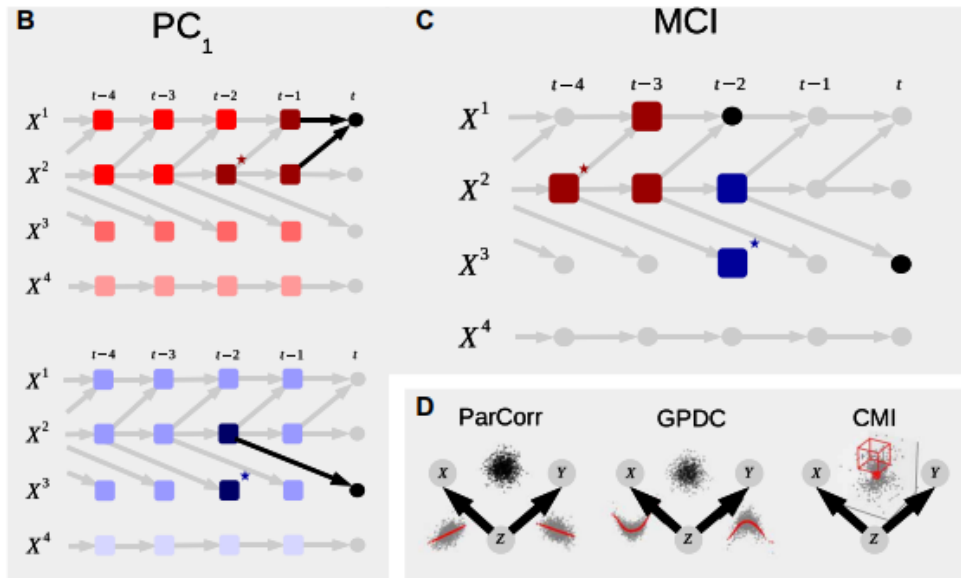
- 다양한 형태의 변수 간 상관관계를 명확히 구분하여 거짓 인과관계를 쉽게 배제할 수 있다는 강점이 존재

- 각 단계가 독립적으로 구성되어 있어 상황에 따라 다양한 검정 방법을 적용할 수 있는 유연성을 가짐

- 인과관계의 크기 또한 다양한 방식으로 계산할 수 있으므로 다각도의 해석이 가능

○ 지연 시간의 설정에 따라 결과가 민감하게 반응한다는 한계가 있으므로 복잡한 시간 지연 구조를 띠는 인과관계 분석에 적합하지 않음

- 직접적인 인과관계는 감지하지만 간접적인 인과관계 식별은 불가능하므로 다른 방법과의 병행을 통한 보완이 필요(Docquier et al., 2024)



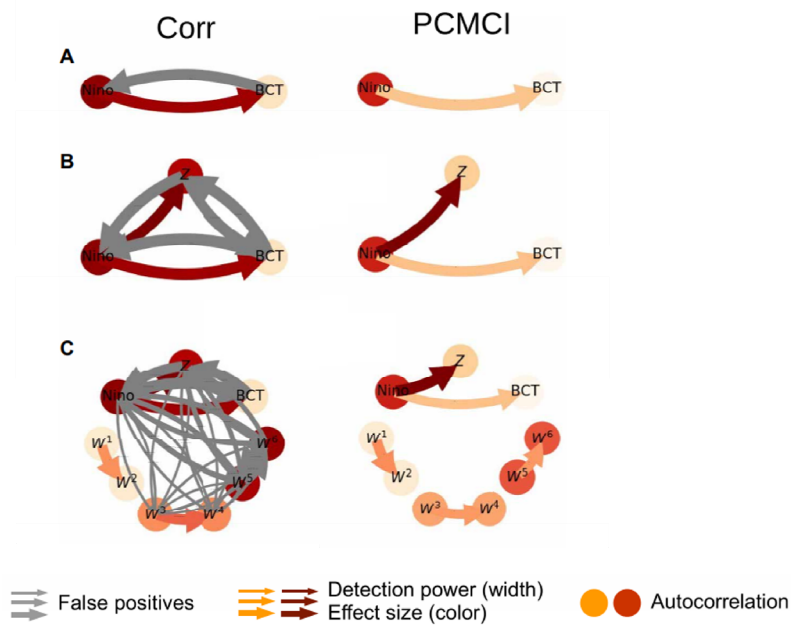
자료: Runge et al.(2019b).

〈그림 2-7〉 PCMCI의 두 단계 개념도

- 두 단계의 분석 과정, 즉 PC(Peter and Clark) 알고리즘과 MCI(Momentary conditional independence) 검정이 합쳐진 방법으로 그 과정은 아래와 같음(그림 2-7 참조)
  - PC 검정은 처음 변수  $X_t^i$ 를 통해 존재 가능한 모든 인과관계 집합인  $t\{RHO\} (X_{t-\tau}^j)$   $\hat{P}(X_t^i)$  설정 후 첫 시행에서 의존성 없는 인과관계들을  $\hat{P}(X_t^i)$ 에서 제거
  - 다음 시행에서 가장 의존도가 높은 변수  $X_{t-\tau}^i$ 을 찾아서  $X_t^i$ 와 조건적 독립 관계임을 확인한 후 독립적일 경우  $\hat{P}(X_t^i)$ 에서 제거하며, 이 과정을 조건적 독립 관계를 확인할  $X_{t-\tau}^i$ 이 남아 있지 않을 때까지 반복
  - 최종  $\hat{P}(X_t^i)$ 에는 인과적 관계뿐 아니라 거짓 양의(false positive) 인과관계가 포함될 수 있으므로 언급한 알고리즘이 완료된 후,  $\hat{P}(X_t^i)$ 에 대해 MCI 검정을 시행
  - MCI 검정은  $X_{t-\tau}^j \rightarrow X_t^j$ 의 인과관계에 대해 두 변수의 조건부 독립성을 확인함으로써 자기상관관계를 비롯한 거짓 양의 관계를 찾아 최종적으로  $\hat{P}(X_t^j)$ 에서 제거
    - 두 변수의 조건부 독립성:  $X_{t-\tau}^i \not\perp\!\!\!\perp X_t^j | \hat{P}(X_t^j) \setminus X_{t-\tau}^i, \hat{P}(X_{t-\tau}^i)$
  - PC와 MCI 두 단계를 거쳐 구한 인과관계는 MCI 통계를 이용해 관계의 크기를 정량화할 수 있음
- PCMCI는 비교적 최근에 개발되었으며 주로 기후 과학에서 복잡 기후 네트워크 (complex

climate network)를 분석하는데 사용함(Runge et al., 2019; Krich et al., 2020; Krich et al., 2022; Ludescher et al., 2021)

- 현재는 이주 패턴(Tárraga et al., 2024), 뇌과학(Silfwerbrand et al., 2024) 등 다양한 분야로 활용성이 확장되고 있음
- Runge et al.(2019a)은 PCMCI를 기후 분야에 적용하여 Nino와 British Columbia 지역 사이의 공기온도(그림 2-8, A 참조)와 임의 변수인 Z, W'(그림 2-8, B와 C 참조)의 관계성을 분석하여 PCMCI가 상관관계 분석에 비하여 높은 Detection Power를 가지고 있음을 밝힘



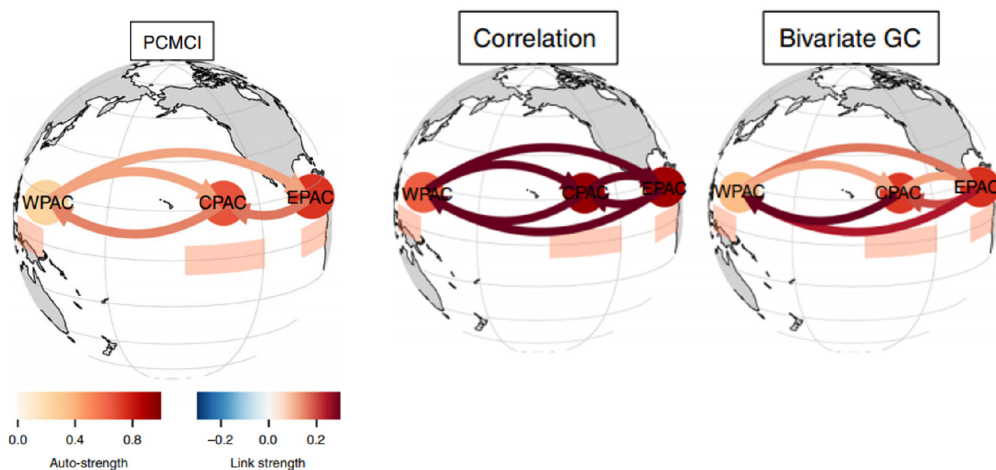
자료: Runge et al.(2019a), <그림 2> 에서 발췌.

