

2 Methodische Vorgehensweise

Anke Hupe, Jörn Sanders, Lucie Chmelikova, Andreas Gattinger, Almut Haub, Johanna Hoppe, Rüdiger Jung, Karin Levin, Solveig March, Hanna Treu, Philipp Weckenbrock

Im Rahmen der vorliegenden Arbeit wurde für jeden Leistungsbereich eine systematische Literaturanalyse durchgeführt. Ziel war es zum einen, alle relevanten Studien der letzten drei Jahrzehnte zu identifizieren, in denen die ökologische und konventionelle Landwirtschaft hinsichtlich der ausgewählten Leistungsbereiche vergleichend analysiert wurden. Zum anderen galt es zu überprüfen, ob sich die theoretisch aus den Produktionsvorschriften ableitbaren Leistungen des Ökolandbaus anhand publizierter Forschungsergebnisse belegen lassen. Neben der Suche nach naturwissenschaftlichen Arbeiten wurde ergänzend eine Literaturrecherche nach Studien durchgeführt, die sich mit den durch die Landwirtschaft verursachten gesellschaftlichen Kosten sowie dem Wert der durch die Landwirtschaft erbrachten gesellschaftlichen Leistungen beschäftigt haben. Nachfolgend wird das methodische Vorgehen dieser Arbeit beschrieben. Ergänzende Informationen befinden sich hierzu auch in einem separaten Materialband.⁶

2.1 Identifikation und Auswahl der Vergleichsstudien und Indikatoren

Identifikation der Studien

Die Identifikation relevanter Vergleichsstudien basierte auf einer abgestimmten, systematischen Literaturrecherche, die themenspezifisch für die einzelnen Leistungsbereiche anhand der festgelegten Suchwörter durchgeführt wurde. Hierfür wurden die wissenschaftlichen Datenbanken *Web of Science*, *Scopus* und *Organic Eprints* verwendet (für den Leistungsbereich Wasserschutz wurden zusätzlich noch die Datenbanken *pubmed* und *KARLA* verwendet). Die Suche in Datenbanken wurde ergänzt durch eine web-basierte Suche (u.a. mit Hilfe des Schneeballverfahrens), um relevante, aber nicht dem klassischen Peer-Review unterliegende Literatur zu erfassen. Im Fokus dieser Suche standen insbesondere Endberichte relevanter wissenschaftlicher Projekte, Tagungsbeiträge und Dissertationen. Für die Leistungsbereiche Klimaschutz und Ressourceneffizienz wurden zudem bestehende Datensätze aus anderen Forschungsarbeiten genutzt. Die thematische Abgrenzung der einzelnen Leistungsbereiche, beispielsweise welche Produktionsbereiche (Ackerbau, Tierhaltung, Wein, Obst, Gewächshaus, etc.) berücksichtigt wurden, erfolgte für jeden Bereich individuell (siehe Kapitel 2.3).

⁶ Der Materialband enthält Informationen zu (a) den verwendeten Schlagwörtern und Teilindikatoren, (b) den Vergleichspaaren mit bibliographischen Angaben sowie (c) relevanten Produktionsvorschriften der EU-Öko-Verordnung und den Richtlinien ausgewählter deutscher Anbauverbände. Der Materialband ist abrufbar unter: <https://www.thuenen.de/de/bw/projekte/leistungen-des-oekolandbaus-fuer-umwelt-und-gesellschaft/>

Auswahl der Studien

Von den im Rahmen der Literaturrecherche identifizierten Studien wurden die Arbeiten ausgewählt, die mindestens folgende Kriterien erfüllten:

- Der geografischer Raum der Studien umfasste die temperierten Klimazonen.
- Die Vergleichsstudien beinhalteten zumindest ein Vergleichspaar (VGP) mit einer ökologischen und konventionellen Variante.
- Die Umstellung der ökologischen Flächen lag mindestens zwei Jahre zurück.
- Der Erscheinungszeitraum der Studien lag zwischen Januar 1990 bis März 2017.
- Die Publikation erfolgte auf Deutsch oder Englisch.

Darüber hinaus wurden für jeden Leistungsbereich ergänzende Auswahlkriterien festgelegt, die im Kapitel 2.3 beschrieben werden. Bei mehrfach publizierten Studienergebnissen wurden jeweils die neuesten und vollständigsten Daten einbezogen, um Doppelbewertungen zu vermeiden. Insgesamt wurden 528 Vergleichsstudien für die Analyse ausgewählt (Tabelle 2.1). Von jeder Studie wurde jeweils ein PDF-Dokument der Veröffentlichung archiviert.

Spezifikation der Indikatoren zur Bewertung der gesellschaftlichen Leistung

Aus den Studien wurden anschließend relevante Ergebnisse extrahiert und in eine Datenmatrix überführt. Für die Auswertung graphischer Abbildungen wurde im Bedarfsfall ein WebPlotDigitizer verwendet (Rohatgi, 2017). Die zusammengestellte Datenbasis umfasst insgesamt 2.816 Vergleichspaare und ermöglicht die Bewertung der gesellschaftlichen Leistung des ökologischen Landbaus anhand von 36 Indikatoren (Tabelle 2.1 - 2.3). Da die Analyse der gesellschaftlichen Leistungen des ökologischen Landbaus primär auf einer Literaturlauswertung basiert, ergab sich die Auswahl der Leistungsindikatoren durch die in den Studien verwendeten Beurteilungskriterien. In den Bereichen Wasserschutz, Biodiversität und Tierwohl setzen sich die Indikatoren aus verschiedenen Teilindikatoren zusammen, die in den Kapiteln 3, 5 und 9 näher beschrieben werden. Die Tierwohlindikatoren beziehen sich auf sieben Tierarten; die für den Leistungsbereich Biodiversität verwendeten Indikatoren auf fünf Artengruppen sowie drei Teilgruppen.