<그림 2-8> PCMCI 방법론 기후 영향 적용 연구 예시

### 라. 그레이저 인과관계(GC: Granger causality)

- 변수 X가 지연된 시간 후의 변수 Y에 대해 예측력을 제공하는지 평가하기 위한 통계적 기법으로 인과관계 규명보다 두 변수의 시간적 선후관계를 밝히는 데 유용하게 활용됨
- 변수 Y의 시계열 데이터에서 X의 과거 정보를 제외했을 때 Y의 다음 시간 단계에 대한 예측 오차가 증가하는지 확인하여 인과성을 검증함

- Y의 예측 오류가 유의미하게 증가하면 X는 Y에 대해 그래인저 인과성이 있다고 간주
- Runge et al.(2019b)은 또 다른 연구에서 이변량 그래인저 인과관계를 기후과학에 적용하여 다양한 지구상 기후변수의 인과관계를 나타냄
  - 〈그림 2-9〉는 열대 기후를 예시로 서태평양(WPAC)의 월별 지표 기압 이상치(pressure anomaly)와 중앙 태평양(CPAC), 동태평양(EPAC)의 지표 온도 이상치(temperature anomaly) 간의 의존 관계를 분석한 사례를 보여 줌
  - 이변량 GC 및 단순 상관성 분석은 지역 간 관계성을 매우 높게 계산하고 동태평양과 서태평양 간의 직접 연관성을 추출한 반면, PCMCI는 실제 워커(Walker) 순환을 가장 효과적으로 식별해 냄
  - 저자는 본 결과를 통해 PCMCI가 공통 원인 및 시차 변수 간의 간접적 연관성을 체계적으로 식별할 수 있음을 밝힘



자료: Runge et al.(2019b), 〈그림 1〉을 저자 수정.

#### 〈그림 2-9〉 PCMCI, 단순 상관관계와 이변량 그래인저 인과관계 기후 분야 적용 연구

- 그래인저 인과관계는 시계열이 가우시안 확률 분포를 따르고 정상성(stationarity)을 가정하기 때문에 분석이 제한적이라는 한계가 있음
- 최근 딥러닝 기법을 통해 복잡한 데이터에서 그래인저 인과성을 추출하는 방법이 생겨나고 있음

마. 탐욕적 동등 탐색(GES: Greedy Equivalence Search)

- 데이터에서 변수 간의 인과 그래프를 학습하기 위해 인과 그래프 구조를 정보 함유량에 따라 점수매겨 인과 그래프 구조를 도출하는 방식(그림 2-2 참조)<sup>12)</sup>
  - 일반적으로 BIC(Bayesian Information Criterion) 또는 AIC(Akaike Information Criterion)와 같은 점수 기준을 사용하여 그래프 적합도를 측정
  - 데이터 구조는 DAG(Directed Acyclic Graph)이라는 형태를 기반으로 하며 이는 인과 관계 모델링뿐 아니라 컴퓨터 내 데이터 흐름 분석 등에 널리 사용됨
  - 가장 높은 점수를 가진 구조를 선택(greedy한 방식)하면서 점진적으로 구조를 개선해 나가며 더 이상 점수 증가가 없을 경우 구조의 확장을 멈춤으로써 최적 구조를 도출
- 그래프 공간을 전부 탐색하지 않고 효율이 가장 높은 인과 구조를 도출하기 때문에 계산 비용이 낮고, 고차원 데이터에서도 효과적으로 작동한다는 장점이 존재
  - 연속 시계열 및 이산형 데이터 모두에서 사용 가능하여 적용 범위가 넓음
- 결과가 구조 그래프로 생산되기 때문에 정확한 방향성을 파악하기에는 제약이 있으며 노이즈가 많거나 데이터 부족시 정확도가 떨어진다는 한계가 존재
  - 확률 모델에 대한 적합도를 점수화하여 평가하는 방식이므로 확률 모델을 계산할 수 있는 데이터 확보가 어려운 경우 결과의 신뢰도가 낮아짐

〈표 2-1〉 인과분석 방식 기능 및 조건 구분, 자료: Runge et al. 2019b, 〈표2〉에서 일부 발췌.

	PC	FCI(MCI)	GES	LiNGAM
데이터에 대한 독립성이 요구되는가	네	네	일부 약 조건 요구	아니요
데이터 분포형에 대한 특정 가정이 필요한가	아니요	아니요	가우시안 모델 또는 다변량 분포를 가정	네
교란변수에 대한 적절한 처리가 가능한가	아니요	네	네	아니요

12) Chikering(2001)이 고안해 낸 인과추론 방식으로 변수의 구조를 결정하는 데 초점을 둠.

# III 가뭄 전파경로 분석 방법론 조사

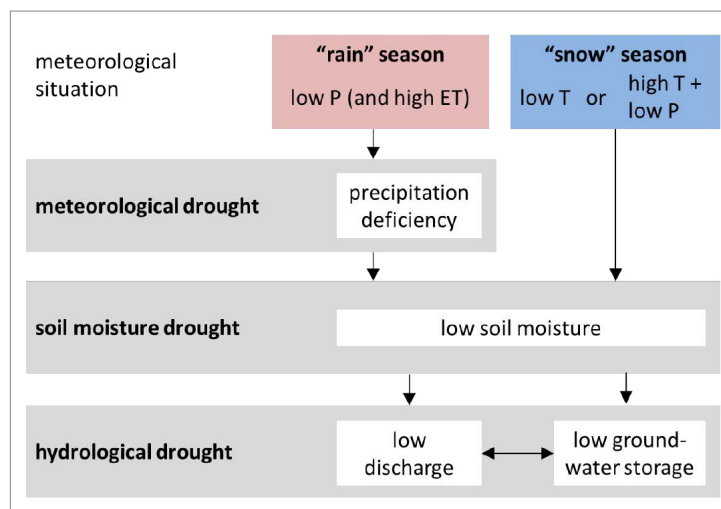
## 1. 가뭄 전파 과정 분석

### 가. 기상-수문학적 가뭄 전파 과정 연구

○ 가뭄의 전파 과정은 복잡한 물리 과정을 포함하고 있으며, 이 과정을 이해하려는 많은 선행 연구가 활발하게 수행됨(Van Loon., 2015; Shi et al., 2021; Zhou et al., 2021)

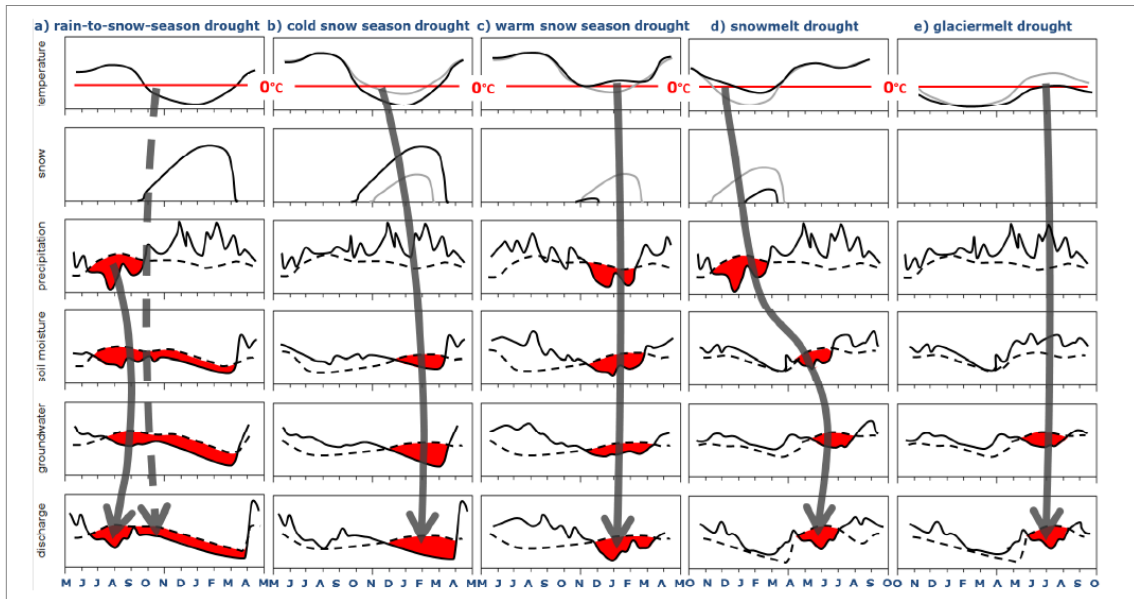
- Van Loon et al.(2013)은 기상학적 가뭄에서 수문학적 가뭄으로의 확산 과정이 강수량, 강설량, 기온 등 기상 요인에 의해 좌우되는 과정을 설명(그림 3-1 참조)