Tabelle 2.1 Anzahl der im Rahmen der Literaturrecherche identifizierten und ausgewählten Studien, Indikatoren und Vergleichspaare

Leistungsbereiche	Anzahl der Studien		Anzahl der verwendeten Indikatoren	Anzahl der verwendeten Vergleichspaare
	identifiziert	verwendet		
Wasserschutz ^a	6.711	98	4	355
Bodenfruchtbarkeit	1.211	56	5	307
Biodiversität ^b	801	75	2	312
Klimaschutz ^c	1.617	119	8	311
Klimaanpassung	372	43	7	278
Ressourceneffizienz	1.986	70	7	780
Tierwohl ^d	474	67	3	473

a) Die Anzahl der Indikatoren bezieht sich auf die Indikatorengruppen (bzw. Stoffgruppen) Stickstoff, Pflanzenschutzmittel, Phosphor und Tierarzneimittel.

b) Die Anzahl der Indikatoren bezieht sich auf Artenzahl und Abundanz, die jeweils in Mittel- und Gesamtwerte (Teilindikatoren) unterschieden wurden. Berücksichtigt wurden fünf Artengruppen: Ackerflora, Samenbank, Saumvegetation, Vögel und blütenbesuchende Insekten (Wildbienen inkl. Hummeln, Honigbienen, Tagfalter, Schwebfliegen und Florfliegen).

c) Zusätzlich wurde ein Datensatz zur Bodenkohlenstoffspeicherung aus Gattinger et al. (2012) genutzt.

d) Die Anzahl der Indikatoren bezieht sich auf Indikatorengruppen (bzw. Tierwohldimensionen) (a) Tiergesundheit (Eutergesundheit, Stoffwechselkrankheiten, Gliedmaßen- und Klauengesundheit sowie Integumentschäden, Fruchtbarkeit und Reproduktion, Mortalitäts- und Abgangsraten sowie sonstige Gesundheitsaspekte), (b) Tierverhalten und (c) Emotionales Befinden. Berücksichtigt wurden sieben Tierarten.

Quelle: Eigene Darstellung

2.2 Auswertungsverfahren und Ergebnisdarstellung

2.2.1 Deskriptiv-statistische Auswertung

In einem ersten Auswertungsschritt wurden die Ergebnisse der Paarvergleiche zunächst deskriptiv statistisch analysiert (Min-Wert, Max-Wert, Mittelwert, Median) und mithilfe von Boxplot-Diagrammen graphisch dargestellt, soweit für einen Indikator mindestens neun Vergleichspaare vorlagen. Die für jeden Leistungsindikator erstellten Diagramme stellen jeweils die prozentuale Abweichung der ökologischen zur konventionellen Variante dar (konv. Variante = 0 %) ⁷. Die Länge der Whisker entspricht dem anderthalbfachen des Interquartilsabstandes (IQR). ⁸ Werte, die anderthalb bis dreifach über dem IQR lagen, wurden als „Ausreißer“ klassifiziert und als Punkte dargestellt. Ab dem dreifachen IQR wurden die Ergebnisse als „Extremwerte“ eingestuft und in der Bildunterschrift aufgeführt.

2.2.2 Klassifikation

In einer zweiten Auswertung wurden die Ergebnisse der ökologischen Variante der einzelnen Paarvergleiche hinsichtlich ihrer relativen Merkmalsausprägung im Vergleich zur konventionellen Variante in drei Gruppen eingeteilt:

- Höhere Umwelt- und Tierwohlleistungen durch ökologischen Landbau (**Öko +**)
- Vergleichbare Umwelt- und Tierwohlleistungen (**Öko =**)
- Niedrigere Umwelt- und Tierwohlleistungen durch ökologischen Landbau (**Öko -**)

Zuvor wurde für jeden Indikator festgelegt, (a) wann ein Unterschied zwischen der ökologischen und konventionellen Variante als eindeutig zu bewerten ist, und folglich von einem Unterschied in der Leistungserbringung auszugehen ist sowie (b) ob dieser Unterschied eine höhere, vergleichbare oder niedrigere Leistungserbringung anzeigt. Ein Unterschied wurde dann als eindeutig definiert, wenn dieser nach statistischer Prüfung in der Studie entweder signifikant war oder – falls keine Signifikanzangaben vorlagen – die Differenz einen bestimmten, indikatorspezifischen Schwellenwert überschritt (siehe Tabelle 2.2). Aufgrund der sehr unterschiedlichen und hohen Anzahl an Teilindikatoren, die für den Leistungsbereich Tierwohl herangezogen wurden, erfolgte die Klassifizierung auf der Grundlage der in den Studien ausgewiesenen Signifikanzwerte bzw. der in den Studien vorgenommenen Bewertung der Ergebnisse (siehe Tabelle 2.3).

⁷ Formel: $([\text{Wert d. Indikators Öko}] - [\text{Wert d. Indikators Konv}] / [\text{Wert d. Indikators Konv}]) * 100$

⁸ Der IQR zeigt den Bereich an, innerhalb dessen 50 % der Daten liegen.

Tabelle 2.2 Indikatoren und Klassifikationskriterien zur Beurteilung der gesellschaftlichen Leistungen des ökologischen Landbaus im Umweltbereich

Leistungs- bereiche	Indikator	Anzahl Studien	Anzahl VGP	Schwellenwerte zur Klassifizierung der gesellschaftlichen Leistung des des ökologischen Landbaus (falls keine Signifikanzangaben vorlagen)	
Wasser	Nitrat	79	226	+/- 20%	
	PSM	12	64		
	Phosphor	-	-	<i>nicht quantitativ ausgewertet</i>	
	TAM	-	-		
Boden	Regen- würmer	Abundanz	21	64	+/- 20%
		Biomasse	17	93	
	Phosphor		14	65	
	Eindringwiderstand		4	44	+/- 1%
	Bodenacidität		30	71	
Biodiversität	Ackerflora	Artenzahl	36	99	+/- 20%
		Abundanz	6	11	
	Acker- Samenbank	Artenzahl	5	15	
		Abundanz	2	6	
	Saum- vegetation	Artenzahl	8	14	
		Abundanz	1	2	
	Vögel	Artenzahl	13	23	
		Abundanz	14	34	
	Insekten	Artenzahl	20	44	
		Abundanz	16	64	
Klimaschutz	Boden / Pflanze	SOC-Gehalt	103	270	+/- 10%
		SOC-Vorrat	52	131	
		C-Speicherung	17	41	
		Lachgasemissionen	13	35	
		Methanemissionen	3	6	
		THG-Gesamt	-	-	
	Milchvieh- haltung	Methanemissionen	-	-	<i>nicht quantitativ ausgewertet</i>
		THG-Gesamt	-	-	
Klima- anpassung	Fruchtfolgeeffekte (C-Faktor)		3	6	+/- 10%
	Anteil organischer Substanz		24	71	
	Aggregatstabilität		22	76	
	Trockenraumdichte		13	30	
	Infiltration		11	28	
	Oberflächenabfluss		9	22	
Bodenabtrag		16	45		
Ressourcen- effizienz	N-Input		38	113	+/- 10%
	N-Effizienz		36	113	
	N-Saldo		38	114	
	Energieinput		55	141	
	Energieeffizienz		37	105	

Quelle: Eigene Darstellung

Tabelle 2.3 Indikatoren und Klassifikationskriterien zur Beurteilung der gesellschaftlichen Leistungen des ökologischen Landbaus im Bereich Tierwohl

Leistungs- bereiche	Indikator	Anzahl Studien	Anzahl VGP	Schwellenwerte zur Klassifizierung der gesellschaftlichen Leistung des des ökologischen Landbaus (falls keine Signifikanzangaben vorlagen)	
Tierwohl	Milchkühe	Tiergesundheit	46	286	Bewertung wurde aus Primärstudien übernommen
		Tierverhalten	3	10	
		Emotionales Befinden	1	3	
	Aufzucht- kälber	Tiergesundheit ^a	5	6	
		Tierverhalten	-	-	
		Emotionales Befinden	1	1	
	Mastrinder	Tiergesundheit	3	49	
		Tierverhalten	-	-	
		Emotionales Befinden	-	-	
	Schweine	Tiergesundheit	8	51	
		Tierverhalten	2	2	
		Emotionales Befinden	-	-	
	Legehennen	Tiergesundheit	6	28	
		Tierverhalten	2	4	
		Emotionales Befinden	3	5	
Schafe Ziegen	Tiergesundheit ^b	5	28		
	Tierverhalten	-	-		
	Emotionales Befinden	-	-		

^{a)} Davon zwei Studien mit 35 VGP zu Schlacht- und Mastrindern (Tiergesundheit) und eine Studie mit 14 VGP zu Mutterkühen und Kälbern

^{b)} Davon vier Studien mit 25 VGP zu Schafen und eine Studie mit 3 VGP zu Ziegen

Quelle: Eigene Darstellung

2.3 Ergänzende Informationen zum Vorgehen bei einzelnen Leistungsbereichen

2.3.1 Wasserschutz

Studienauswahl und Datenerfassung

- Im Rahmen der Literatursuche wurden je nach Indikator sowohl Experimentalstudien als auch Modell- und LCA-Untersuchungen berücksichtigt, die die Ergebnisse sowohl auf die Fläche als auch auf den Ertrag beziehen.
- Ergebnisse, die sich auf Sonderkulturen wie Gemüse, Obst und Wein, den Unterglas- oder Reisanbau beziehen, wurden ausgeschlossen.

Datenauswertung und Ergebnisdarstellung

- Für die deskriptiv-statistische Auswertung der Stoffgruppe *Stickstoff* wurden nur die Vergleichspaare berücksichtigt, die einen Wert für den Stickstoffaustrag pro Hektar, berechnet auf Basis von Sickerwasserproben, aufwiesen. Dieser Indikator gilt als aussagefähigster Indikator zur Quantifizierung der tatsächlichen Stickstoffausträge.
- Zum Großteil wurden die Ergebnisse in den Vergleichsstudien in kg Nitrat (NO_3)/ha und kg Nitratstickstoff ($\text{NO}_3\text{-N}$)/ha angegeben. Diese beiden Ergebnisformen wurden für die Boxplot-Darstellung in NO_3 in $\text{NO}_3\text{-N}$ umgerechnet.
- Neben $\text{NO}_3\text{-N}/\text{ha}$ und NO_3/ha wurden in einigen Vergleichsstudien die Ergebnisse für den ausgetragenen Stickstoffgehalt als kg Gesamt-N/ha angegeben. Teilweise wurde von den Autoren nicht genau definiert, um welche Stickstoffverbindungen es sich handelt (bspw. Seeger et al., 1997; Filipinski et al., 2009). Von anderen Autoren wurde wiederum neben der Bestimmung von Nitrat und Ammonium (kg $\text{N}_{\text{min}}/\text{ha}$) in Sickerwasserproben auch von einer Bestimmung von organischem Stickstoff berichtet. Trotz der unterschiedlichen Einheiten wurden die Ergebnisse der Arbeiten in einer Boxplot-Darstellung zusammengefasst.