- 기상학적 가뭄은 강수량 감소 또는 강수량 감소와 증발산 증가의 합에 의해 좌우되며 장기간의 강수 부족이 수문학적 시스템의 물 부족을 야기
- 지하수 배수능력 저하, 토양 이상 상태, 식물 증발산량 감소는 토양 수분 저장량의 감소를 야기하며 이는 복사량 증가, 풍속 증가에도 영향을 받음
- 토양 수분량 감소는 지하수의 재충전을 막아 지하수위를 낮추며 이는 하천 유량을 줄이는 수문학적 가뭄에도 영향을 미침



자료: Van Loon et al.(2013).

〈그림 3-1〉 기상학적 가뭄에서 수문학적 가뭄으로의 전파에 대한 흐름도



자료: Van Loon et al.(2015), <그림 2>.13)

<그림 3-2> 다양한 기후 조건에서 기상학적 가뭄의 토양수분, 수문학적 가뭄 전파 과정

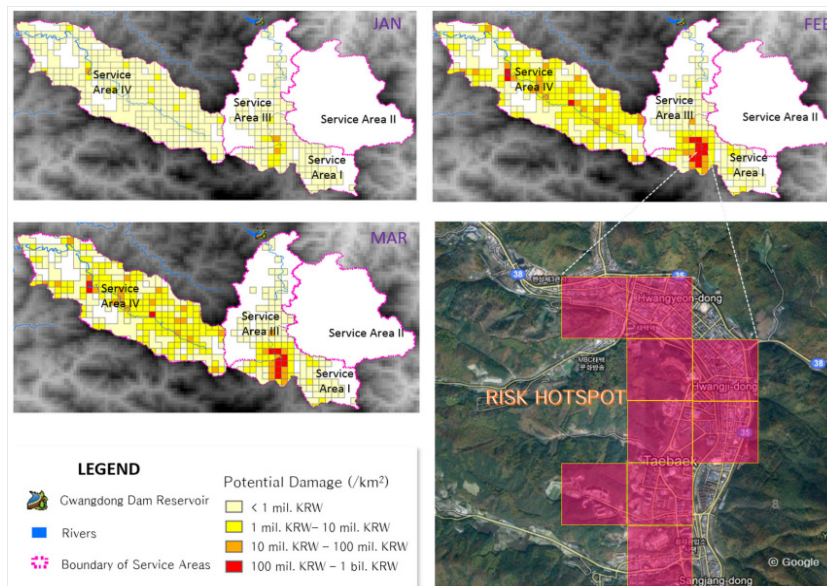
- 많은 연구에서 가뭄의 전파 과정은 원리상 인과관계를 보이며, 두 가뭄 이벤트 사이에 시간적 지연이 발생한다는 가정(예: 강수 부족은 유출량 감소를 초래하지만 유출량 감소는 강수 부족을 초래하기 어려움)을 기반으로 분석됨
  - 시계열 분석 및 시뮬레이션 모의를 통해 다양한 조건에서 가뭄의 전파 과정 경로 연구가 진행됨(그림 3-2 참조)
  - 변수를 Large-scale 기후모델을 통해 모의 후 선형 또는 비선형 상호관계 분석을 수행

#### 나. 기상-수문-사회경제 가뭄 전파 연구

- 사회경제적 가뭄은 동일한 강수 부족 상황에서도 지역의 기술, 경제 및 사회 특성에 따라 그 위험도 및 피해 규모가 달라질 수 있어 복잡한 시스템을 이해할 수 있는 다양한 방법론이 요구됨
  - 김지은 등(2019)은 가뭄에 대한 사회경제 인자(예: 생활 수급자수, 상수도 요금격차지표, 지하수 이용량, 1인당 가용 수자원량, 상수도 보급률 등)를 기반으로 지역에 맞는 평가 인자를 식별한 후 이에 기반하여 가뭄 취약성 평가를 수행

13) Van Loon et al.(2015)은 시계열 모의 과정을 통해 강수 및/또는 온도의 이상치(anomaly)가 수문 순환 과정을 통해 다양한 가뭄으로 전파되는 과정을 보여 줌.

- 분석 과정에서 주성분 분석<sup>14)</sup> 및 엔트로피 기법<sup>15)</sup>을 적용하여 평가 인자를 선정
- 이상은 등(2015)은 광동댐 저수지역을 대상으로 가뭄에 대한 사회경제적 위험도 평가 기법을 개발함으로써 전파경로에 대한 이해를 높이는 데 기여함
- 용수 공급을 위한 급수지역에서 용수의 가용량을 산정하기 위해 시계열 분석을 사용



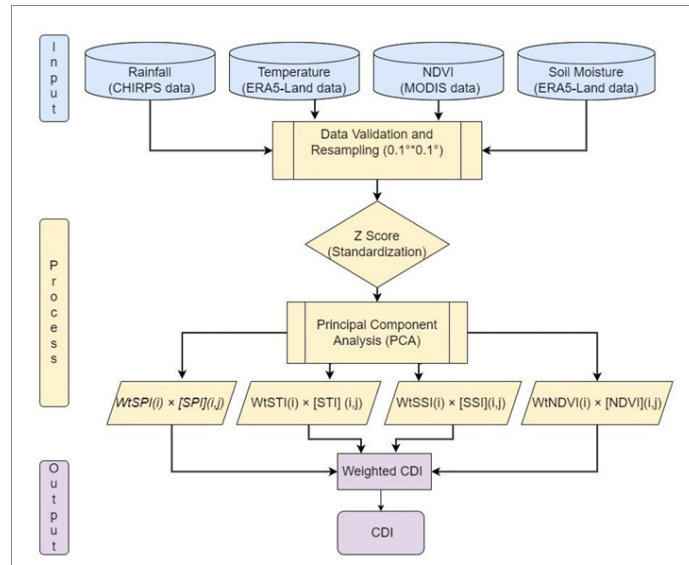
자료: 이상은 외(2015), <그림 6>.

<그림 3-3> 광동댐 저수지 주변의 사회, 경제적 영향을 고려한 가뭄 위험도 평가 지도

- Kulkarni et al. (2024)은 기상학적 가뭄이 사회경제적 가뭄에 미치는 영향을 정확히 평가하는 것이 중요한 과제를 강조
  - 한 가지 지표에 의존하는 기존 지수의 한계를 보완하기 위해 새로운 통합가뭄지수(CDI: Combined Drought Index)를 개발하고 기존 지표와 비교 검토함
  - 기존에는 SPI(기상학적 가뭄), SSI(수문학적 가뭄), NDVI(농업적 가뭄)가 연구되었으며 이 지수들을 결합하여 통합가뭄지수를 개발함
  - 통합가뭄지수는 PCA(Principal Component Analysis)를 이용하여 각 인자를 중요도 순으로 통합하여 개발됨
  - 사회경제적 가뭄은 GDIS(Geocoded Disaster Dataset) 기록을 이용하여 심각도를 정량화함

14) 주성분 분석(principal component analysis): 상관관계가 있는 평가인자들을 조합하여 인자들이 변수에 대한 정보를 최대한 포함할 수 있게 하는 다변량 통계 기법임(김지은 등, 2019).