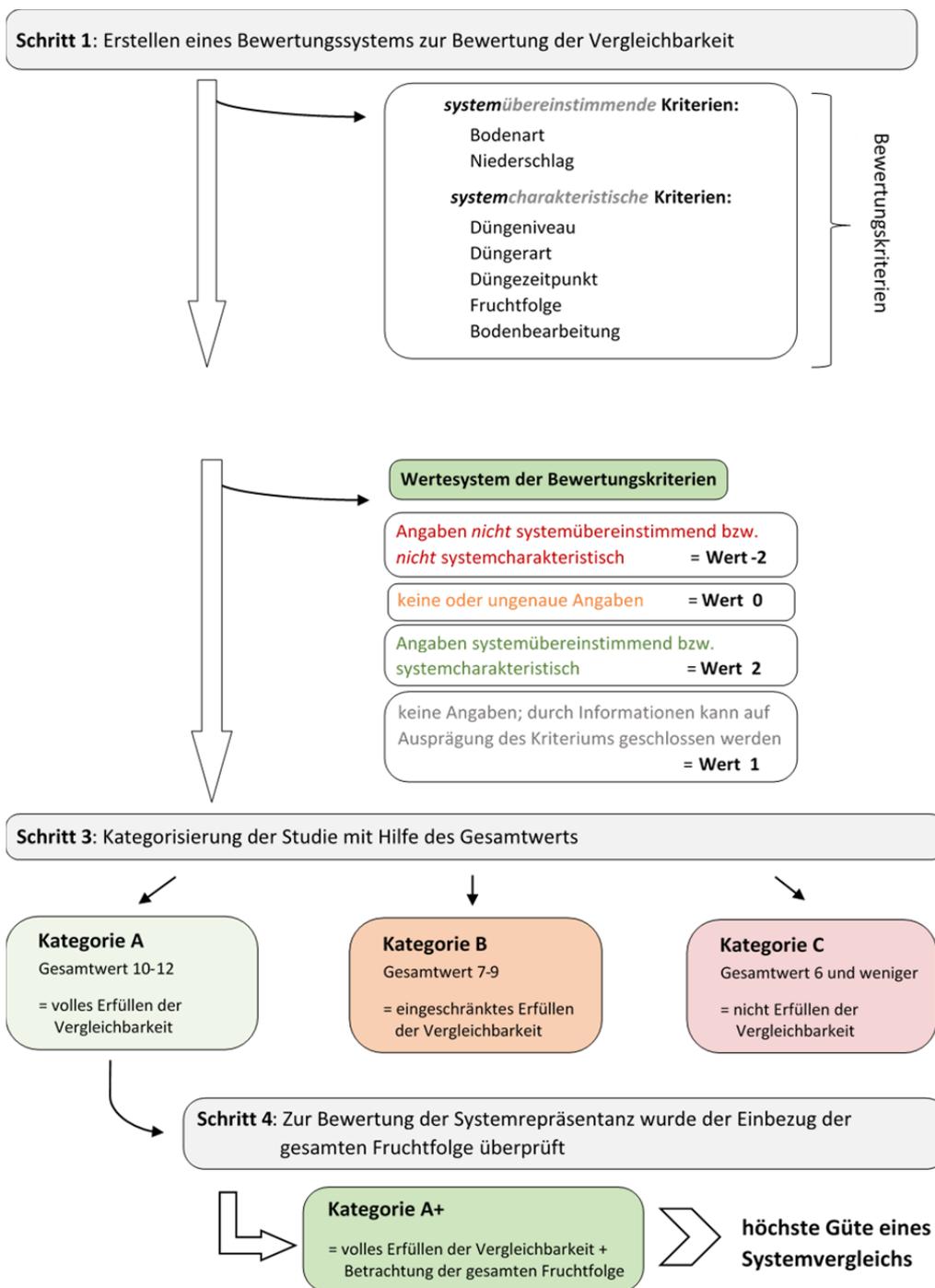
Dieses Vorgehen ist begründbar, da die Grundwasserbelastung mit Stickstoffverbindungen im Allgemeinen nahezu vollständig durch Nitrat verursacht wird (Rohmann und Sontheimer, 1985). Nur auf sehr sandigen und durchlässigen Böden sowie bei niedrigen pH-Werten kann auch eine Ammoniumauswaschung zum Stickstoffeintrag in das Grundwasser beitragen (Scheffer et al., 2010). Diese befindet sich allerdings auf einem sehr niedrigen Niveau von ca. 1,5 kg Ammoniumstickstoff/ha, wie die Studie von Syväsalo et al. (2006) zeigt. Von einer Auswaschung von organischem Stickstoff wird in der Literatur nicht berichtet. Aufgrund dessen wurde unterstellt, dass der Anteil an organischem Stickstoff und Ammoniumstickstoff zu vernachlässigen ist. Weiterhin ist davon auszugehen, dass, wenn die Autoren den ausgetragenen Stickstoff berechnen, auch alle Formen des N-Austrags in den jeweiligen Sickerwasserproben bestimmt wurden. Dieses ist wie bereits erwähnt in den meisten Fällen $\text{NO}_3\text{-N}$ und in nur wenigen Fällen zusätzlich organischer Stickstoff bzw. Ammoniumstickstoff.

- Für die Stoffgruppe *Stickstoff* erfolgte die Klassifizierung der Vergleichspaare unter Verwendung des aussagefähigsten Indikators (bzw. Untersuchungsmethode) einer Studien; die Aussagefähigkeit wurde absteigend folgendermaßen definiert: berechneter N-Austrag > Sickerwasserproben > Bodenproben > Feldbilanzen > Hoftorbilanzen.
- Aufgrund zu weniger Vergleichspaare wurden die Ergebnisse für die Leistungsindikatoren Pflanzenschutzmittel, Phosphor und Tierarzneimittel nicht deskriptiv-statistisch ausgewertet. Aufgrund gänzlich fehlender Vergleichsstudien wurden die Unterschiede zwischen der ökologischen und konventionellen Bewirtschaftung hinsichtlich des Austrags von Tierarzneimitteln nur qualitativ bewertet. Da die ermittelten absoluten Unterschiede bei den Austräge von Phosphor in das Grundwasser sich auf einem sehr niedrigen Niveau befanden und zudem nur eine geringe Anzahl an Studien mit nur bedingt aussagekräftigen Teilindikatoren vorlagen, wurde auch für diesen Indikator keine quantitative Klassifizierung vorgenommen.