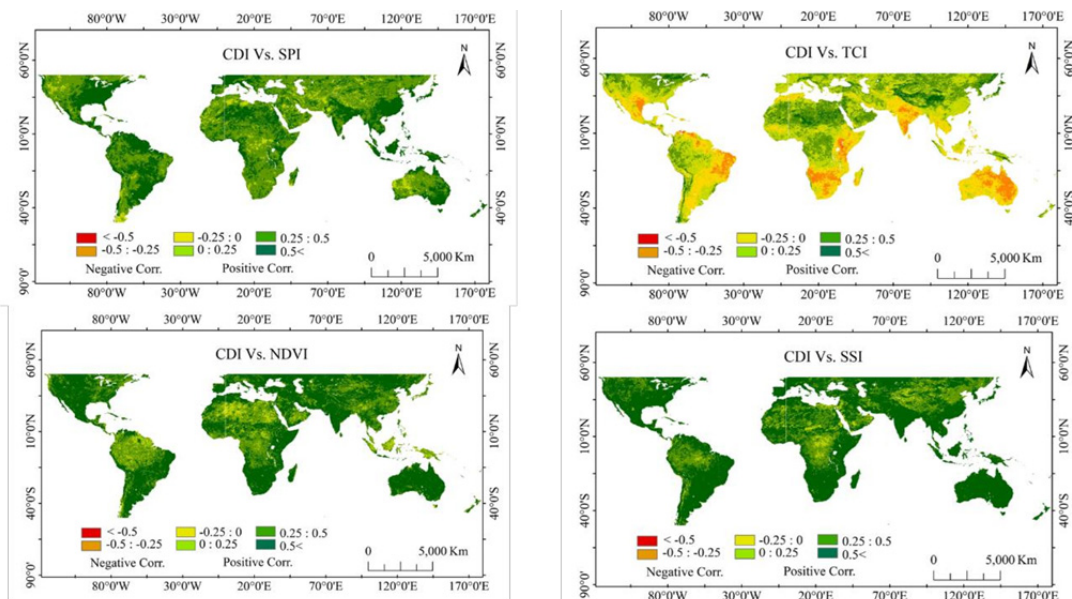
15) 엔트로피 기법: 엔트로피를 통해 불확실성 및 변수가 포함하고 있는 정보량을 계산함으로써 주어진 데이터의 속성을 이용하여 독립변수가 종속변수에 미치는 가중치를 추정하는 방법임(김지은 등, 2019).



자료: Kulkarni et al (2015), <그림 2>.

<그림 3-4> CDI 기반 가뭄 계산 과정에 대한 순서도

- 연구 결과는 통합가뭄지수(CDI)가 GDIS에 기록된 사회경제적 가뭄 사건의 95%를 성공적으로 감지한 결과를 보임으로써 사회경제적 가뭄 탐지에 높은 잠재력을 나타냄
- 동일한 CDI값, 즉 유사 가뭄 상황에서 개발도상국(예: 아프리카, 남아시아)은 선진국에 비해 큰 사회경제적 취약성을 보이는 것을 확인하고 인프라의 중요성을 조언함



자료: Kulkarni et al.(2021), <그림 9>.

<그림 3-5> Pearl River Basin의 1개월(a), 3개월(b), 6개월(c), 12개월(d) 지속기간의 가뭄 전파

## 2. 인과추론을 적용한 가뭄 전파 연구

- 가뭄 전파경로 분석에 인과추론 방식을 적용한 연구는 최근에 시작되었으며 많은 연구를 통해 방법론의 높은 잠재력을 확인
  - Shi et al.(2021)의 연구 결과는 연관분석에 비해 인과추론이 가뭄 전파 분석 결과에 높은 신뢰성을 가지고 있음을 보임
  - Wu et al.(2024)은 가뭄의 전파 과정을 이해하는 데 데이터 기반의 인과추론 방식이 기존 물리 기반 분석 방법의 한계를 극복할 수 있는 대안이라고 설명

### 가. 가뭄 유형별 지수 선정

- 인과추론 분석은 각 변수의 특성 파악을 위한 시계열이 필요함에 따라 가뭄 심각도의 정량적 평가를 위한 적절한 가뭄지수 선정 및 계산이 필수
  - 기간별 가뭄지수를 계산함으로써 가뭄 전파 과정에서 발생하는 시간 지연 특성을 파악할 수 있음
  - 선행 연구들은 다양한 가뭄지수 계산을 통하여 가뭄들의 인과관계를 확인(Wu et al., 2024; Shiau, 2021; Kulkarni et al. 2024)
- 가뭄지수는 기후 또는 수문 인자를 이용하여 가뭄의 심각도를 평가하는 수치로 특정 기간 동안의 가뭄 상태를 측정하기 위해 고안됨
  - 다양한 시간 척도에서 기후 모니터링을 통해 장기 가뭄 내에 단기 습윤 기간 또는 장기 습윤 기간 동안 내에 단기 가뭄 기간 등 비선형적인 추이의 식별이 가능
  - 가뭄 심각도, 위치, 발생 시기 및 지속 기간 등 지역의 가뭄 특성을 다각도로 정량화할 수 있다는 특성을 지님
    - 심각도는 가뭄지수가 특정 기간 내에 지역의 정상 상태에서 얼마나 벗어났는지를 나타냄
    - 발생 위치는 가뭄지수의 임계값 설정을 통하여 가뭄이 시작되고 종료되며 영향을 받은 지리적 범위를 파악
    - 발생 시기와 지속 기간은 가뭄의 시작과 종료 시기를 파악하는 데 사용되며 가뭄은 특히 시작 시기를 파악하기가 매우 어려움

〈표 3-1〉 전파경로 분석에 활용 가능한 가뭄지수 목록<sup>16)</sup>

가뭄지수	가뭄 형태	필요 기상인자	비고
Standardized Precipitation Index (SPI)	기상학적 가뭄	강수량	세계기상기구(WMO)에서 가뭄 산정에 기본이 되는 지수로 중요성 강조
Standardized Precipitation Evapotranspiration Index (SPEI)	기상학적 가뭄	강수량, 온도	계산 결과는 SPI와 유사하나, 온도 요소가 고려됨, 완전한 시계열 데이터가 필요
Palmer Drought Severity Index (PDSI)	토양수분 또는 수문학적 가뭄	강수량, 온도, 유효수분량 <sup>17)</sup>	계산 과정이 복잡하고 완전한 시계열 데이터가 필요
Self-Calibrated Palmer Drought Severity Index (sc-PDSI)	토양수분 또는 수문학적 가뭄	강수량, 온도, 유효수분량	계산 과정이 복잡하고 완전한 시계열 데이터가 필요
Standardized Water-level Index (SWI)	수문학적 가뭄	지하수위	SPI와 유사한 계산방식을 따르나 강수량 대신 지하수위를 이용
Streamflow Drought Index (SDI)	수문학적 가뭄	하천유량	SPI와 유사한 계산방식을 따르나 강수량 대신 하천유량을 이용
Standard Runoff Index (SRI)	수문학적 가뭄	유출량	SPI와 유사한 계산 방식을 따르나 수문학적 가뭄을 정의하는 데 더 적합
Multivariate Standardized Drought Index (MSDI)	복합적 가뭄	여러 데이터 동시 사용	가뭄의 복잡한 특성을 다차원적으로 평가. 사회경제적 가뭄 분석이 가능
Combined Drought Indicator (CDI)	복합적 가뭄	모델변수, 강수량 등	가뭄의 복잡한 특성을 다차원적으로 평가. 사회경제적 가뭄 분석이 가능

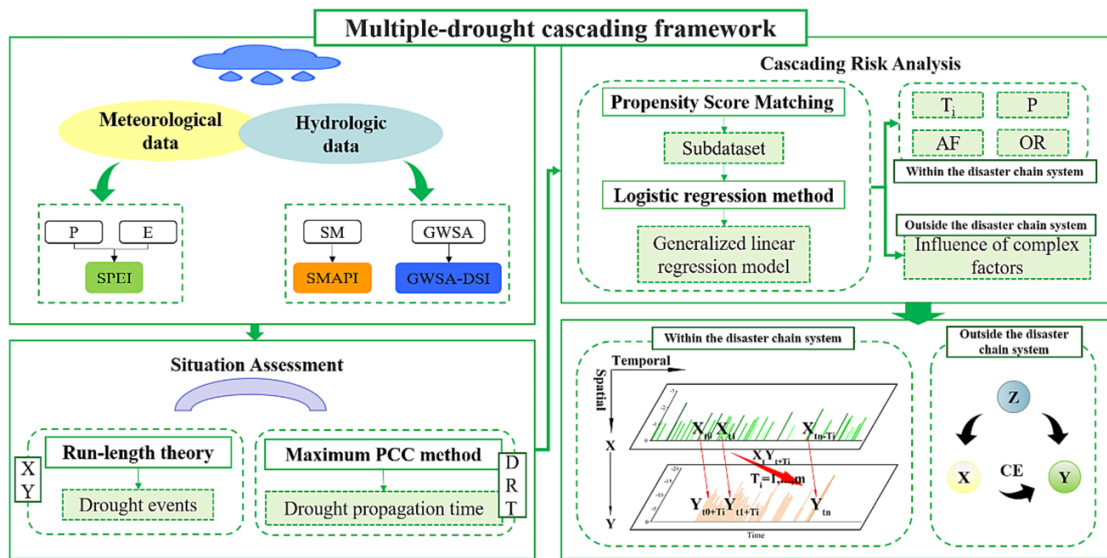
자료: WMO(2023), 보고서 〈표 1〉을 수정.