Gütebeurteilung

- Ein differenzierter Systemvergleich zwischen konventioneller und ökologischer Landwirtschaft setzt voraus, dass sowohl die konventionelle als auch die ökologische Betriebsform in der Untersuchung vergleichbar und repräsentativ ist. Exemplarisch wurde deshalb für den Indikator *Stickstoffaustrag* eine Gütebeurteilung der ausgewählten Studien hinsichtlich der Vergleichbarkeit und Systemrepräsentanz der ökologischen und konventionellen Untersuchungsvarianten durchgeführt. Die Vergleichbarkeitsbewertung der in den Studien untersuchten konventionellen und ökologischen Bewirtschaftung wurde durch die Prüfung der systemübereinstimmenden Faktoren (Klima, Boden) sowie der systemcharakteristischen Faktoren (Düngeniveau, Düngeart, Fruchtfolgegestaltung) vorgenommen. Um die Repräsentativität der Varianten zu überprüfen, wurde untersucht, ob die Ergebnisse sich auf die gesamte Fruchtfolge beziehen und somit das gesamte System erfasst wurde.
- Wie Abbildung 2.1 zeigt, gliedert sich diese Bewertung in vier Schritte. In **Schritt 1** wurde ein Bewertungssystem erstellt. Es umfasst die Definition systemübereinstimmender sowie systemcharakteristischer Kriterien und weitere notwendige Informationen zur Versuchsdurchführung für den Gewässerschutz. Ausschlaggebend für die Erstellung des Bewertungssystems war die Fragestellung, welche Faktoren den Stickstoffaustrag in das Grundwasser beeinflussen. Dazu zählen hauptsächlich Bodenart, Niederschlagsmenge, Düngeniveau, Düngerart und Fruchtfolgegestaltung. Weiterhin haben auch die Faktoren Bodenbearbeitung und Düngezeitpunkt einen Einfluss auf den Stickstoffaustrag. Die standortabhängigen Faktoren Bodenart und Niederschlag sind nicht zu beeinflussende Faktoren. Dagegen werden die Faktoren Düngeniveau, Düngerart und Fruchtfolgegestaltung durch das Bewirtschaftungssystem stark beeinflusst. Die systemübereinstimmenden Faktoren Bodenart und Niederschlag müssen in beiden untersuchten landwirtschaftlichen Varianten übereinstimmen, damit eine Vergleichbarkeit gegeben ist. Zum anderen wird der Systemvergleich durch die systemcharakteristischen Faktoren Düngeniveau, Düngerart und Fruchtfolgegestaltung geprägt, die für das jeweilige Anbausystem charakteristisch gestaltet sein müssen. Nur wenn die genannten Faktoren in einem Systemvergleich in dieser Weise gestaltet sind, können die in der Studie untersuchten Varianten als vergleichbar bezeichnet werden.
- In **Schritt 2** erfolgte eine Gewichtung der Bewertungskriterien. Angaben zu den Kriterien in den Studien, die als nicht-systemübereinstimmend bzw. -charakteristisch bewertet wurden, erhielten den Wert -2. Wurden keine oder ungenaue Angaben getätigt, wurde der Wert 0 vergeben. Handelte es sich um systemübereinstimmende bzw. -charakteristische Angaben so entsprach dies dem Wert +2. Ein Sonderfall ergab sich, wenn zwar keine direkten Angaben zu den Kriterien vorlagen, jedoch auf Grund von Informationen aus dem Text auf die Ausprägung des Kriteriums geschlossen werden konnte. Hier wurde der Wert +1 vergeben (siehe Tabelle 2.4). Diese Gewichtung wurde für jedes der sieben Bewertungskriterien (Bodenart, Niederschlagsmenge, Düngeniveau, Düngerart, Fruchtfolgegestaltung, Bodenbearbeitung, Düngezeitpunkt) durchgeführt, woraus ein Gesamtwert für jede Studie errechnet werden konnte.

Abbildung 2.1 Schematische Darstellung der Gütebeurteilung der Vergleichbarkeit und der Systemrepräsentanz



Quelle: Eigene Darstellung

Tabelle 2.4 Bewertungskriterien und Gesamtwert-Punkte der Gütebewertung

Bewertungskriterium	Wert
- Systemübereinstimmend, systemcharakteristisch	2
- Keine explizite Angaben, aber Hinweise, dass eine Systemübereinstimmung vorliegt	1
- Keine/ungenauere Aussage	0
- Nicht systemübereinstimmend	-2

Quelle: Eigene Zusammenstellung

- Die Gewichtung ist notwendig, um den Einfluss fehlender oder unzureichender Angaben bzgl. der systemcharakteristischen Faktoren nicht überzubewerten. Wenn das systemcharakteristische Kriterium beispielsweise die Fruchtfolge ist, so ist es durch eine Gewichtung möglich, Studien, welche die Fruchtfolge nur teilweise bzw. unvollständig darstellen und Studien die keine systemcharakteristische Fruchtfolge beinhalten, zu unterscheiden. Ohne eine Gewichtung würden unvollständige, aber möglicherweise systemcharakteristische Fruchtfolgen und systemuntypische Fruchtfolgen gleichwertig behandelt. Die Gewichtung ermöglicht somit eine Differenzierung und angemessene Bewertung der aufgestellten Kriterien.
- Der auf Grundlage der Bewertungskriterien und Addition der Punkte ermittelte *Gesamtwert* jeder Studie bzw. jedes Vergleichspaars diente in **Schritt 3** dazu, die Studien in die drei Kategorien A, B und C einzustufen. Studien bzw. Vergleichspaare mit einem Gesamtwert von 8 bis 10 wurden in Kategorie A eingestuft. Dies entspricht dem vollen Erfüllen der Vergleichbarkeit. In Kategorie B (eingeschränkte Vergleichbarkeit), wurden Studien mit einem Gesamtwert von 5 bis 7 eingestuft. Wiesen die Studien einen Gesamtwert von 4 und weniger auf, entsprach dies der Kategorie C und somit einem Nicht-Erfüllen der Vergleichbarkeit.
- In **Schritt 4** wurden Studien bzw. Vergleichspaare aus der Kategorie A, in Kategorie A+ eingestuft wenn sie gesamte Fruchtfolge untersuchten und somit neben der Vergleichbarkeit auch das Kriterium der Systemrepräsentanz erfüllten.