16) WMO(2023)에서 발표한 가뭄지수 목록 중 인과추론에 적절한 지수를 선택한 후 번역함.

17) 유효수분량(AWC: Available water content)은 토양이 식물에 이용 가능하도록 저장할 수 있는 물의 양을 나타내는 지표를 일컫음.

나. 가뭄지수 데이터에 인과추론 방법 적용

- Wu B. et al. (2024)은 인과추론을 적용하여 가뭄이 전파되는 과정을 정량화하고 시스템에 영향을 미치는 복잡한 요인들이 이 과정에 미치는 영향을 연구
  - 가뭄의 전파를 연쇄 재난(cascading diaster) 효과로 정의하고 정량화를 시도
  - 기존 연구들이 가뭄의 전파 과정에서 발생하는 외부 요인의 영향으로 인과관계를 명확하게 이해하기 어렵다는 한계를 극복하기 위해 인과추론 방법론을 적용
- (연구 목적) 기후변화와 식생 복원이 지표 수문순환을 하면서 가뭄 전파에 영향을 미치는 과정에 대한 이해
  - 중국의 황토고원 구릉지 및 협곡지역의 대표적인 강 유역인 옌허(Yanhe) 유역을 대상으로 함
  - 기상학적 가뭄(지수: SPEI), 토양수분 가뭄(지수: SMAPI), 지하수 가뭄(GWSA-DSI)에 대한 전파 과정에 초점을 맞춤

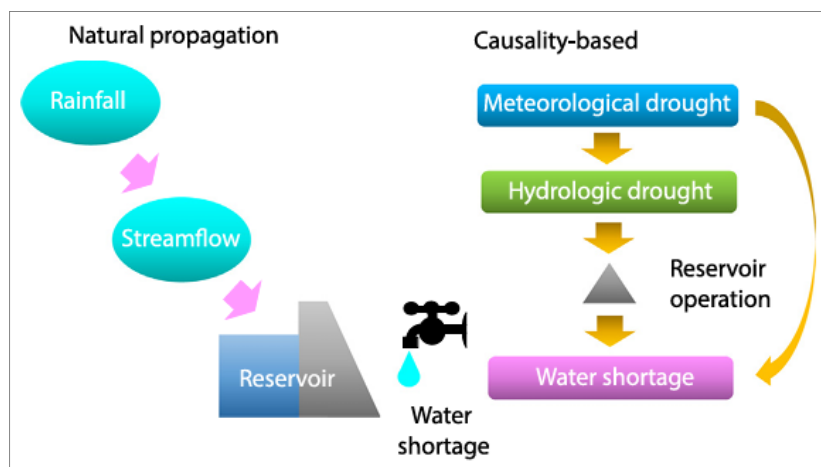


자료: Wu et al.(2024), <그림 2>.

<그림 3-6> 가뭄 전파 과정 분석 순서도

- (방법론) 인과 모델 구성을 위해 ‘다중 가뭄 연쇄 프레임워크’ 제안-성향 매칭 점수(propensity-matching scoring)와 로지스틱 회귀(Logistic regression) 기반
  - 시스템 외부의 복잡한 요인의 영향을 분리시킴으로써 가뭄 전파를 ‘연쇄 재난(disaster chain system)의 관점’에서 평가

- (연구 결과) 해당 지역에서 기상학적 가뭄과 토양수분 가뭄이 지하수 가뭄에 비해 강한 인과 관계가 존재함을 발견
  - 기상학적 가뭄과 토양수분 가뭄은 관측 기간 동안 완화된 반면 지하수 가뭄의 강도는 증가함
  - 기후변화와 식생 복원이 토양수분 가뭄과 지하수 가뭄에 잇따른 영향을 미친다는 점을 발견하고 식생 복원의 중요성을 강조

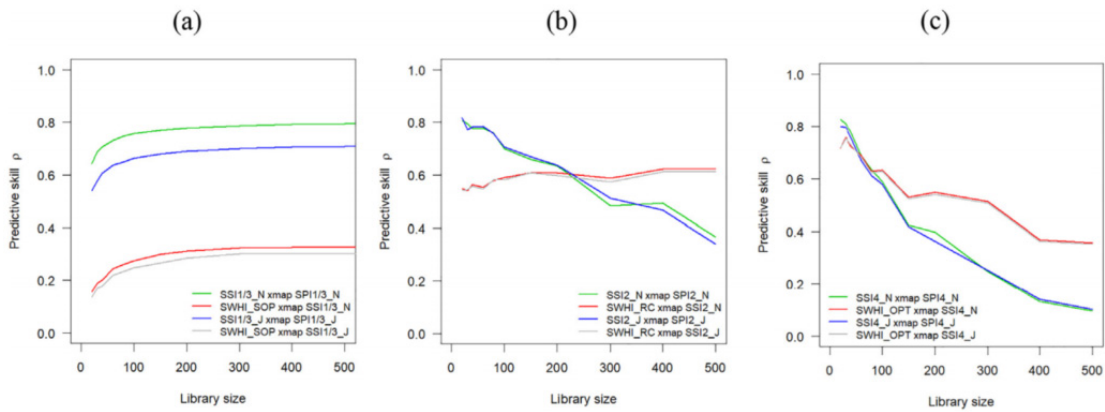


자료: Shiau(2024).

〈그림 3-7〉 저수지 운영 방식과 기상-수문학적 가뭄 전파 과정 관계에 대한 연구 흐름도

- Shiau(2023)은 대만의 난후아(Nanhua) 저수지 주변을 대상으로 기상-수문-사회적 가뭄의 전파 경로 및 전파 시간을 인과추론 방법을 적용하여 분석
  - 사회적 가뭄은 저수지 운영 방식에 따른 사용 가능한 물의 부족 정도로 정의
- (연구 목적) 기상학적 가뭄에서 수문학적 가뭄의 인과관계를 탐색하고 이 관계가 다양한 저수지 운영 방식에 따라 물 부족을 일으키는 과정을 분석
  - 대만 남부에 위치한 타이난(Tainan)과 카오슝(Kaoshung) 지방의 대표 물 공급원인 난후아(Nanhua) 저수지와 지아시안(Jiaxian) 위어 시스템을 대상으로 함
  - 기상학적 가뭄(지수: SPI), 수문학적 가뭄(지수: SSI), 사회적 가뭄(지수: SWHI) 간 인과 관계 추론에 주목함
- (방법론) 수렴 교차 매핑(CCM)을 적용하여 가뭄 간 인과관계를 추론
  - 다양한 시간 범위(1달- 5달)와 시간의 지연 효과를 함께 고려하여 인과관계를 관찰함

- (연구 결과) 기상-수문학적 10일 지속 기간 가뭄의 1/3개월 기간(그림 3-8a 참조)과 수문-사회적 2개월 지속 기간 가뭄 기간 사이에 유의미한 인과관계를 발견(그림 3-8b 참조)
  - 계산된 수렴 교차 매핑 변수의 크기가 증가할 때 예측력이 높은 값으로 수렴하면 두 인자들의 인과관계가 유의미한 것으로 판단함
- 〈그림 3-8c〉에 제시된 지속 기간이 4개월인 가뭄에서는 인과관계가 포착되지 않음
- Shi H 등(2021)은 중국 Pearl River Basin을 대상으로 기상-수문학적 가뭄의 전파 과정을 수렴 교차 매핑을 활용하여 분석함
  - 기상학적 가뭄(지수: SPI), 수문학적 가뭄(지수: SRI)를 대상으로 인과관계를 분석