2.3.2 Bodenfruchtbarkeit

Studienauswahl und Datenerfassung

- Bei der Literaturrecherche wurden ausschließlich Ackerbausysteme berücksichtigt. Der Fokus der Suche lag dabei auf Flächen, die zur Produktion von primären Grundnahrungspflanzen genutzt werden, d.h. in erster Linie (a) Getreidearten wie Weizen, Mais, Roggen, Gerste, Hafer, (b) Leguminosen wie Ackerbohnen, Erbsen, Lupinen, (c) Ölfrüchte wie z. B. Raps, Sonnenblumen sowie (d) Wurzel- und Knollenfrüchte wie z. B. Kartoffeln und Zuckerrüben.
- Obwohl Grünlandflächen einen bedeutenden Anteil der landwirtschaftlichen Nutzfläche einnehmen, wurden diese bei der Literatursuche in der vorliegenden Studie ausgeschlossen. Diese Eingrenzung wurde vorgenommen, weil sich Ackerbausysteme und Grünlandflächen in ihrer Ökologie und in der Art der Bewirtschaftung erheblich voneinander unterscheiden.
- Es wurden ferner nur Freilandstudien verwendet; Versuche in Gewächshäusern etc. wurden ausgeschlossen.
- Es wurde eine Eingrenzung auf Europa und Nordamerika vorgenommen. Innerhalb der Kontinente wurde die Datenauswahl auf die Bereiche der klimatischen Zone C und D beschränkt (Klassifikationen nach (Köppen und Geiger, 1954; Peel et al., 2007). Die Studienauswahl in Europa beinhaltete die Klimazonen Cfb und Dfb. In Nordamerika wurden Studien ausgewählt, die nach Schulz (2016) der Ökozone der sogenannten „Feuchten Mittelbreiten“ zugeordnet werden können.

2.3.3 Biodiversität

Studienauswahl und Datenerfassung

- In die Literaturlauswertung wurden Untersuchungen zur Flora (Ackerflora und Acker-Samenbank sowie Saumvegetation) einbezogen. Des Weiteren wurden Vergleichsstudien zu Vögeln (mit besonderem Augenmerk auf die Feldlerche) sowie zu bestimmten blütenbesuchenden Insekten (Wildbienen inklusive Hummeln, Honigbienen, Tagfalter, Schwebfliegen und Florfliegen) ausgewählt. In allen Studien wurden die Artenzahl und/oder Abundanz zur Charakterisierung der Biodiversität herangezogen, so dass die Auswertung auf Grundlage dieser beiden Indikatoren erfolgte.
- In Bezug auf die genannten Indikatoren erfolgte in den meisten Studien eine Differenzierung auf unterschiedlichen Ebenen und Skalen. Vielfach waren Mittelwerte und/oder Gesamtanzahlen angegeben. Aus diesem Grund wurde bei der Auswertung folgende Unterteilung der Indikatoren vorgenommen: mittlere und Gesamt-Artenzahl sowie mittlere und Gesamt-Abundanz. In bestimmten Fällen wurden Einzelwerte zusammengefasst. Dies war zum Beispiel der Fall, wenn Artengruppen funktionell untergliedert und die Ergebnisse getrennt dargestellt wurden. Auch bei Angabe einzelner Monatswerte wurden diese zusammengefasst, um eine Überbewertung von Einzelergebnissen zu vermeiden.

- Der im Juli 2017 durchgeführte systematische Suchdurchlauf wurde im weiteren Verlauf um einzelne relevante Publikationen mit späterem Erscheinungsdatum ergänzt. Alle zusätzlichen Quellen wurden den gleichen Auswahlkriterien unterzogen.
- Berücksichtigt wurden Studien, die sich auf die gemäßigten Klimazonen beziehen, d.h. alle C-Klimate, Dfa, Dfb, Dwa, Dwb, BSk. Ein Abgleich der Studien erfolgte nach der Klimaklassifikation von Köppen-Geiger über Google Earth-Karte (Kottek et al., 2006).
- Studien, die sich auf den Reisanbau, die Aquakultur, die Gewächshausproduktion oder auf Topfversuche beziehen, wurden nicht berücksichtigt. Ferner wurden Untersuchungen mit ausschließlich modellbasierten Ergebnissen sowie Arbeiten mit deutlichen methodischen Mängeln (z. B. stark abweichende Stichprobengröße zwischen ökologischen und konventionellen Vergleichspaaren) ausgeschlossen. Studien, die keine Angabe zur Dauer der ökologischen Bewirtschaftung nach Umstellung machten, wurden bei sonstiger methodischer Eignung einbezogen.

Datenauswertung und Ergebnisdarstellung

- Um eine statistisch aussagefähigere Bewertung zu erhalten, wurden in Studien ohne Aussagen zur Signifikanz der Ergebnisse aber mit angegebenen Standardfehlern (SE) die Unterschiede zwischen den Mittelwerten der Vergleichspaare mit Hilfe des Standardfehlers verifiziert. Ab einer Abweichung von $3 \times SE$ wurden sie als signifikant bewertet und entsprechend gekennzeichnet. Diese Bewertung diente als Vergleichswert für die Güte der 20 %-Grenze. Veränderungen in der Zuordnung der Paarvergleiche zu den Kategorien (Öko + / Öko = / Öko -) wurden auf dieser Basis in den Tabellen nicht vorgenommen.

2.3.4 Klimaschutz

Studienauswahl und Datenerfassung

- Für den Zeitraum bis 2011 berücksichtigt die Arbeit einen Datensatz über die Speicherung des Bodenkohlenstoffs in der ökologischen und konventionellen Landwirtschaft, der für die Metaanalyse von Gattinger et al. (2012) verwendet wurde. Darüber hinaus basiert die Analyse auf weiteren Daten, die nach 2011 publiziert wurden.
- Diese Datensatz wurde ergänzt um Ergebnisse wissenschaftlicher Studien, die seit 2012 veröffentlicht wurden. Graue Literatur wurde nicht verwendet.
- Es wurden nur Freilandversuche (keine Gewächshaus-, Topf- oder Inkubationsstudien) in die Analyse mit einbezogen.

Datenauswertung und Ergebnisdarstellung

- Um eine verlässliche Datengrundlage zu gewährleisten, wurde weitestgehend darauf verzichtet, Werte zu berechnen; stattdessen wurden nur gemessene Werte genutzt. Die einzigen Umrechnungen, die vorgenommen wurden, waren:

- bei Quellen, die nur die organische Bodensubstanz (SOM, soil organic matter) nicht aber den organischen Bodenkohlenstoff (SOC, soil organic carbon) angaben, wurde letzterer folgendermaßen berechnet: $SOC \% = SOM \% / 1,72$
- wenn sowohl die Vorräte an Bodenkohlenstoff zu Beginn (t_0) als auch zu Ende der Messperiode (t_n) in t/ha als Messwerte vorlagen und die Dauer der Messperiode in Jahren (n) angegeben war, konnte daraus die Kohlenstoffspeicherung folgendermaßen errechnet werden: $(t_n - t_0) / n = \text{Kohlenstoffspeicherung (kg C/ha und Jahr)}$
- Da bei den Paarvergleichen für den Indikator Kohlenstoffspeicherung sowohl positive als auch negative Werte vorkamen, konnten nicht wie für die anderen Indikatoren prozentuale Unterschiede gebildet werden. Stattdessen wurden die Differenzen in absoluten Werten (kg C/ha und Jahr) angegeben.
- Aufgrund des Mangels an experimentellen Vergleichsstudien zu ertragsskalierten Treibhausgasemissionen, konnte dieser Indikator nur qualitativ beurteilt werden.
- Obwohl Methan (CH_4) aus der enterischen Fermentation der Wiederkäuer vorwiegend aus der Rinderhaltung die zweitgrößte Einzelquelle der landwirtschaftlichen THG-Emissionen in Deutschland und weltweit darstellt, gibt es bislang keine experimentellen Vergleichsuntersuchungen ökologischer und konventioneller Rinderhaltung. Daher muss an dieser Stelle auf LCA-Vergleichen und anderen Modellierungen zurückgegriffen werden. Da diesen unterschiedlichen methodischen Herangehensweisen und Systemgrenzen zugrunde liegen, wurde hierzu nur eine qualitative Beurteilung durchgeführt.

Metaanalyse

- Für den Indikator *Organischer Bodenkohlenstoff-Gehalt* wurde ergänzend eine statistische Meta-Analyse durchgeführt. Hierfür wurden aus einigen Studien mehrere Vergleichspaare gebildet. Wenn beispielsweise in einem einheitlichen Kontext drei ökologisch (Öko 1, Öko 2, Öko 3) und eine konventionell bewirtschaftete Fläche (Konv. 1) miteinander verglichen wurden, konnten folgende drei Paare gebildet werden: Öko 1 - Konv. 1, Öko 2 - Konv. 1 und Öko 3 - Konv. 1. Das Vergleichspaar Konv. 1 floss also dreimal in die Berechnung ein, es wurden somit zwei *Pseudoreplikationen* oder *nicht unabhängige Daten* gebildet. In einer eigens hierfür erzeugten Spalte wurde für jeden Eintrag festgehalten, ob es sich um unabhängige und nicht unabhängige Daten handelt.
- Oftmals werden in der Literatur unterschiedliche Einheiten für dieselben Indikatoren verwandt (z.B. kg/ha und Tag, kg/ha und Jahr, g/m³ und Stunde, etc.). In diesen Fällen mussten die Daten auf dieselbe Einheit umgerechnet werden. Aus Transparenzgründen wurde dies in der Eingabetabelle vermerkt.
- Für die Auswertung der Daten im Rahmen der Metaanalyse wurde das Metafor-Package der Software R (Viechtbauer, 2010) verwendet.

- Durch die Bildung und Analyse von Untergruppen zur Laufzeit des Paarvergleichs sowie Beprobungshorizont wurden zwei erklärende Variablen überprüft und kausale Zusammenhänge getestet.
- Die drei wichtigsten Angaben, die in die Metaanalyse einfließen waren: (a) der Mittelwert als ein Maß für den Indikator „Bodenkohlenstoffgehalt auf ökologisch und konventionell bewirtschafteten Flächen“, b) die Standardabweichung als ein Maß für die Einheitlichkeit der gemessenen Werte sowie (c) die Zahl der unabhängigen Messungen (n) als ein Maß dafür, wie zuverlässig der Mittelwert die tatsächlichen Verhältnisse wiedergibt. Aus diesen Angaben wurde die Response Ratio (RR) berechnet, die ein Maß für die Differenz des Indikators zwischen ökologisch und konventionell bewirtschafteten Flächen ist.

2.3.5 Klimaanpassung

Studienauswahl und Datenerfassung

- Studien außerhalb des Zeitraumes 1990 - März 2017 wurden einbezogen (vor 1990 ergänzende Suche; nach März 2017 systematisch in Web of Science und Scopus). Darüber hinaus wurde keine weitere Literatur mit einbezogen.
- Acker- und Futterbaufruchtfolgen, Sonderkulturen wie Gemüse, Obst und Wein sowie Studien zum Anbau unter Glas wurden mit einbezogen. Berücksichtigt wurden ferner auch Ergebnisse aus Modellrechnungen.
- Es wurden Indikatoren eingezogen, die direkt den Oberflächenabfluss und die Erosion abbilden sowie Indikatoren, die in der Allgemeinen Bodenabtragungsgleichung (ABAG) eine bewirtschaftungsabhängige Rolle spielen. Insgesamt konnten sieben Indikatoren einbezogen werden. Weitere wichtige Indikatoren wie die Bodenbedeckung waren in zu wenigen Studien angegeben.

Datenauswertung und Ergebnisdarstellung

- Für den Indikator *C_{org}-Gehalt* wurde, wenn nötig, soil organic matter (SOM) in *C_{org}* umgerechnet ($C_{org} (\%) = \text{SOM} (\%) / 1,72$).
- Die deskriptiv-statistische Auswertung und die Klassifikation der Leistungen des ökologischen Landbaus erfolgte unter Verwendung der jeweils aussagefähigsten Untersuchungsmethode innerhalb der Studien, die absteigend folgendermaßen definiert wurde:
 - Aggregatstabilität: Anteil wasserstabile Aggregate > geometrische mittlere Durchmesser > Perkolations > Anteil Makroaggregate; Anteil wasserstabile Makroaggregate > Anteil wasserstabile Mikroaggregate; geometrische mittlere Durchmesser Wasser > geometrische mittlere Durchmesser trocken.
 - Infiltration: Infiltration > gesättigte hydraulische Leitfähigkeit (nur in situ Messungen).

- Oberflächenabfluss: Oberflächenabfluss gesamt > Schwebstoffe/gelöste Feststoffe > C_{org} im Oberflächenabfluss (Sedimente).
- Bodenabtrag: Bodenabtrag gemessen > Bodenabtrag modelliert > Bodenabtrag beobachtet.