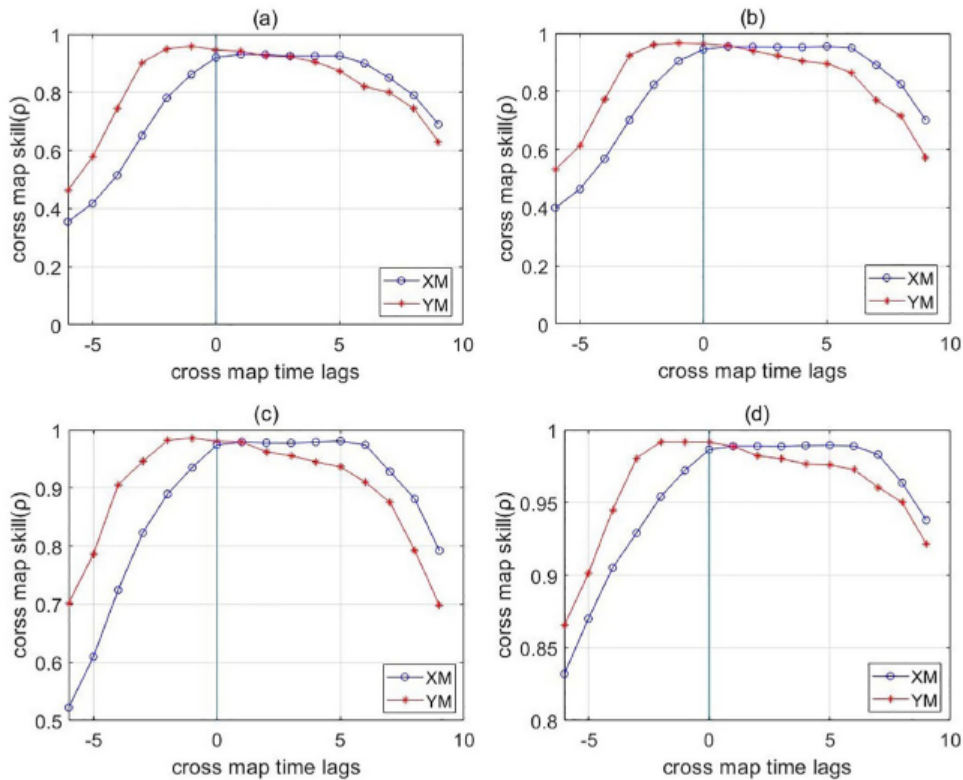


자료:Shiau(2024), 〈그림 6〉.

〈그림 3-8〉 기상-수문학적 가뭄(a), 수문 - 사회적 가뭄(b)의 CCM 인과추론 결과<sup>18)</sup>

- (연구 목적) 기상학적 가뭄에서 수문학적 가뭄의 인과관계를 탐색하고 이 관계가 다양한 저수지 운영 방식에 따라 물 부족을 일으키는 과정을 분석
  - 가뭄 지속 시간별, 전파 시간별 인과관계 추론에 주목하였으며 지속 기간 1, 3, 6, 12개월 가뭄을 대상으로 연구
- (방법론) ERA-5 월별 강수량, 유출량 데이터에 수렴 교차 매핑(CCM) 방식을 적용하여 인과관계를 분석
  - 수렴 교차 매핑 방식의 변수 간 시간적 지연 및 방향성 감지 능력이 가뭄 전파 기간을 분석하는 데 효과적이라는 점에 착안함
- (연구 결과) 수렴 교차 매핑 분석을 통하여 연구 지역의 기상학적 가뭄에서 수문학적 가뭄으로의 전파 기간이 약 5~6개월 정도 소요됨을 확인함

18) 그림에서 X "xmap" Y는 Y가 X의 원인이 됨을 의미함.



자료: Shi et al.(2021), <그림 3>.

<그림 3-9> Pearl River Basin의 1개월(a), 3개월(b), 6개월(c),12개월(d) 지속기간의 기상-수문학적 가뭄 전파 과정의 인과추론 결과<sup>19)</sup>

- 연구 결과 모든 지속 기간의 기상학적 가뭄에서 수문학적 가뭄으로의 전파(XM) 경로가 전파 소요 시간(1~5개월) 사이에서 상관관계가 높게 나타남을 확인(그림 3-9 참조)
- 수문학적 가뭄에서 기상학적 가뭄으로의 전파(YM)는 전파 소요시간이 반대 (0~-2개월) 일 때 상관관계가 높게 나타나 기상-수문학적 가뭄 전파 방향에 신뢰성이 강화됨
- 수렴 교차 매핑 방식은 거짓 양방향 인과관계 도출이라는 기존 상관분석이 가진 한계를 극복할 수 있는 잠재력이 있는 것을 확인
- 이 방식은 인과관계 검출 성능 뿐 아니라 전파 시간에 대한 정확한 측정을 통해 분석 대상의 물리적 메커니즘을 이해하는 데 기여할 것으로 기대
- 다만 장기간의 관측 데이터가 필요하고 가뭄이 가진 계절성의 영향을 완전히 배제하지 못한다는 한계를 극복할 방안이 필요함

19) XM에 cross map time lags가 양수인 경우는 기상학적 가뭄이 먼저 발생하고 이것이 그 후에 발생하는 수문학적 가뭄의 원인이 됨을 의미함.

---

## IV 향후 연구 방안

---

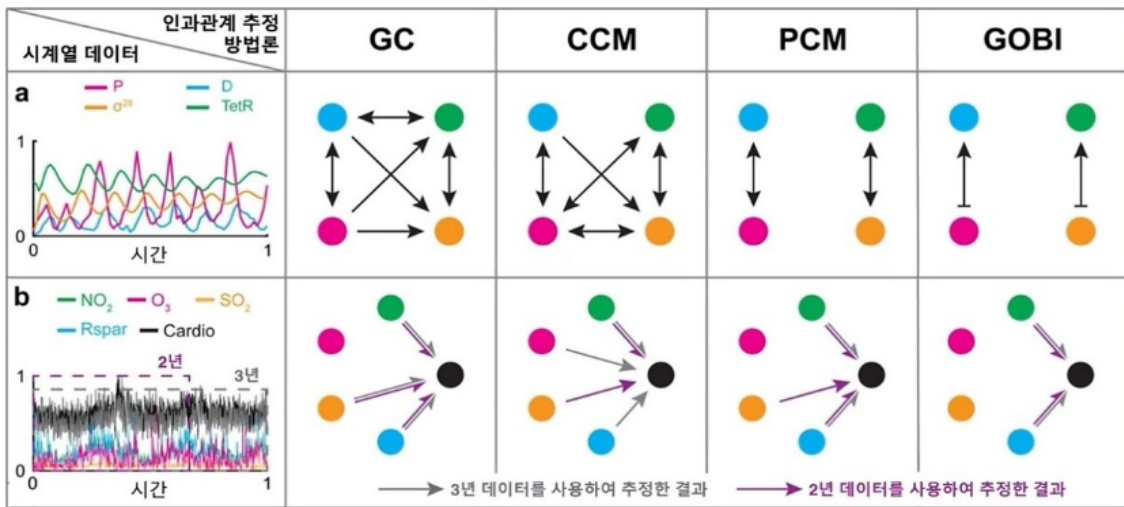
### 1. 인과추론 모델 개발 시 고려 사항

- 각 인과추론 방법론의 가정 사항 및 한계점을 정확히 이해하고 모델을 개발하는 것이 분석 결과의 왜곡된 해석과 잘못된 결론을 막을 수 있음<sup>20)</sup>
  - 시계열 기반의 인과관계 발견 기법은 두 변수의 시계열이 비슷한 주기로 변화하면 인과관계가 있다고 잘못 판단할 가능성이 있음
  - 허위변수(spurious variable) 존재 시 직접적인 인과관계와 간접적 인과관계 구분이 어려우므로 변수들 간의 조건부 독립성 확인을 통하여 잘못된 인과관계를 탐지
- 수리 모델(시계열을 이용한 통계 기반 기법)은 데이터가 짧거나 데이터 내의 노이즈가 많을 경우 결과의 신뢰성이 낮아질 수 있으므로 충분한 데이터 확보가 전제되어야 함
  - 수리 모델은 모델의 정확성에 따라 인과관계의 올바른 추정이 좌우되나, 복잡한 모델의 경우 과도한 계산 자원(computation resource)이 소요된다는 한계 존재
- 데이터 사이즈, 노이즈 정도, 비선형성, 비정상성 등 데이터의 특성을 파악하여 적절한 인과추론 모델을 선택해야 함
  - 데이터의 가정사항 위반에 대한 정확한 이해를 위하여 민감도 분석 등 필요에 따라 추가 분석 필요
- 새로운 방식으로 특정 모델에 대한 가정 없이 시계열로부터 인과관계를 추정하는 방법론을 추가로 고려할 필요가 있음<sup>21)</sup>
  - 최근 여러 통계 검정법을 기반으로 복잡한 계산 없이 미분방정식을 이용하여 인과관계를 추정하는 방법이 개발됨(general ODE-based inference, Park, Ha, and Kim, 2023)
  - 현재 다른 분야 응용 사례가 없는 신규 방법이라 본 보고서에서 자세히 다루지 않았으나 향후 연구에 적용 가능성이 있음

---

20) 과학지식백과(2023)에서는 인과관계 추정에서 발생하는 잠재적 문제를 설명함.

21) Park, Ha, and Kim(2023)은 일반 미분방정식을 이용한 인과발견 방법론을 개발함.



자료: Park, Ha, and Kim (2023), p.14, <그림 5>.

<그림 4-1> GC, CCM, PCM과 GOBI의 성능 비교

## 2. 향후 모델 개발 방안

### 가. 가뭄지수 선정의 중요성

- 적절한 가뭄지수의 선정은 가뭄의 특성을 파악하는 중요한 결정요소이고 연구 대상에 맞는 지수 결정이 전파경로 분석 연구의 출발점임
  - 데이터 유무, 연구 지역 특성에 따라 가뭄 종류별로 가뭄 지수를 선별해야 함
    - 예를 들어 기상학적 가뭄 계산을 위해 강수량, 증발산을 이용한 SPEI를 적용할 수 있으나 증발산 계산에 필요한 데이터 부족 시 강수량만 이용하는 SPI를 사용할 수 있음
- 사회경제적 가뭄은 타 가뭄에 비해 가뭄 지수 개발이 부족한 상태이므로 추후 면밀한 조사 및 장단점 파악을 통해 적절한 지수를 선택하는 것이 필요함
  - 본 연구에서는 기후 및 수문 변수에 사회경제 데이터를 통합한 MSDI를 적절한 대안으로 제시(표1)
    - MSDI는 GDP, 농업 생산량, 물 소비량, 인구 밀도 등을 기후 및 수문 변수에 결합할 수 있으므로 사회경제적 가뭄 분석에 적합한 지표로 이용할 수 있음
  - CDI 역시 기후 및 수문 변수에 사회경제 데이터를 통합하여 사회경제적 가뭄 지표로 이용할 수 있는 잠재력 보유
    - 앞서 언급한 선행 연구(Kulkarni et al., 2024)에서와 같이 가뭄 재난 관련 기록을 결합한 가뭄 지수 개발이 가능

- 선형 연구들과 같이 사회경제적 가뭄은 다면적 특성 파악이 중요하므로 복합가뭄지표를 이용하는 것이 적절함

#### 나. 인과관계 분석론 적용

- 연구 지역에 맞는 지수 선정 후 데이터 성격에 맞는 적절한 인과관계 분석 방법을 적용하여 가뭄 전파경로를 분석
  - PCMCI 이용 시 시계열의 조건부 독립성 검정 수행 후 인과 그래프 구성(그림 2-2 참조)을 통해 가뭄 간 전파경로를 파악
  - 탐욕적 동등성 탐색(GES) 이용 시 각 변수 쌍(pair)에 대한 상호 정보량(Mutual information)을 계산한 후 정보량 순으로 인과 그래프를 확장하고 정보량이 가장 많은 인과 그래프를 최종 전파 형태로 결정함으로써 가뭄 전파경로를 이해
  - GOBI 방법 이용 시 가뭄 시스템을 표현하는 미분방정식 모델 구축 후 실제 관측 데이터를 적용하여 매개변수 추정 후 모델 시뮬레이션을 통해 전파 과정을 분석
- 인과추론 분석 후 모델 검증을 비롯한 후처리 과정을 통해 잘못된 인과관계 판단을 줄이고 결과 신뢰성을 확보
  - 모델의 입력변수가 결과에 미치는 영향에 대한 민감도를 분석해 중요한 변수들을 식별하고 과대 추정되는 변수 간 관계에 대한 신뢰성을 평가함
  - 모델 결과의 불확실성을 정량화하고 신뢰구간을 설정하여 모델이 생산하는 결과를 다양한 조건별로 신뢰구간을 확보

#### 다. 정책 제안의 기초

- 가뭄 인과추론 분석은 가뭄 간의 관계에 대한 이해를 높임으로써 정책 제안의 중요한 기초자료로 활용할 수 있음
  - 인과추론 방식을 기반으로 기상-수문-사회경제적 가뭄 전파 과정에서 취약한 경로를 파악함으로써 이를 단계 맞춤형 대응 정책 수립에 기초 근거로 활용할 수 있음
  - 가뭄 전파경로의 핵심 요인들을 식별하여 제한된 자원을 효과적으로 배분하기 위한 정책의 우선순위를 결정하는 데 기초 자료로 활용할 수 있음
  - 가뭄의 복합적인 영향에 대한 이해 증진을 통해 농업, 수자원, 환경, 경제 등 여러 분야를 아우르는 통합적인 가뭄 대응 체계를 구축할 수 있음

## 참고문헌

### [국내문헌]

- 관계부처합동(2023), 「제3차 국가 기후위기 적응 강화대책」.
- 김지은 외(2019), “구성분 분석 및 엔트로피 기법을 적용한 사회·경제적 가뭄 취약성 평가”, 「한국수자원학회지」, 52(6), 한국수자원학회, pp.441-449.
- 김지은 외(2021), “구조방정식모형과 베이지안 네트워크를 활용한 가뭄 영향인자의 민감도 분석”, 「대한토목학회논문집」, 42(1), 대한토목학회, pp.11-21.
- 류재나 외(2015), 「인과지도(Causal Loop)를 활용, 미래 물수급관리 정책 지원을 위한 기초연구」, 한국환경정책·평가연구원.

### [국외문헌]

- Bhaskar, A. et al.(2017), “Quantitative Assessment of Drivers of Recent Global Temperature Variability: An Information Theoretic Approach”, *Climate Dynamics*, 49(1), pp.3877-3886.
- Clark, A. T. et al.(2015), “Spatial Convergent Cross mapping to Detect Causal Relationships from Short Time Series”, *Ecology*, 96(5) pp.1174-1181.
- Chickering, D. M.(2022), “Optimal Structure Identification With Greedy Search”, *Journal of Machine Learning Research*, 3, pp.507-554.
- Castri, L., et al. (2022). Causal Discovery of Dynamic Models for Predicting Human Spatial Interactions, *10.48550/arXiv.2210.16535*.
- Corander, J., Hanage, W. P., and J. Pensar (2022). “Causal Discovery for the Microbiome”, *The Lancet Microbe*, 3(11), pp.881-887.
- Delussu, F. et al.(2023), “The Limits of Human Mobility Traces to Predict the Spread of COVID-19: A Transfer Entropy Approach”, *PNAS Nexus*, 2(10), pgad302.
- Docquier, D. et al.(2024), “A Comparison of Two Causal Methods in the Context of Climate Analyses”, *Nonlinear Processes in Geophysics*, 31(1), pp. 115-136.
- Glymour, C., Zhang, K., & Spirtes, P. (2019). “Review of Causal Discovery Methods Based