2.3.6 Ressourceneffizienz

Studienauswahl und Datenerfassung

- Um Überschneidungen zu den anderen Leistungsbereichen zu vermeiden, wurden im Bereich Ressourceneffizienz die Nitratauswaschung (Auswertung erfolgt im Bereich Wasserschutz) und die Treibhausgasemissionen (erfolgt im Bereich Klimaschutz) nicht behandelt.
- Für die Auswertungen wurden nur die Studien berücksichtigt, die Daten auf der Systemebene von Fruchtfolgen oder der Fruchtart Weizen beinhalten. Weizen wurde gewählt, weil die Kultur in den Studien am häufigsten vertreten ist. Die Analyse ist somit auf den Pflanzenbau fokussiert.
- Die Werte mussten den Praxisbedingungen und rechtlichen Vorgaben entsprechen. Die ungedüngten „Nullvarianten“ in Feldexperimenten wurden nicht als Varianten des ökologischen Landbaus einbezogen.
- Von jeder Studie wurden folgende Informationen erfasst: Datenursprung, Betriebstyp, Fruchtarten und Anzahl der Fruchtarten in der Fruchtfolge, % Anteil der Leguminosen in der Fruchtfolge, Bodenbedingungen.
- Soweit verfügbar wurden zum Stickstoff- und Energieinput folgende zusätzliche Angaben aus den Studien extrahiert:
 - Stickstoffinput: Angaben zur N_2 -Fixierleistung der Leguminosen, Gesamt-N in organischen und mineralischen Düngern, atmosphärischer N-Deposition und N im Saatgut
 - Energieinput: Angaben zum direkten Energieinput (Kraftstoffe) und zum indirekten Energieinput (Maschinen, Geräte, Saat- und Pflanzgut, organische/mineralische Dünger, Pflanzenschutzmittel)

Datenauswertung und Ergebnisdarstellung

- In den Energiebilanzen wurden entsprechend der Methodik der energetischen Prozessanalyse die Sonnenenergie und die menschliche Arbeitskraft nicht als Energieinput einbezogen.
- Die Werte der verwendeten Indikatoren (N-Input, N-Output, N-Saldo, Energieinput, Energieoutput) wurden flächenbezogen in den Einheiten kg/ha und Jahr (Stickstoffbilanz) und GJ/ha und Jahr (Energiebilanz) dargestellt.

- Als produktbezogene Indikatoren wurden die Energieeffizienz (Energieoutput/Energieinput) und Stickstoffeffizienz (Stickstoffoutput/Stickstoffinput) verwendet.
- Werte, die in den Studien für eine Fruchtfolge oder für eine Untersuchungsperiode angegeben wurden, wurden auf ein Jahr umgerechnet.
- Unterschiedliche Einheiten (z. B. Btu, kcal) wurden auf die oben genannten Einheiten umgerechnet. Falls es möglich war und notwendige Informationen vorhanden waren, wurden auch produktbezogene Werte umgerechnet.
- Wenn nicht alle Indikatoren in der Publikation vorhanden waren, aber gleichzeitig andere Indikatoren vorlagen, die zur Berechnung der fehlenden Indikatoren nötig waren, wurden diese berechnet und in der Tabelle markiert.
- Eine Stickstoff- oder Energieeffizienz über 100 % wurde für die Analyse auf 100 % gesetzt.
- Für die meisten Vergleichsstudien lagen keine statistischen Tests aufgrund des Versuchsdesigns und der Fragestellung vor.

2.3.7 Tierwohl

Studienauswahl und Datenerfassung

- Es wurden ausschließlich Studien berücksichtigt, die Daten aus Praxisbetrieben analysiert hatten, um die Komplexität der landwirtschaftlichen Praxis abzubilden, in der viele verschiedene Einflussfaktoren miteinander interagieren. Somit wurden Untersuchungen mit dem Fokus auf Versuchsanstellungen, in denen nur einzelne Aspekte bzw. Vorgaben zur ökologischen Tierhaltung „simuliert“ wurden, von der Auswertung ausgeschlossen.
- Die Vergleiche der Untersuchungen mussten auf tatsächlich ermittelten Werten für die Indikatoren des Tierwohls basieren. Es wurden auch Studien aufgenommen, die diese „gemessenen Werte“ in statistische Analysen einfließen ließen und somit eine abschließende Bewertung vornahmen, ob Unterschiede zwischen den Wirtschaftsweisen vorlagen. Da im Bereich Tierwohl keine deskriptiv-statistische Analyse durchgeführt wurde (siehe Begründung unten), war es nicht notwendig, dass die Werte für jedes Vergleichspaar in der Publikation ausgewiesen wurden. Wenn sich der Unterschied zwischen den Wirtschaftsweisen laut Angaben der Studienautoren als signifikant erwies, wurde dieses in der tabellarischen Darstellung im Ergebnisteil berücksichtigt.
- Vergleichsstudien wurden berücksichtigt, wenn sie mindestens zwei Betriebe pro Wirtschaftsweise miteinander verglichen und somit mindestens zwei konventionell - ökologische Vergleichspaare aufwiesen. Das war in zwei der verwendeten Vergleichsstudien der Fall. Die anderen berücksichtigten mindestens fünf Betriebe einer Wirtschaftsweise in ihren Untersuchungen.

Datenauswertung und Ergebnisdarstellung

- Es wurde in der Literatur eine Vielzahl von Indikatoren für einzelne Dimensionen des Tierwohls, zumeist bezüglich Tiergesundheit, identifiziert. Gleichzeitig war die Anzahl an Studien bzw. Vergleichspaaren je Indikator sehr gering, bei teils sehr unterschiedlichen statistischen Kennzahlen für dieselben Indikatoren. Deshalb lagen nur für wenige Einzelindikatoren neun Vergleichspaare vor. Eine deskriptiv statistische Auswertung wurde deshalb nicht durchgeführt. Bei einer entsprechenden Auswertung hätte zudem ausschließlich die Tierwohldimension Gesundheit adressiert werden können und diese auch lediglich mit einigen wenigen Tiergesundheitsbereichen respektive Tierarten.