- on Graphical Models”, *Frontiers in Genetics*, 10(524).
- Takens F., D. A. Rand and L.-S. Young (ed) (1981), Detecting Strange Attractors in Turbulence”, *Dynamical Systems and Turbulence, Lecture Notes in Mathematics*, Vol.898, Springer-Verlag, p.366-381.
- Ludescher, J. et al.(2021), “Network-Based Forecasting of Climate Phenomena”, *Proceedings of the National Academy of Sciences*, 118(47), e1922872118.
- Haig, B. D.(2003), “What is a Spurious Correlation?”, *Understanding Statistics*, 2(2), pp.125-132.
- Hlinka, J. et al.(2013), “Reliability of Inference of Directed Climate Networks Using Conditional Mutual Information”, *Entropy*, 15(6), pp.2023-2045.
- Huffaker, R. and Andrew F. (2019), “Reconstructing Systematic Persistent Impacts of Promotional Marketing with Empirical Nonlinear Dynamics”, *PloS One*, 14(9) e0221167.
- Park, S. H., S. Ha and J. K. Kim(2023), “A General Model-Based Causal Inference Method Overcomes the Curse of Synchrony and Indirect Effect”, *Nat Commun*, 14, 4287.
- Kadir, M. et al.(2020), “Exploring Causes of Streamflow Alteration in the Medjerda River, Algeria”, *Journal of Hydrology: Regional Studies*, 32, 100750.
- Krich, C. et al.(2020), “Estimating Causal Networks in Biosphere: Atmosphere Interaction with the PCMCI Approach”, *Biogeosciences*, 17(4) pp. 1033-1061.
- Krich, C. et al.(2022), “Decoupling between Ecosystem Photosynthesis and Transpiration: A Last Resort against Overheating”, *Environmental Research Letters*, 17(4), 044013.
- Kulkarni, S. et al.(2024), “Global Assessment of Socio-Economic Impacts of Subnational Droughts: A Comparative Analysis of Combined Versus Single Drought Indicators”, *Hydrol. Earth Syst. Sci. Discuss*, [preprint], DOI: <https://doi.org/10.5194/hess-2024-245>, in review.
- Kushwaha, N. and E. D. Lee.(2023), “Discovering the Mesoscale for Chains of Conflict”, *PNAS Nexus*, 2(7), pgad228.
- Ludescher, J. et al.(2021), “Network-Based Forecasting of Climate Phenomena”, *Proceedings of the National Academy of Sciences*, 118(47), e1922872118.

- Medina, M. et al.(2019), “Nonlinear Dynamics in Treatment Wetlands: Identifying Systematic Drivers of Nonequilibrium Outlet Concentrations in Everglades STAs”, *Water Resources Research*, 55(12), pp.11101-11120.
- Runge, J. et al.(2019a), “Detecting and Quantifying Causal Associations in Large Nonlinear Time Series Datasets”, *Science Advances*, 5(11), eaau4996.
- Runge, J. et al.(2019b), “Inferring Causation from Time Series in Earth System Sciences”, *Nature Communications*, 10(1), 2553.
- Shannon, C. E.(1948), “A Mathematical Theory of Communication”, *The Bell System Technical Journal*, 27(3), pp.379-423.
- Shen, X., et al. (2020). “Challenges and Opportunities with Causal Discovery Algorithms: Application to Alzheimer’s Pathophysiology”, *Scientific Reports*, 10(1), 2975.
- Shi, H. et al.(2022), “A New Perspective on Drought Propagation: Causality”, *Geophysical Research Letters*, 49, e2021GL096758.
- Shiau, J.-T.(2023), “Causality-Based Drought Propagation Analyses among Meteorological Drought, Hydrologic Drought, and Water Shortage”, *Science of the Total Environment*, 888, 164216.
- Shin, C., P. A. Dirmeyer, and B. Huang(2023), “A Joint Approach Combining Correlation and Mutual Information to Study Land and Ocean Drivers of U.S. Droughts: Methodology”, *J. Climate*, 36, pp.2795-2814.
- Silfwerbrand, L. et al.(2024), “Directed Causal Effect with PCMCI in Hyperscanning EEG Time Series”, *Frontiers in Neuroscience*, 18, 1305918.
- Sugihara, George et al.(2012), “Detecting Causality in Complex Ecosystems”, *Science*, 338(6106), pp.496-500.
- Sung, K., Bohrer, G., and Stagge, J. H. (2024). “Centennial-scale intensification of wet and dry extremes in North America”, *Geophysical Research Letters*, 51, e2023GL107400.
- Tárraga, J. M. et al.(2024), “Causal Discovery Reveals Complex Patterns of Drought-Induced Displacement”, *Science*, 27.9.
- Van Loon A. F.(2013), On the Propagation of Drought. How Climate and Catchment

- Characteristics Influence Hydrological Drought Development and Recovery, *PhD thesis, Wageningen University, Wageningen, NL*
- Van Loon et al.(2015), “Hydrological Drought Types in Cold Climates: Quantitative Analysis of Causing Factors and Qualitative Survey of Impacts”, *Hydrol. Earth Syst. Sci.*, 19, pp.1993–2016
- Vicente, R. et al.(2011), “Transfer Entropy: A Model-Free Measure of Effective Connectivity for the Neurosciences”, *Journal of Computational Neuroscience*, 30.(1), pp.45-67.
- Wang, Y. et al.(2018), “Detecting the Causal Effect of Soil Moisture on Precipitation Using Convergent cross Mapping”, *Scientific Reports*, 8(1), 12171
- World Meteorological Organization(WMO) and Global Water Partnership(GWP)(2016), “Handbook of Drought Indicators and Indices”, M. Svoboda and B.A. Fuchs, Eds., *Integrated Drought Management Programme(IDMP)*, Integrated Drought Management Tools and Guidelines Series 2.
- Wu B, et al.(2024), “A Multiple-Drought Cascading Framework Based on Causal Inference”, *Journal of Hydrology*, Vol.630, 130657.
- Yazdani, A., and Boerwinkle, E.(2014). “Causal Inference in the Age of Decision Medicine” *Journal of Data Mining in Genomics & Proteomics*, 6(1), 163
- Zanga, A., Ozkirimli, E., and Stella, F.(2022), “A Survey on Causal Discovery: Theory and Practice”, *International Journal of Approximate Reasoning*, 151, pp.101-129.
- Zigler C.M., et al. (2016), “HEI Health Review Committee. Causal Inference Methods for Estimating Long-Term Health Effects of Air Quality Regulations”, *Res Rep Health Eff Inst.*, 187, pp.5-49.
- Zhang, H. et al.(2021), “Understanding the Soil Temperature Variability at Different Depths: Effects of Surface Air Temperature, Snow Cover, and the Soil Memory”, *Advances in Atmospheric Sciences*, 38, pp.493-503.
- Zhao, Z., Na Z., and Na Y.(2021), “Association, Correlation, and Causation among Transport Variables of PM<sub>2.5</sub>”, *Frontiers in Physics*, 9, 684104.

[온라인 자료]

과학지식백과(2023), “인과관계 추정 정확도를 높인 새로운 방법론”, [https://www.ibs.re.kr/cop/bbs/BBSMSTR\\_000000000901/selectBoardArticle.do?nttId=24225&pageIndex=1&searchCnd=&searchWrd=](https://www.ibs.re.kr/cop/bbs/BBSMSTR_000000000901/selectBoardArticle.do?nttId=24225&pageIndex=1&searchCnd=&searchWrd=), 검색일: 2024.11.25.

국가가뭄정보포털, “[쉽게보는가뭄] 가뭄이란/특징/영향”, <https://www.drought.go.kr/menu/m30/m31.do>, 검색일: 2024.11.24

환경부 보도자료(2024.4.29), “기상청, 정부합동 2023년 이상기후보고서 발간”, <https://www2.korea.kr/briefing/pressReleaseView.do?newsId=156627835&pWise=sub&pWiseSub=C4>, 검색일: 2024.11.1.

City of Las Vegas(2022.8.16), “Lake Mead Water Shortage”, <https://www.lasvegasnevada.gov/News/Blog/Detail/lake-mead-water-shortage>, 검색일: 2024.11.25.

Meghanath, G.(2024.11.5), “Causal Analysis Overview: Causal Inference versus Experimentation versus Causal Discovery”, *Data Science at Microsoft*, <https://medium.com/data-science-at-microsoft/causal-analysis-overview-causal-inference-versus-experimentation-versus-causal-discovery-d7c4ca99e3e4>, 검색일: 2024.11.25.

※ 본 책자는 환경표지 인증을 받은 용지로 인쇄되었습니다.



## 가뭄의 기상-수문-사회경제 영향경로 분석을 위한 인과추론(Causality) 방법론 검토

**KEI** 한국환경연구원  
Korea Environment Institute

(30147) 세종특별자치시 시청대로 370 세종국책연구단지 B동(과학·인프라동)  
전화 044-415-7777 팩스 044-415-7799 www.kei.re.kr



ISBN 979-11-5980-961-3